Lecture Notes on Data Engineering
and Communications Technologies 2

Leonard Barolli
Fatos Xhafa
Kangbin Yim *Editors*

Advances on Broad-Band Wireless Computing, Communication and Applications

Proceedings of the 11th International Conference on Broad-Band Wireless Computing, Communication and Applications (BWCCA–2016) November 5–7, 2016, Korea

Springer

Lecture Notes on Data Engineering and Communications Technologies

Volume 2

The aim of the book series is to present cutting edge engineering approaches to data technologies and communications. It publishes latest advances on the engineering task of building and deploying distributed, scalable and reliable data infrastructures and communication systems.

The series has a prominent applied focus on data technologies and communications with aim to promote the bridging from fundamental research on data science and networking to data engineering and communications that lead to industry products, business knowledge and standardisation.

More information about this series at http://www.springer.com/series/15362

Leonard Barolli · Fatos Xhafa
Kangbin Yim
Editors

Advances on Broad-Band Wireless Computing, Communication and Applications

Proceedings of the 11th International Conference on Broad-Band Wireless Computing, Communication and Applications (BWCCA-2016) November 5–7, 2016, Korea

Springer

Editors

Leonard Barolli
Fukuoka Institute of Technology
Fukuoka
Japan

Fatos Xhafa
Technical University of Catalonia
Barcelona
Spain

Kangbin Yim
Department of Information Security Engineering
Soonchunhyang University
Asan-si
Korea (Republic of)

ISSN 2367-4512 ISSN 2367-4520 (electronic)
Lecture Notes on Data Engineering and Communications Technologies
ISBN 978-3-319-49105-9 ISBN 978-3-319-49106-6 (eBook)
DOI 10.1007/978-3-319-49106-6

Library of Congress Control Number: 2016956190

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Welcome Message of BWCCA-2016 International Conference Organizers

Welcome to the 11-th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA-2016), which will be held in conjunction with the 11-th 3PGCIC-2016 International Conference from November 5 to November 7, 2016 at Soonchunhyang (SCH) University, Asan, Korea.

This International Conference is a forum for sharing ideas and research work in the emerging areas of broadband and wireless computing. Information networking are going through a rapid evolution. Different kinds of networks with different characteristics are emerging and they are integrating in heterogeneous networks. For these reasons, there are many interconnection problems which may occur at different levels of the hardware and software design of communicating entities and communication networks. These kinds of networks need to manage an increasing usage demand, provide support for a significant number of services, guarantee their QoS, and optimize the network resources.

The success of all-IP networking and wireless technology has changed the ways of living the people around the world. The progress of electronic integration and wireless communications is going to pave the way to offer people the access to the wireless networks on the fly, based on which all electronic devices will be able to exchange the information with each other in ubiquitous way whenever necessary.

The aim of this conference is to present the innovative research and technologies as well as developments related to broadband networking, and mobile and wireless communications. BWCCA-2016 received 195 paper submissions and based on review results, we accepted 53 papers (about 27% acceptance ratio) for presentation in the conference and publication in the Springer Lecture Notes on Data Engineering and Communication Technologies Proceedings.

The organization of an International Conference requires the support and help of many people. A lot of people have helped and worked hard to produce a successful BWCCA-2016 technical program and conference proceedings. First, we would like to thank all authors for submitting their papers, Program Committee Members and reviewers who carried out the most difficult work by carefully evaluating the submitted papers.

This year in conjunction with BWCCA-2016 we have 7 International Workshops that complemented BWCCA-2016 program with contributions for specific topics. We would like to thank the Workshop Co-Chairs and all workshops organizers for organizing these workshops.

We thank Shinji Sakamoto, Donald Elmazi and Yi Liu, Fukuoka Institute of Technology (FIT), Japan, as Web Administrator Co-Chairs and Dr. Makoto Ikeda, FIT, Japan, as Finance Chair for their excellent work.

We would like to express our gratitude to Prof. Makoto Takizawa, Hosei University, Japan and Prof. Kyoil Suh, Soonchunhyang University, Korea as Honorary Co-Chairs of BWCCA-2016 for their support and help.

We give special thanks to Prof. Nobuo Funabiki, Okayama University, Japan for kindly accepting to be Keynote Speaker of BWCCA-2016.

Finally, we would like to thank the Local Arrangement Team for making excellent local arrangement for the conference.

We hope you will enjoy the conference and have a great time in Asan, Korea.

BWCCA-2016 International Conference Organizers

BWCCA-2016 General Co-Chairs

Leonard Barolli, Fukuoka Institute of Technology (FIT), Japan
Fatos Xhafa, Technical University of Catalonia, Spain
Kangbin Yim, Soonchunhyang University, Korea

BWCCA-2016 Program Committee Co-Chairs

Yunyoung Nam, Soonchunhyang University, Korea
Tetsuya Shigeyasu, Prefectural University of Hiroshima, Japan
Marek R. Ogiela, AGH University of Science and Technology, Krakow, Poland

Welcome Message from BWCCA-2016 Workshops Co-Chairs

Welcome to the Workshops of the 11-th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA-2016), which will be held in conjunction with the 11-th 3PGCIC-2016 International Conference from November 5 to November 7, 2016 at Soonchunhyang (SCH) University, Asan, Korea.

This year 7 workshops will be held in conjunction with BWCCA-2016 International Conference. The workshops are very important part of the main conference and they cover specific topics related to next generation networks, network traffic analysis, sensor technologies, smart environments, complex systems, wireless communication, mobile networks and multimedia networking.

BWCCA-2016 workshops are listed in following:

1. The 18-th International Symposium on Multimedia Network Systems and Applications (MNSA-2016)
2. The 9-th International Workshop on Next Generation of Wireless and Mobile Networks (NGWMN-2016)
3. The 7-th International Workshop on Methods, Analysis and Protocols for Wireless Communication (MAPWC-2016)
4. The 7-th International Workshop on Cloud, Wireless and e-Commerce Security (CWECS-2016)
5. The 5-th International Workshop on Robot Interaction, Control, Communication and Cooperation (RI3C-2016)
6. The 3-rd International Workshop on Secure Cloud Computing (SCC-2016)
7. The 3-rd International Workshop on Large Scale Networks and Applications (LSNA-2016)

These workshops bring to the researchers conducting research in specific themes the opportunity to learn from this rich multi-disciplinary experience.

The Workshop Chairs would like to thank the workshop organizers for their great efforts and hard work in proposing the workshop, selecting the papers, the interesting programs and for the arrangements of the workshop during the conference days.

We hope you enjoy the workshops programs and proceedings.

BWCCA-2016 Workshops Co-Chairs

Cheongghil Kim, Namseoul University, Korea
Lidia Ogiela, AGH University of Science and Technology, Krakow, Poland
Elis Kulla, Okayama University of Science, Japan

BWCCA-2016 Organizing Committee

Honorary Chairs

Makoto Takizawa, Hosei University, Japan
Kyoil Suh, Soonchunhyang University, Korea

General Co-Chairs

Leonard Barolli, Fukuoka Institute of Technology, Japan
Fatos Xhafa, Universitat Politècnica de Catalunya, Spain
Kangbin Yim, Soonchunhyang University, Korea

Program Committee Co-Chairs

Yunyoung Nam, Soonchunhyang University, Korea
Tetsuya Shigeyasu, Prefectural University of Hiroshima, Japan
Marek R. Ogiela, AGH University of Science and Technology, Krakow, Poland

Workshop Co-Chairs

Cheongghil Kim, Namseoul University, Korea
Lidia Ogiela, AGH University of Science and Technology, Krakow, Poland
Elis Kulla, Okayama University of Science, Japan

Finance Chairs

Makoto Ikeda, Fukuoka Institute of Technology, Japan

Web Administrator Chairs

Shinji Sakamoto, Fukuoka Institute of Technology, Japan
Donald Elmazi, Fukuoka Institute of Technology, Japan
Yi Liu, Fukuoka Institute of Technology, Japan

Local Organizing Co-Chairs

Sunyoung Lee, Soonchunhyang University, Korea
Hwamin Lee, Soonchunhyang University, Korea
Yunyoung Nam, Soonchunhyang University, Korea

Track Areas

1. Wireless Networks and Applications

Chairs:

Hsing-Chung Chen, Asia University, Taiwan
Safdar Hussain Bouk, Kyungpook National University, Korea
Jing Li, Xidian University, China

PC Members:

Jyh-Horng Wen, Tunghai University, Taiwan
Baojiang Cui, Beijing University of Posts And Telecommunications, China
Zheli Liu, Nankai University, China
Tainhan Gao, National Pilot Software College, China
Yung-Fa Huang, Chaoyang University of Technology, Taiwan
Chia-Hsin Cheng, National Formosa University Yunlin County, Taiwan
Tzu-Liang Kung, Asia University, Taiwan
Shu-Hong Lee, Chienkuo Technology University, Taiwan
Ho-Lung Hung, Chienkuo Technology University, Taiwan
Gwo-Ruey Lee, Lung-Yuan Research Park, Taiwan
Chung-Wen Hung, National Yunlin University of Science & Technology University, Taiwan
Nadeem Javaid, COMSATS Institute of Information Technology, Pakistan
Ahmed Naseem Alvi, COMSATS Institute of Information Technology, Pakistan
Syed Hassan Ahmed, Kyungpook National University, Korea
Abdul Wahid, COMSATS Institute of Information Technology, Pakistan
Muhammad Azfar Yaqub, Kyungpook National University, Korea
Juan Fang, Intel Corporation, USA
Chensi Zhang, Xidian University, China
Xuewen Liao, Xian Jiaotong University, China
Xiangbin Yu, Nanjing University of Aeronautics, China

2. Ad-Hoc and Mesh Networks

Chairs:

Elis Kulla, Okayama University of Science, Japan
Dongkyun Kim, Kyungpook National University, Korea
Makototo Ikeda, Fukuoka Institute of Technology, Japan

PC Members:

Admir Barolli, Aleksander Moisiu University of Durres, Albania
Tetsuya Oda, Fukuoka Institute of Technology, Japan
Evjola Spaho, Polytechnic University of Tirana, Albania
Arjan Durresi, IUPUI, USA
Tomoya Enokido, Rishho University, Japan
Akio Koyama, Yamagata University, Japan
Keita Matsuo, Fukuoka Institute of Technology, Japan
Isaac Woungang, Ryerson University, Canada
Noriki Uchida, Fukuoka Institute of Technology, Japan
Mimoza Durresi, Europian University of Tirana, Albania
Fumiaki Sato, Toho University, Japan

3. Cloud and Service Computing

Chairs:

Hwamin Lee, Soonchunhyang University, Korea
Florin Pop, Polytechnic University of Bucharest, Romania
Yilei Wang, Shangdong University, China

PC Members:

Hwa-Min Lee, Soonchunhyang University, Korea
Dae-Won Lee, Seokyoung University, Korea
Jong-Hyuk Lee, Samsung Electronics, Korea
Sung-Ho Chin, LG Electronics, Korea
Ji-Su Park, Korea University, Korea
Jae-hwa Chung, Korea National Open University, Korea
Ciprian Dobre, Polytechnic University of Bucharest, Romania
Sergio L. Toral Marín, University of Seville, Spain
Nik Bessis, Edge Hill University, UK
Makoto Ikeda, Fukuoka Institute of Technology, Japan

Fatos Xhafa, Technical University of Catalonia, Spain
Hao Wang, Shandong Normal University, China
Chengyu Hu, Shandong University, China
Xiaomei Yu, Shandong Normal University, China
Xiangwei Zheng, Shandong Normal University, China
Zhenhua Chen, University of Science and Technology, China

4. Multimedia and Web Applications

Chairs:

Chul Sur, Pusan University of Foreign Studies, Korea
Tomoyuki Ishida, Ibaraki University, Japan
Kenzi Watanabe, Hiroshima University, Japan

PC Members:

Jung Soo Rhee, Busan University of Foreign Studies, Korea
Sang Uk Shin, Pukyong National University, Korea
Youngho Park, Pukyong National University, Korea
Tetsuro Ogi, Keio University, Japan
Hideo Miyachi, Tokyo City University, Japan
Noriki Uchida, Fukuoka Institute of Technology, Japan
Yasuo Ebara, Osaka University, Japan
Nobuyuki Kukimoto, Kyoto University, Japan
Kaoru Sugita, Fukuoka Institute of Technology, Japan
Noriyasu Yamamoto, Fukuoka Institute of Technology, Japan
Yoshiaki Hori, Saga University, Japan
Takashi Yamanoue, Fukuyama University, Japan

5. Security and Privacy

Chairs:

Changhoon Lee, Seoul University of Science and Technology, Korea
Ryuya Uda, Tokyo University of Technology, Japan
Baojiang Cui, Beijing University of Posts and Telecommunications, China

PC Members:

Sang-Soo Yeo, Mokwon University, Korea
Soocheol Kim, Chung-Ang University, Korea
Kihong Park, Mokwon University, Korea
Sanghyun Seo, ETRI, Korea
Jongsung Kim, Kookmin University, Korea

Hangbae Chang, Chung-Ang University, Korea
Nobutaka Kawaguchi, Hitachi, Ltd., Japan
Masayuki Terada, NTT DOCOMO, Inc., Japan
Yoshihiro Kita, Tokyo University of Technology, Japan
Jianxin Wang, Beijing Forestry University, China
Jie Cheng, Shandong University, China
Shaoyin Cheng, University of Science and Technology of China, China
Jingling Zhao, Beijing University of Posts and Telecommunications, China

6. Network Protocols and Performance Analysis

Chairs:

Hyobeom Ahn, Kongju University, Korea
Francesco Palmieri, Second University of Naples, Italy
Akio Koyama, Yamagata University, Japan

PC Members:

Taekyoung Kwon, Yonsei University, Korea
Suyeon Lee, Baeseok Culture University, Korea
Youngwany Lee, Fareast University, Korea
Minoru Uehara, Toyo University, Japan
Fumiaki Sato, Toho University, Japan
Tomoyuki Nagase, Hirosaki University, japan
Tomoya Enokido, Risho University, Japan
Aniello Castiglione, University of Salerno, Italy
Massimo Ficco, Second University of Naples, Italy
Alessio Merlo, University of Genoa, Italy
Mauro Migliardi, University of Padova, Italy

7. Intelligent Computing

Chairs:

Jiwon Yoon, Korea University, Korea
Tomasz Hachaj, Pedagogical University of Cracow, Poland
Tetsuya Oda, Fukuoka Institute of Technology, Japan

PC Members:

Kangbin Yim, SCH University, Korea
Hiroaki Nishino, Oita University, Japan
Makoto Ikeda, Fukuoka Institute of Technology, Japan
Akio Koyama, Yamagata University, Japan

Takuo Suganuma, Tohoku University Japan
Salvatore Vitabile, University of Palermo, Italy
Katarzyna Koptyra, AGH University of Science and Technology, Poland
Adam Piórkowski, AGH University of Science and Technology, Poland
Paweł Hachaj, Cracow University of Technology, Poland
Marek Ogiela, AGH University of Science and Technology, Poland
Lidia Ogiela, AGH University of Science and Technology, Poland

8. Mobile and Vehicular Networks

Chairs:

Jeong Hyun Yi, Soongsil University, Korea
Bhed Bista, Iwate Prefectural University, Japan
Danda B. Rawat, Georgia Southern University, USA

PC Members:

Lei Chen, Georgia Southern University, USA
Gongjun Yan, University of Southern Indiana, USA
Houbing Song, West Virginia University, USA
Kayhan Zrar Ghafoor, Koya University, Iraq
Jiahong Wang, Iwate Prefectural University, Japan
Shigetomo Kimura, University of Tsukuba, Japan
Chotipat Pornavalai, King Mongkut's Institute of Technology Ladkrabang, Thailand
Evjola Spaho, Polytechnic University of Tirana, Albania
Wenjia Lei, New York Institute of Technology, USA
Chandra Bajracharya, Georgia Southern University, USA
Ghalib Asadullah, KICS UET Lahore, Pakistan
Yaser Jararweh, Jordan University of Science and Technology, Jordan

9. Distributed Algorithms and Systems

Chairs:

Hae-Duck Joshua Jeong, Korean Bible University, Korea
Tomoya Enokido, Rissho University, Japan

PC Members:

Jiyoung Lim, Korean Bible University, Korea
Jong-Suk Ruth Lee, KISTI, Korea
Francesco Palmieri, University of Salerno, Italy
Cuong Viet Dinh, Ho Chi Minh City University of Science, Vietnam

Hsing-Chung Jack Chen, Asia University, Taiwan
Woo-Seok Hyun, Korean Bible University, Korea
Gangman Yi, Gangneung-Wonju National University, Korea
Eric Pardede, La Trobe University, Australia
Vamsi Krishna Paruchuri, University of Central Arkansas, USA
Andrzej Wilczyński, Cracow University of Technology, Poland
Minoru Uehara, Toyo University, Japan
Akio Koyama, Yamagata University, Japan
Leonard Barolli, Fukuoka Institute of Technology, Japan
Fatos Xhafa, Technical University of Catalonia, Spain
Makoto Takizawa, Hosei University, Japan

10. Database and Data Mining

Chairs:

Seungmin Rho, Sungkyul University, Korea
Agustinus Borgy Waluyo, Monash University, Australia

PC Members:

Muhammad Sajjad, Islamia College Peshawar NWFP, Pakistan
Irfan Mehmood, Sejong University, Korea
Mucheol Kim, Sungkyul University, Korea
Sanghyun Seo, Sungkyul University, Korea
Yusuke Gotoh, Okayama University, Japan
Kin Fun Li, University of Victoria, Canada
David Taniar, Monash University, Australia
Wenny Rahayu, La Trobe University, Australia
Eric Pardede, La Trobe University, Australia
Tomoya Enokido, Rissho University, Japan

11. Ubiquitous and Pervasive Computing

Chairs:

Howon Kim, Pusan University, Korea
Ryo Nishide, Ritsumeikan University, Japan
Isaac Woungang, Ryerson University, Canada

PC Members:

Ian Piumarta, Ritsumeikan University, Japan
Kazuya Murao, Ritsumeikan University, Japan
Gregor Schiele, University of Duisburg-Essen, Germany

Taku Noguchi, Ritsumeikan University, Japan
Gaurav Indra, University of Delhi, India
Andrea Ceccarelli, University of Florence, Italy
Alagan Anpalagan, Ryerson University, Canada
Wei Lu, Keene State College, USA
Sanjay K. Dhurandher, University of Delhi, India
Neelanjana Dutta, Missouri University of Science and Technology, USA
Luca Caviglione, CNIT, Italy
Sriram Chellappan, Missouri University of Science and Technology, USA
Leandro Buss Becker, Universidadae Federal de Santa Catarina, Brazil
Glaucio Carvalho, Ryerson University, Canada
Deepak Sharma, University of Delhi, India

12. IoT, Sensor and Body Networks

Chairs:

Yang-Sun Lee, Mokwon University, Korea
Nik Bessis, Edge Hill University, UK
Zahoor Ali Khan, Higher Colleges of Technology, UAE

PC Members:

Jae-Myung Choi, Mokwon Unviersity, Korea
Mucheol Kim, Sungkyul University, Korea
Sang-Hyun Seo, Sungkyul University, Korea
Sang Oh Park, KISTI, Korea
Taeshik Shon, Ajou University, Korea
Woong Cho, Jungwon University, Korea
Jongsung Kim, Kookmin University, Korea
Jaehak Yu, ETRI, Korea
Eleana Asimakopoulou, Hellenic National Defence College, Greece
Marcello Trovati, University of Derby, UK
Bill Karakostas, VLTN, Belgium
Kevin Curran, Ulster University, UK
Federico Barrero, University of Seville, Spain
Nadeem Javaid, COMSATS IIT, Pakistan
Chaudhary Muhammad Imran, King Saud University, Saudi Arabia
Umar Qasim, University of Alberta, Canada
Farrukh Khan, King Saud University, Saudi Arabia
Hamed Aly, Acadia University, Canada

BWCCA-2016 Reviewers

Ahn Hyobeom
Ali Khan Zahoor
Barolli Admir
Barolli Leonard
Bessis Nik
Bista Bhed
Bouk Safdar Hussain
Caballé Santi
Castiglione Aniello
Chellappan Sriram
Chen Hsing-Chung
Chen Xiaofeng
Cui Baojiang
Di Martino Beniamino
Dobre Ciprian
Durresi Arjan
Enokido Tomoya
Ficco Massimo
Fiore Ugo
Fujioka Hiroyuki
Fun Li Kin
Gentile Antonio
Gotoh Yusuke
Hachaj Tomasz
Hussain Farookh
Hussain Omar
Javaid Nadeem
Jeong Joshua
Ikeda Makoto
Ishida Tomoyuki
Kikuchi Hiroaki
Kim Howon
Kolici Vladi

Koyama Akio
Kulla Elis
Lee Changhoon
Lee Hwamin
Lee Kyungroul
Lee Yang-Sun
Li Jing
Loia Vincenzo
Matsuo Keita
Kim Dongkyun
Koyama Akio
Kryvinska Natalia
Nishide Ryo
Nishino Hiroaki
Oda Tetsuya
Ogiela Lidia
Ogiela Marek
Palmieri Francesco
Paruchuri Vamsi Krishna
Pop Florin
Rahayu Wenny
Rawat Danda
Rho Seungmin
Shibata Yoshitaka
Sato Fumiaki
Spaho Evjola
Suganuma Takuo
Sugita Kaoru
Sur Chul
Takizawa Makoto
Taniar David
Terzo Olivier
Tokuyasu Tatsushi
Uchida Noriki
Uehara Minoru
Uda Ryuya
Venticinque Salvatore
Vitabile Salvatore
Waluyo Agustinus Borgy
Wang Xu An
Wang Yilei
Watanabe Kenzi
Woungang Isaac
Xhafa Fatos
Yim Kangbin
Yi Jeong Hyun
Yoon Jiwon

Welcome Message from MNSA-2016 International Symposium Co-Chairs

It is our great pleasure to welcome you to the 18-th International Symposium on Multimedia Network Systems and Applications (MNSA-2016), which will be held in conjunction with the 11-th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA-2016) at Soonchunhyang (SCH) University, Asan, Korea from November 5 to November 7, 2016.

This international symposium is a forum for sharing ideas and research work in the emerging areas of multimedia networking and systems.

Networks of today are going through a rapid evolution and the growing popularity of wired and wireless networks, multimedia network systems and applications are changing our daily life. In the last few years, we have observed an explosive growth of multimedia computing, communication and applications. This revolution is transforming the way people lives, works and interacts with each other, and is impacting the way business, education, entertainment, and health care are operating. Presently, a lot of research on high-speed networks and multimedia communication is going on. The papers included in this symposium cover aspects of IoT, multimedia applications, DTNs, network protocols, distributed computing systems and wireless networks.

Many people contributed to the success of MNSA-2016. First, we would like to thank the organizing committee of BWCCA-2016 International Conference for giving us the opportunity to organize the symposium. We would like to thank all authors for submitting their research work and for their participation. We are looking forward to meet them again in the forthcoming editions of the workshop.

We would like to express our appreciation to MNSA-2016 reviewers who carefully evaluated the submitted papers. Finally, we would like to thank the Local Arrangement Chairs for the local arrangement of the workshop.

We hope you will enjoy the workshop and have a great time in Asan, Korea.

MNSA-2016 International Symposium Organizing Committee

MNSA-2016 Symposium Organizers

Makoto Takizawa, Hosei University, Japan
Leonard Barolli, Fukuoka Institute of Technology, Japan

MNSA-2016 Program Co-Chairs

Tomoya Enokido, Rissho University, Japan

MNSA-2016 Organizing Committee

Symposium Co-Chairs

Makoto Takizwa, Hosei University, Japan
Leonard Barolli, Fukuoka Institute of Technology, Japan

Symposium PC Chair

Tomoya Enokido, Rissho University, Japan

Program Committee Members

Testuya Shigeyasu, Prefectural University of Hiroshima, Japan
Shintaro Imai, Iwate Prefectural University, Japan
Takuya Yoshihiro, Wakayama University, Japan
Motoi Yamagiwa, University of Yamanashi, Japan
Kazunori Ueda, Kochi University of Technology, Japan
Markus Aleksy, ABB AG, Germany
Irfan Awan, University of Bradford, UK
Bhed Bahadur Bista, Iwate Prefectural University, Japan
Yusuke Gotoh, Okayama University, Japan
Hui-Huang Hsu, Tamkang University, Taiwan
Rei Itsuki, Hiroshima International University, Japan
Satoru Izumi, Tohoku University, Japan
Akio Koyama, Yamagata University, Japan
Tomotaka Kozuki, Hiroshima International University, Japan
Toshiaki Osada, Tohoku Bunka Gakuen University, Japan
Fumiaki Sato, Toho University, Japan
Takuo Suganuma, Tohoku University, Japan
Hideyuki Takahashi, Tohoku University, Japan
Atsushi Takeda, Tohoku Gakuin University, Japan
Noriki Uchida, Fukuoka Institute of Technology, Japan
Misako Urakami, Oshima National College of Maritime Technology, Japan
Masaaki Yamanaka, Hiroshima International University, Japan
Muhammad Younas, Oxford Brookes University, UK
Fatos Xhafa, Technical University of Catalonia, Spain

Welcome Message from NGWMN-2016 International Workshop Co-Chairs

Welcome to the 9-th International Workshop on Next Generation of Wireless and Mobile Networks (NGWMN-2016), which will be held in conjunction with the 11-th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA 2016) at at Soonchunhyang (SCH) University, Asan, Korea from November 5 to November 7, 2016.

The aim of this workshop is to present the innovative researches, methods and algorithms for wireless networks, sensor networks and ubiquitous computing. The next generation of wireless and mobile networks is expected to allow a single mobile user to access heterogeneous wireless and mobile networks. Therefore, this workshop will provide a timely technical forum for the dissemination of new results in this exciting research area and is devoted to the architectures, protocols, resource management, mobility management, and scheduling in integrated wireless and mobile networks.

Many people have kindly helped us to prepare and organize the NGWMN-2016 workshop. First, we would like to thank the authors who submitted high quality papers and reviewers who carefully evaluated the submitted papers. We would like to give our special thanks to General Co-Chairs of BWCCA-2016 for their strong encouragement and guidance to organize this workshop. We would like to thank all of the PC members for their serious review works in order to make successful organization of NGWMN-2016.

Finally, we would like to thanks the Local Organizing Committee of BWCCA-2016 for excellent arrangement.

We hope you will enjoy the conference and have a great time in Asan, Korea

NGWMN-2016 Co-Chairs

Leonard Barolli, Fukuoka Institute of Technology, Japan
Hsing-Chung Chen (Jack Chen), Asia University, Taiwan
Kangbin Yim, SCH University, Korea

NGWMN-2016 Organizing Committee

Workshop Co-Chairs

Leonard Barolli, Fukuoka Institute of Technology, Japan
Hsing-Chung Chen (Jack Chen), Asia University, Taiwan
Kangbin Yim, SCH University, Korea

Program Committee Members

Muhammad Younas, Oxford Brookes University, UK
Awan Irfan, University of Bradford, UK
Makoto Ikeda, Fukuoka Institute of Technology, Japan
David Taniar, Monash University, Australia
Kin Fun Li, University of Victoria, Canada
Fatos Xhafa, Technical University of Catolonia, Spain
Vamsi Paruchuri, University of Central Arkansas, USA
Neng-Yih Shih, Asia University, Taiwan
Yeong-Chin Chen, Asia University, Taiwan
Akio Koyama, Yamagata University, Japan
Ming-Shiang Huang, Asia University, Taiwan
Isaac Woungang, Ryerson University, Canada
Arjan Durresi, Indiana University Purdue University Indianapolis, USA
Jyh-Horng Wen, Tunghai University, Taiwan
Cheng-Ying Yang, Department of Computer Science, University of Taipei, Taiwan
Tzu-Liang Kung, Asia University, Taiwan
Yung-Fa Huang, Chaoyang University of Technology, Taiwan
Chia-Hsin Cheng, National Formosa University Yunlin County, Taiwan
Neng-Yih Shih, Asia University, Taiwan
Jyu-Wei Wang, Asia University, Taiwan

Message from MAPWC-2016 International Workshop Organizers

Welcome to the 7-th International Workshop on Methods, Analysis and Protocols for Wireless Communication (MAPWC-2016), which will be in conjunction with the 11-th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA-2016) at Soonchunhyang (SCH) University, Asan, Korea from November 5 to November 7, 2016.

Wireless communications are characterized by high bit-error rates and burst errors, which arise due to interference fading, shadowing, terminal mobility, and so on. Since the traditional design of the algorithms, methods and protocols of the wired Internet did not take wireless networks into account, the performance over wireless networks is largely degraded. Especially, the multi-hop communication aggravates the problem of wireless communication even further. To solve these problems, there has been increased interest to propose and design new algorithms and methodologies for wireless communication.

The aim of this workshop is to present the innovative researches, methods and numerical analysis for wireless communications and wireless networks. The workshop contains high quality research papers, which were selected carefully by Program Committee Members.

It is impossible to organize such a successful program without the help of many individuals. We would like to express our appreciation to the authors of the submitted papers, and to the program committee members, who provided timely and significant review.

We hope all of you will enjoy MAPWC-2016 and find this a productive opportunity to exchange ideas with many researchers.

MAPWC-2016 International Workshop Organizers

MAPWC-2016 Workshop Chair
Leonard Barolli, Fukuoka Institute of Technology, Japan

MAPWC-2016 Workshop PC Co-Chairs
Makoto Ikeda, Fukuoka Institute of Technology, Japan
Hiroshi Maeda, Fukuoka Institute of Technology, Japan

MAPWC-2016 Organizing Committee

Workshop Chair

Leonard Barolli, Fukuoka Institute of Technology, Japan

Workshop PC Chair

Makoto Ikeda, Fukuoka Institute of Technology, Japan
Hiroshi Maeda, Fukuoka Institute of Technology, Japan

Program Committee Members

Arjan Durresi, Indiana University Purdue University Indianapolis (IUPUI), USA
Koki Watanabe, Fukuoka Institute of Technology, Japan
Shinichi Ichitsubo, Kyushu Institute of Technology, Japan
Zhi Qi Meng, Fukuoka University, Japan
Irfan Awan, Bradford University, UK
Tsuyoshi Matsuoka, Kyushu Sangyo University, Japan
Fatos Xhafa, Technical University of Catalonia, Spain
Kiyotaka Fujisaki, Fukuoka Institute of Technology, Japan

Web Administrator Co-Chairs

Tetsuya Oda, Fukuoka Institute of Technology, Japan
Shinji Sakamoto, Fukuoka Institute of Technology, Japan

Message from CWECS-2016 International Workshop Organizers

Welcome to Asan, Korea and the 7-th International Workshop on Cloud, Wireless and e-Commerce Security (CWECS-2016), which is held in conjunction with the 11-th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA-2016) at Soonchunhyang (SCH) University, Asan, Korea from November 5 to November 7, 2016.

Computer network and communication have been a part of our everyday life. People use them to contact others almost anytime anywhere. However, hackers due to business benefits, enjoying their skill/professional achievement or some other reasons very often attack, intrude or penetrate our systems. This is the key reason why computer/network security has been one of the important issues in computer research. Many researchers have tried to do their best to develop system security techniques and the methods to protect a system. But system attacks still occur worldwide every day. In fact, current system security technology is far away from prefect and should be continuously improved.

This workshop aims to present the innovative researches, methods and applications for cloud, wireless and e-commerce security. Other network related papers are also welcomed. The workshop contains high quality research papers, which were selected carefully by Program Committee Members.

It is impossible to organize such a successful program without the help of many individuals. We would like to express our appreciation to the authors of the submitted papers, and to the program committee members, who provided timely and significant reviews.

We hope all of you will enjoy CWECS-2016 and find this a productive opportunity to exchange ideas with many researchers.

CWECS-2016 International Workshop Organizers

CWECS-2016 Workshop Co-Chairs

Fang-Yie Leu , Tunghai University, Taiwan
Aniello Castiglione, University of Salerno, Italy
Chu-Hsing Lin, Tunghai University, Taiwan

CWECS-2016 Workshop PC Co-Chairs

Ilsun You, Korean Bible University, Korea
Fuu-Cheng Jiang, Tunghai University, Taiwan
Yi-Li Huang, Tunghai University, Taiwan

CWECS -2016 Organizing Committee

Workshop Co-Chairs

Fang-Yie Leu , Tunghai University, Taiwan
Aniello Castiglione, University of Salerno, Italy
Chu-Hsing Lin, Tunghai University, Taiwan

Workshop PC Co-Chairs

Ilsun You, Korean Bible University, Korea
Fuu-Cheng Jiang, Tunghai University, Taiwan
Yi-Li Huang, Tunghai University, Taiwan

Program Committee Members

Alessandra Sala, University of California Santa Barbara, USA
Antonio Colella, Italian Army, Italy
Chin-Cheng Lien, Soochow University, Taiwan
Chin-Ling Chen, Chaoyang University of Technology, Taiwan
Chiu-Ching Tuan, National Taipei University of Technology, Taiwan
Claudio Soriente, Universitat Politecnica de Madrid, Spain
Francesco Palmieri, Second University of Naples, Italy
Fuw-Yi Yang, Chaoyang University of Technology, Taiwan
I-Long Lin, Central Police University, Taiwan
Jason Ernst, University of Guelph, Canada
Jinn-Ke Jan, National Chung Hsing University, Taiwan
Lein Harn, University of Missouri Kansas City, USA
Sen-Tang Lai, Shih Chien University, Taiwan
Sergio Ricciardi, Technical University of Catalonia, Spain
Shiuh-Jeng Wang, Central Police University, Taiwan
Ugo Fiore, University of Naples, Italy
Heru Susanto, University of Brunei, Brunei

Message from RI3C-2016 International Workshop Organizers

Welcome to the 5-th International Workshop on Robot Interaction, Control, Communication and Cooperation (RI3C-2016), which will be held in conjunction with the 11-th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA 2016) at Soonchunhyang (SCH) University, Asan, Korea from November 5 to November 7, 2016.

Robots are being steadily introduced into modern everyday life and are expected to play a key role in the near future. Typically, the robots are deployed in situations where it is too dangerous, expensive, tedious, and complex for humans to operate. Although many of the real-life applications may only need a single robot, a large number of them require the cooperation, coordination and communication of a team of robots to accomplish a certain task. The use of multiple robots of overlapping capabilities offers a cost-effective and more robust solution. This redundancy in the robots' capabilities makes the overall system more flexible and fault-tolerant.

This workshop focuses on the emerging field of robot interaction, communication and cooperation bringing together research and application of methodology from robotics, human factors, human-computer interaction, interaction design, cognitive psychology, education and other fields to enable robots to have more natural and more rewarding interactions, communication and cooperation with humans throughout their spheres of functioning.

The design of an efficient collaborative multi-robot framework that ensures the autonomy and the individual requirements of the involved robots is a very challenging task. Developing operational multi-robot teams involves research on a number of topics such as fault tolerant cooperative control, adaptive action selection, distributed control, robot awareness of team member actions, improving efficiency through learning, inter-robot communication, action recognition, local versus global control, and metrics for measuring the success.

The aim of this workshop is to present the innovative researches, technologies and new concepts, services and application software of robotic systems.

The organization of an International Workshop needs the help of many people. We would like to express our appreciation to the authors of the submitted papers, and to the program committee members.

We hope all of you will enjoy RI3C-2016 program and your stay in Asan, Korea.

RI3C-2016 Workshop Organizers

RI3C-2016 Workshop Chair

Leonard Barolli, Fukuoka Institute of Technology, Japan

RI3C-2016 Workshop PC-Chairs

Keita Matsuo, Fukuoka Institute of Technology, Japan
Hiroyuki Fujioka, Fukuoka Institute of Technology, Japan

RI3C-2016 Organizing Committee

Workshop chair

Leonard Barolli, Fukuoka Institute of Technology, Japan

Workshop PC-Chairs

Keita Matsuo, Fukuoka Institute of Technology, Japan
Hiroyuki Fujioka, Fukuoka Institute of Technology, Japan

Program Committee Members

Tatsushi Tokuyasu, Fukuoka Institute of Technology, Japan
Akio Koyama, Yamagata University, Japan
Kaoru Fujioka, Fukuoka Women's University, Japan
Tetsuya Morizono, Fukuoka Institute of Technology, Japan
Junpei Arai, Yamagata College of Industry and Technology, Japan
Arjan Durresi, Indiana University Purdue University at Indianapolis (IUPUI), USA
Fatos Xhafa, Catalonia Technical University, Spain
Vladi Kolici, Polytechnic University of Tirana, Albania

Web Administrator Co-Chairs

Tetsuya Oda, Fukuoka Institute of Technology, Japan
Shinji Sakamoto, Fukuoka Institute of Technology, Japan

Message from SCC-2016 International Workshop Organizers

Welcome to the 3-rd International Workshop on Secure Cloud Computing (SCC-2016) which will be in conjunction with the 11-th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA-2016) at Soonchunhyang (SCH) University, Asan, Korea from November 5 to November 7, 2016.

As cloud computing becomes prevalent, more and more organizations outsource the expensive computing and storage into the cloud servers. It brings appealing benefits including relief of the burden for storage management, universal data access with independent geographical locations, and avoidance of capital expenditure on hardware, software, and personnel maintenances, etc. Despite the tremendous benefits, outsource storage inevitably suffers from some new security challenges, such as security and privacy of outsourced data. To address these issues, there has been increased interest to propose and design new algorithms and methodologies for secure cloud computing.

This workshop covers the latest advances in securing cloud storage and cloud computing that lead to gain competitive advantages in business and academia scenarios. Industry and academic researchers, professionals and practitioners are invited to exchange their experiences and present their ideas in this field. The workshop contains high quality research papers, which were selected carefully by Program Committee Members. The main topics of interest of SCC-2016 include but are not limited to the following:

- Security infrastructure and framework of cloud computing
- Coding and cryptography for secure cloud
- Remote data integrity and possession
- Distributed computation and access control on encrypted data
- Privacy preserving technologies in cloud computing
- Security and privacy in outsourcing data and computation
- Dependability, availability and forensics in cloud
- Secure data sharing, secure data replication
- Security in Cloud and Grid Systems

It is impossible to organize such a successful program without the help of many individuals. We would like to express our appreciation to the authors of the submitted papers, and to the program committee members, who provided timely and significant review.

We hope all of you will enjoy SCC-2016 and find this a productive opportunity to exchange ideas with many researchers.

SCC-2016 International Workshop Organizers

SCC-2016 Workshop Chair
Xiaofeng Chen, Xidian University, China

SCC-2016 Workshop PC Chair
Jin Li, Guangzhou University, China

SCC-2016 Organizing Committee

Workshop Chair

Xiaofeng Chen, Xidian University, China

Workshop PC Chair

Jin Li, Guangzhou University, China

Program Committee Members

Fangguo Zhang, Sun Yat-sen University, China
Xinyi Huang, Fujian Normal University, China
Jianwei Liu, Beihang University, China
Zhenjie Huang, Zhangzhou City University, China
Joseph K. Liu, Institute for Infocomm Research, Singapore
Yong Yu, University of Wollongong, Australia

Web Administrator Co-Chairs

Tetsuya Oda, Fukuoka Institute of Technology, Japan
Shinji Sakamoto, Fukuoka Institute of Technology, Japan

Message from LSNA-2016 International Workshop Organizers

Welcome to the 3-rd International Workshop on Large Scale Networks and Applications (LSNA-2016), which will be held in conjunction with the 11-th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA 2016) at Soonchunhyang (SCH) University, Asan, Korea from November 5 to November 7, 2016.

There are many network applications in various areas of Human life, thus there are many security threats by various attacks such as viruses and information interception. The network applications have attracted continuous attentions from both academia and industry. Especially, the large scale networks (e. g. social networks and wireless sensor network) aggravate even futher the problem of cyber attack. To address these issues of large scale networks, there has been increased interest to propose and design new algorithms and methodologies for network security and applications.

The aim of this workshop is to serve as a forum to present current and future work as well as to exchange research ideas in this field. The workshop invites authors to submit their original and unpublished work that demonstrate current research in all areas of large scale networks and applications.

It is impossible to organize such a successful program without the help of many individuals. We would like to express our appreciation to the authors of the submitted papers, and to the program committee members, who provided timely and significant review.

We hope all of you will enjoy LSNA-2016 and find this a productive opportunity to exchange ideas with many researchers.

LSNA-2016 International Workshop Organizers

LSNA-2016 Workshop Chair
Xiaofeng Chen, Xidian University, China

LSNA-2016 Workshop PC Chair
Debiao He, Wuhan University, China

LSNA-2016 Organizing Committee

Workshop Chair

Xiaofeng Chen, Xidian University, China

Workshop PC Chair

Debiao He, Wuhan University, China

Program Committee Members

Duncan Wong, City University of Hong Kong, China
Xinyi Huang, Fujian Normal University, China
Jingwei Liu, Xidian University, China
Joseph K. Liu, Institute for Infocomm Research, Singapore
Zheli Liu, Nankai University, China
Patrick P. C. Lee, The chinese University of Hong Kong
Yong Yu, University of Wollongong, Australia

Web Administrator Co-Chairs

Tetsuya Oda, Fukuoka Institute of Technology, Japan
Shinji Sakamoto, Fukuoka Institute of Technology, Japan

BWCCA-2016 Keynote Talk

Prof. Nobuo Funabiki, Okayama University, Japan

Java Programming Learning Assistant System: JPLAS

Abstract

As a useful and practical object-oriented programming language, Java has been used in many practical systems including enterprise servers, smart phones, and embedded systems, due to its high safety and portability. Then, a lot of educational institutes have offered Java programming courses to foster Java engineers. We have proposed and implemented a Web-based Java Programming Learning Assistant System called JPLAS, to assist such Java programming educations. JPLAS supports three types of problems that have different difficulties to cover a variety of students: 1) element fill-in-blank problem, 2) statement fill-in-blank problem, and 3) code writing problem. For 1), we have proposed a graph-theory based algorithm to automatically generate element fill-in-blank problems that have unique correct answers. For 2) and 3), we have adopted the test-driven development (TDD) method so that the answer codes from students can be automatically verified using test codes for their self-studies. In this talk, we introduce outlines of JPLAS and its application results to the Java programming course in our department. Besides, we introduce some new features of JPLAS including the offline answering function and the coding rule learning function.

Contents

Part I 11th International Conference on Broad-Band Wireless Computing, Communication and Applications (BWCCA-2016)

Part II The 18-th International Symposium on Multimedia Network Systems and Applications (MNSA-2016)

Part III The 9-th International Workshop on Next Generation of Wireless and Mobile Networks (NGWMN-2016)

Part IV The 7-th International Workshop on Methods, Analysis and Protocols for Wireless Communication (MAPWC-2016)

Part V The 7-th International Workshop on Cloud, Wireless and e-Commerce Security (CWECS-2016)

Part VI The 5-th International Workshop on Robot Interaction, Control, Communication and Cooperation (RI3C-2016)

Part VII The 3-rd International Workshop on Secure Cloud Computing (SCC-2016)

Part VIII The 3-rd International Workshop on Large Scale Networks and Applications (LSNA-2016)

Part I
11th International Conference on Broad-Band Wireless Computing, Communication and Applications (BWCCA-2016)

C++ Memory Check tool based on Dynamic Binary Instrumentation Platform

Jing Ling Zhao[1], Lei He[1,3], Bing He[2]

[1]School of Computer Science, Beijing University of Posts and Telecommunications, Beijing

zhaojingling@bupt.edu.cn

[2]State Grid Electric Power Company of Sichuan province Tianfu power supply company, Chengdu

feuereis@qq.com

[3]National Engineering Laboratory for Mobile Network Security, Beijing

lhestz@163.com

Abstract. In software development, to detect the presence of defects in the software as soon as possible, would greatly reduce the extent of losses arising. In this paper, focus on the memory-use error in C++ program, designed and implemented a memory check tools named ShadowCheck, based on dynamic binary instrumentation platform, which is platform-cross, efficiency and accuracy. In this paper, introduced dynamic binary instrumentation platform and the memory layout of Linux first, then explained how the ShadowCheck works, at last, summarized the efficiency and accuracy of ShadowCheck.

1. Introduction

With the continuous development of computer technology, application program has extensive deep into all walks of life, as one of the most widely used programming language, C++ was designed with a bias toward system programming and embedded, resource- constrained and large systems, with performance, efficiency and flexibility of use as its design highlights.

Due to the limitations of its own software engineering, as well as the flexibility of C++ itself, so the program will inevitably lead to a variety of defects and bugs, these bugs mostly caused by the incorrect use of memory, such as buffer overflows, memory leaks, read undefined objects, and so on. These kind of bugs often causes undefined behavior and do not immediately make the program error or crash. For instance, the buffer overflow bug, only when the contaminated data is used, the program will show abnormal. Due to the delay makes such kinds of errors are difficult to find and repair.

Under normal circumstances, simply rely on manual code review to Troubleshoot such errors is very inefficient, and further, there is a high false positive rate while checking for software defects with static analysis [1] tools, and some software companies in order to protect their own use, and will not disclose the source code, which makes the source code analysis to check for bugs more difficult.

Compared with static analysis, binary code based dynamic defect detection [2] has a higher accuracy rate. And with binary code oriented software defect detection, we can dynamically change the binary executable code of program and the value of register, this brought great convenience for defect detection.

On the other hand, binary executable code oriented defect detection tools are more difficulty to implement, and the code is always platform-specific and difficulty to port [3]. Furthermore, direct manipulation of the binary executable code makes the code optimization becomes more difficult, so that the efficiency of the program dropped significantly.

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_1

In order to solve the above problems, it has developed DBI (Dynamic Binary Instrumentation) platform [4] that allows us to write robust portable code that is binary executable oriented at a higher level of abstraction.

2. DynamoRIO Introduction

DynamoRIO is a system for runtime code manipulation that is efficient, transparent, and comprehensive, able to observe and manipulate every executed instruction in an unmodified application running on a stock operating system and commodity hardware

DynamoRIO uses a callback mechanism to operate with binary executable code, by setting a callback on the events we interest, we can easily manipulate the binary executable code of the target program. Meanwhile DynamoRIO instruction is a cross-platform binary instrumentation platform, which allows us to write programs more portable.

2.1 DynamoRIO System Details

DynamoRIO operates by shifting an application's execution from its original instructions to a code cache, where the instructions can be freely modified.

DynamoRIO occupies the address space with the application and has full control over execution, taking over whenever control leaves the code cache or when the operating system directly transfers control to the application (kernel-mediated control transfers).

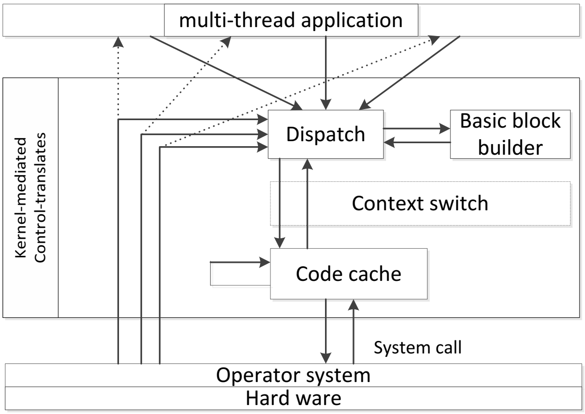


Fig.1. DynamoRI0 system detail

DynamoRIO copies the application code one dynamic basic block at a time into its basic block code cache. A block that directly targets another block already resident in the cache is linked to that block to avoid the cost of returning to the DynamoRIO dispatcher.

Frequently executed sequences of basic blocks are combined into traces, which are placed in a separate code cache. DynamoRIO makes these traces available via its interface for convenient access to hot application code streams.

The following figure shows the flow of control between the components of DynamoRIO and its code caches.

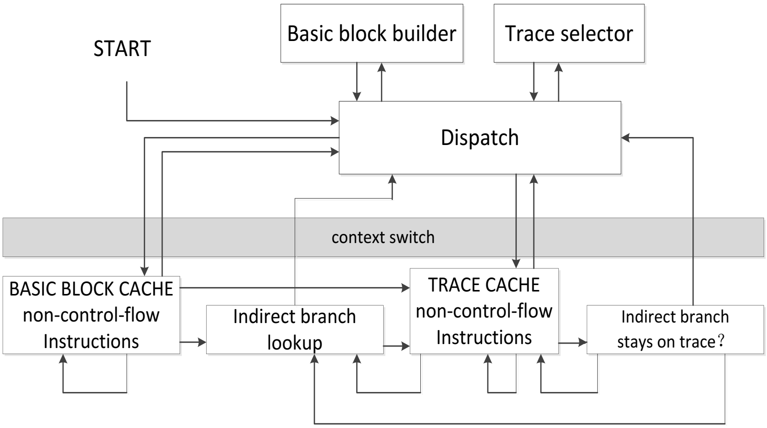


Fig.2. Control flow of DynamoRIO

2.2 The efficiency of DynamoRIO

DynamoRIO operates in user mode on a target process. It acts as a process virtual machine, interposing between the application and the operating system.

Generally speaking, interpreting the binary would be about 300 times slower than executing the binary code directly on the hardware. DynamoRIO use code cache, direct link and indirect branch significantly reduce the cost of interpreting.

2.2.1 Code cache

With using the instruction cache, the code which may execute frequently would be stored in the cache, and being execute as same as the native code. The Code cache mechanism significantly improved the execution speed. Experiments have shown that the time overhead costs by interpret could be reduced to about 25 times by code cache.

2.2.2 Direct link

By copying the basic block to code to the code cache and execute them as same as native code can improve a lot of executing efficiency, but we still need to interpret the jump instruction because while control flow jumps to some address, no one knows whether the jump address is in the same basic code block and whether the new basic code block is cached. So we may have to switch the context to DynamoRIO and query where to jump and it is necessary to cache the new basic code block.

DynamoRIO uses direct-link technique to link the basic code block, so if control flow jumps to a cached basic code block from another cache basic block by jump instruction, do not need to switch context anymore.

So, if the both basic code blocks are cached, the runtime cost could be reduced to about 3 times.

2.2.3 Indirect branch

Unlike jump instruction, the conditional branch instruction can't determine where to jump while basic code block was cached. DynamoRIO sets an indirect branch look up block to find where to jump. Compared with execute the code with conditional branch instruction directly, execute the same code through DynamoRIO only takes 1.2 times runtime.

2.3 Transparent

DynamoRIO should not affect the client process itself, but for the DynamoRIO and the client process are running in the user space together, it is always difficulty to be fully transparent.

For transparency, DynamoRIO done a lot of work.

- To avoid library conflict DynamoRIO do not use any link library, it use system call directly instead.
- DynamoRIO use its own heap memory.
- DynamoRIO use its own input and output program to avoid conflict. And write the output to file instead writing it to standard output.
- DynamoRIO do not create thread, it just depends on the thread created by client process and user its own thread stack.

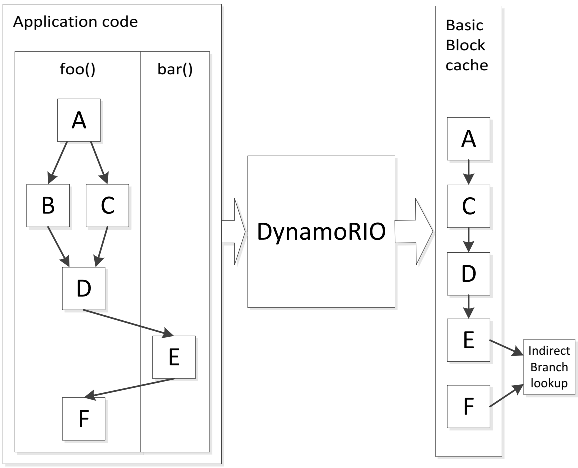


Fig.3. Code cache

3. Memory Check Tools for C++

C++ is a flexibility programming language,which is famous for its efficiency and performance. And the flexibility makes it very easy to make mistake while using C++, even for efficiency programmer. Most of these mistakes are caused by incorrect use of memory.

Traditional C++ detection tools, for example valgrind [5] and PIN [6], have false positive rate, low efficiency, poor portability shortcomings.

In order to help programmers, find such hidden bug, this paper design and implemented a cross-platform memory use check tool based DynamoRIO for C++ (ShadowCheck), which is efficient and accurate. The ShadowCheck uses shadow memory to recording the use of memory, and record the changes by the callback function.

3.1 Shadow Memory

ShadowCheck records the use of memory as shadow memory. In this paper, just like traditional tools, defined three states for every byte of memory; they are defined, undefined and unreachable [8].

Take Linux's virtual memory structure [7] for example, from lower to higher addresses are global variables and binary code text, heap memory, stack memory, the command-line arguments and environment variables.

For initialization, we defined the memory upper than stack is always unreachable, and would never modify it. For heap memory, we define the whole heap memory unreachable in the initialization stage. While the client program running, we could change the range of stack and heap, just like show in figure x. The shadow memory can only be one of the three states, we have not defined read-only state, because in the view of binary executable code, we can't differentiate if a byte of memory located in the heap or stack is writable.

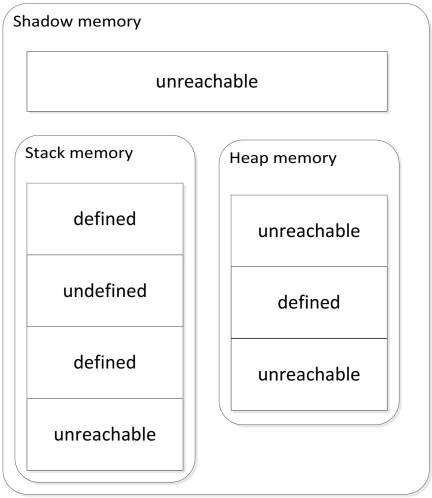


Fig.4. Linux memory layout

The ShadowCheck would check and change the state of shadow memory each time when every key instruction, such as malloc call, assignment instruction and address access instruction.

- When the stack grows, mark the corresponding as undefined.
- After each assign instruction executed, mark the corresponding as defined.
- While the function call asking for heap memory returned successful, mark the returned address as undefined. Then add a record to the global heap memory table.

- When the function releasing heap memory, mark the released memory as unreachable. Then erase the corresponding in the heap memory table.

But if we shadowed every byte of the entire memory space, would take away to many memories. ShadowCheck saved the shadow memory in three ways. If the byte is reachable, shadow each byte for one bit, which is enough to present the two reachable states, defined and undefined. For the unreachable memory, in this paper, it does not record the unreachable memory, instead, it keeps a hash map for every reachable byte. This has two advantages; First, ShadowCheck can find the reachable memory with high-efficiency. Second, if we can't find some address in the hash map, it is always unreachable, in spite of the access to that unreachable memory may not case segment fault, but that is still certainly invalid. The tool would record the call stack and stop the client process. At last, ShadowCheck keeps an address table to record the bound of memory block.

3.2 Error detection

How to detect the invalid memory use with shadow memory is the key point for ShadowCheck. Using an appropriate detection algorithm can reduce the false positive rate and improve efficiency. ShadowCheck can check the following types of errors.

3.2.1 Memory leak

Only the memory without any record can be mentioned as memory leak. Because after the client process exits, kernel would retrieve most of resource, include memory. So, if the client has keep any record of a block of memory, including the address in the middle.

3.2.2 Buffer overflow

While ShadowCheck keep the bound for every memory block, it is easy to check if an assign instruct write into a memory overflowed. This is useful to check buffer overflow, but when write to a struct or class, it is difficult to determine whether buffer overflow occurred. ShadowCheck records the assignment instruct which covers both defined and undefined memory as may buffer overflow.

3.2.3 Reading and writing to unreachable memory.

To determine whether an address is reachable is easy. Check whether the address in upper the stack or in the heap but can't be found in the heap memory table can check an address is reachable or not.

3.2.4 Reading the undefined memory

This problem is a little compose, because many operator system requires memory alignment, while the client process reply for memory or read memory, several more bytes maybe copied. ShadowCheck would not treat these cases as read undefined memory, except those two cases blow, the beginning of the address is undefined, the memory with undefined bytes is used as arguments for system call or conditional switch branch.

3.2.5 Library functions and system calls

Generally speaking, there are two ways to handle library functions and system calls, one is replacing them with own version, the other is wraps around the calls. To replace the library and system call cost too much for a cross-platform tool, ShadowCheck wraps those function calls with callbacks that check the state of shadow memory.

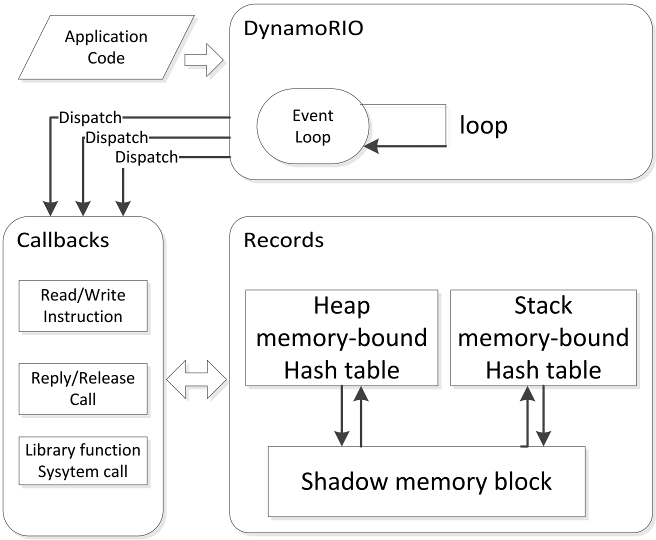


Fig.5. Framework of ShadowCheck

4. Efficiency and accuracy

4.1 Efficiency

For many binary insertion tool, the efficiency has been a difficult. Interpreting the binary executable code by software always slower hundreds of times than execute it directly with hardware.

ShadowCheck is built based on DynamoRIO, which is already very efficiency. Furthermore, we implemented ShadowCheck with a series of optimization methods.

- Use one byte to shadow four bytes, and encoded the related operating into binary manipulation.
- We hashed function name to integer data, increased the speed while manipulating call stack.
- Use extra hashtable to store the bound of important variables.
- In some case, use a pair of tag to record the use of memory, this would be fast a lot than shadow every byte.

We tested the efficiency and compared with the well-known tools, Dr. Memory.

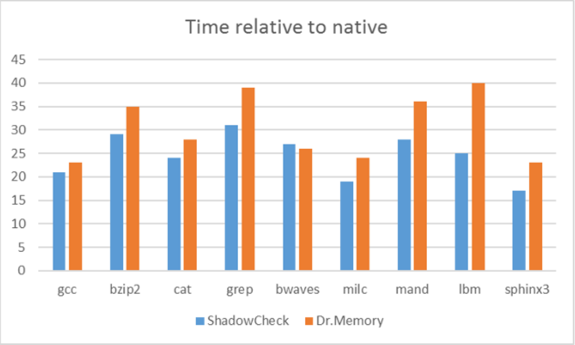


Fig.6. Efficiency compared with dr. memory

4.2 Accuracy

Dynamic test has a higher accuracy than static analysis. And the error judge algorithm used by ShadowCheck can obviously reduce false positives.

- For any memory access instruction, check the shadow memory whether is defined. If the access is used to system call or breach switch instruction, the check will be stricter.
- For each memory apply/release call, check the memory hashtable whether it is valid.
- For any instruction associated with pointer, record it and check whether memory leak occurred.

ShadowCheck divides the errors into three categories; clearly error, possible error, debug information. In most cases, error would be found in the clearly error.

5. Conclusion

This paper concentrate on memory use errors for C++ codes. Design and implemented a memory check tools based on binary dynamic instrumentation platform. By shadow the using of memory, can detect the memory use error for high accuracy.

6. Acknowledgement

This work is supported by State Grid Electric Power Company of Sichuan province Tianfu power supply company and National Engineering Laboratory for Mobile Network Security.

References

1. Ball T, Rajamani S K. The SLAM project: debugging system software via static analysis[C]// 2002:1-3.
2. Hurty W C. Dynamic analysis of structural systems using component modes[J]. Aiaa Journal, 2012, 3(4):678-685.
3. Turkboylari M. Implementation of a secure computing environment by using a secure bootloader, shadow memory, and protected memory: US, US 7313705 B2[P]. 2007.

4. Lyu Y H, Hong D Y, Wu T Y, et al. DBILL: An Efficient and Retargetable Dynamic Binary Instrumentation Framework using LLVM Backend[J]. Acm Sigplan Notices, 2014, 49(7):141-152.

5. Nethercote N, Seward J. Valgrind: a framework for heavyweight dynamic binary instrumentation[J]. Acm Sigplan Notices, 2015, 42(6):89-100.

6. Luk C K, Cohn R, Muth R, et al. 9 8 Pin: Building Customized Program Analysis Tools with Dynamic Instrumentation[C]// ACM Sigplan 2005 Conference on Programming Language Design and Implementation, Chicago, Il, Usa, June. 2005:190-200.

7. Pettersson T. Cryptographic key recovery from Linux memory dumps[C]// Chaos Communication Camp. 2007.

8. He Y, Shu H, Xiong X. Protocol Reverse Engineering Based on DynamoRIO[C]// International Conference on Information and Multimedia Technology. IEEE Computer Society, 2009:L1191-L1194.

A program behavior recognition algorithm based on assembly instruction sequence similarity

Baojiang Cui[1], Chong Wang[2] , GuoWei Dong[3], JinXin Ma[4]

[1] School of Computer Science, Beijing University of Posts and Telecommunications,
National Engineering Laboratory for Mobile Network Security, China
cui_bjiang@163.com

[2] School of Computer Science, Beijing University of Posts and Telecommunications,
National Engineering Laboratory for Mobile Network Security, China
wangchong756@126.com

[3] China Information Technology Security Evaluation Center, Beijing 100085,China
dgw2008@163.com

[4] China Information Technology Security Evaluation Center, Beijing 100085,China
majinxin2003@126.com

Abstract. The analysis on assembly instruction sequence plays a vital role in the field of measuring software similarity, malware recognition and software analysis, etc. This paper summarizes the features of assembly instructions, builds a six-group model and puts forward an algorithm of calculating similarity of assembly instructions. On that base a set of methods of calculating similarity of assembly instruction sequence are summarized. The preliminary experimental results show that it has high efficiency and good effect.

1 Introduction

In recent years, malicious software has a increasing spread, a dizzying variety and fast pace of change, which produce a great deal of trouble and loss for users. A report[1] in Business Software Alliance recently shows that 430 million new pieces of malware were discovered in 2015,up 36 percent from 2014 and organizations experiences some form of malware attack every seven minutes. Thus it can be seen, how to identify the behavior of the program has become an important research field of information security and plays a significant role in the program similarity and malware analysis field.

The traditional program identification technology can be divided into static recognition[2,3] and dynamic recognition[4,5]. Static recognition means using a disassembler to turn the executable program code into assembly language and discover certain acts by matching the sequence of bytes and extracting signatures and constant feature of the program algorithm. Static recognition has an

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_2

advantage of low overheads, but it can t recognize the use of multi-state, deformation, encryption, confusion and other means of malicious programs.

Dynamic recognition is in an isolated emulation environment to run suspicious files, scanning system calls and analyzing instruction steam and data available. Thereby it can substantially eliminate the effects of Packers and code obfuscation. The drawback is the high overhead, low efficiency and low accuracy.

This paper presents a program behavior recognition algorithm based on assembly instruction sequence similarity. In this work we record the assembly instruction steam, extract and analysis the information of instructions and generalise the model of assembly instruction sextuple. Based on this model, we design a matching algorithm for the assembly instruction steam sequence. Ultimately, we complete the induction which is from the assembly instructions to the abstract logic behavior of program from bottom to top, using dynamic program behavior feature extraction technology, to achieve the purpose of identifying the unknown program.

2 Structure of assembly instruction

Typically, the general format of assembly instructions are:

[label field:] opcode field [source operand, destination operand] [; comment]

Among them, the square brackets [] means that contents are optional, depending on the circumstances.

(1) Label field: It is located at the beginning of the statement and represents the address of this statement. Numeral itself consists of one to eight letters and numbers, representing the command position in memory.

(2) Opcode field: It represents the function of the instruction and indicates that the operation to be executed by a computer.

(3) Operand: Operand is the assembly instructions operating data or address that data is located. This part can be divided into operand itself, the address of operand or the information related to the operands.

(4) Comments: Comments begin with semicolon (;). It illustrates the program features to enhance readability.

3 Abstract coding of assembly instructions

Compared to static assembly instructions, the system function call sequences and the like, dynamic assembly instructions flow covers the entire information of the program behavior during the execution. It is an ideal object of analysis which we research to identify the program behavior. Through the above analysis shows that the structure of the assembly instructions, simple assembly instructions structures do not have abstraction. The semantic information that assembly instructions implied can not be identified and calculated by programs. Therefore, in order to extract the assembly instructions operational semantics, we design a abstract code to describe assembly instructions and name the model of assembly instruction sextuple In.

$$\mathrm{In} = < \mathrm{op}_{\mathrm{code}}, \mathrm{op}_{\mathrm{num}}, \mathrm{psw}, \mathrm{op}_{\mathrm{wr}}, \mathrm{deep}, \mathrm{time} > \qquad (3\text{-}1)$$

This model fully considers the abstract feature information between different instructions. After extracting the semantic component feature of each assembly instruction, we quantify each element in the model of assembly instruction sextuple, using quantitative value to express the information in each part.

1) operation code

$\mathrm{op}_{\mathrm{code}}$ includes the operation code itself and their type. First, the operation code can be divided into six kinds according to their semantics, including data transfer instructions(MOV, LEA), arithmetic instructions(ADD, SUB), bit operation instructions(AND, OR), string operation instructions (STOS, CMPS), jump control instructions(JA, CALL), advanced control instructions(CLC, BOUND). Each category can be divided into subcategories in accordance with the address of the operand and the specific operating behaviors in the table below.

Category	Subcategory
Data transfer	General data transfer operation、Stack data transfer operation、Data Exchange operation
Arithmetic	Add and sub、Multiplication and division、Extended operations
Bit operation	General bit operation、Bit test operation、Bit scanning operation、Shift operation
String operation	String transfer operation、String store operation、String compare operation、
Jump control	Function call、Unconditional jump、Conditional jump、Loop control

Table 1 Classification of operation code

In quantifying the operation code, if two operation codes are totally same, it will return 1 in operation coding function. If two operation codes belong in the same category, which means there are similar places in meaning , it will return 0.5 in operation coding function. If they are not in the same category, it will return 0.1.

2) operand coding

$\mathrm{op}_{\mathrm{num}}$ contains the number of operands of the instruction and three characteristics of the operand:

□ Type：Operand is a field of assembly instructions. There are three types of operand can be put in this field, which are immediate, memory and register. In addition, memory space is divided into stack and non-stack space.

□ Importance: If memory or register as an independent operand, will be marked as "important"; on the contrary, if it is as an integral part of the operand address, were marked as "not important." This similarity will be quantized in subsequent alignments.

□ Tags: Tags here represent some information related to operands, recording associated memory or register. For example, EAX is a 32-bit general-purpose registers, EAX will be marked as 32, and joined the AX, AH, AL in the label. In quantifying operands, first we check each read-write mode of operands are same or not. Important will be marked if two operands are totally same and the returned value will be 0.2. If they are not same, the returned value will be 0.1.Next we check their storage type, If they are same, the result will increase by 0.1, otherwise it does not affect the result.

3) Flag register coding

Flag registers are called the program status word. We can know your current state of the CPU by them. For example, OF represents the overflow flag, it is used to check the result from addition or substraction operation is overflowed or not. psw is a nine-tuple structure which describes the flag register set. It is written $< cf, pf, af, zf, sf, tf, if, df, of >$. It corresponds to nine flag registers, describing the impact operation code makes. If the operation code affected one register, the returned value increases by 1/9.

4) Read and write operations coding

$\mathrm{op_{wr}}$ is a triad $< w, r, t >$ which describes read-write mode in the assembly instruction, representing the operand of read , write and take . For example, the assembly instruction MOV EBX，DWORD PTR [EAX+0X4]. There are four operands in it. The read-write mode of the first one EAX is t. The read-write mode of the second one 0x4 is r. Then two operands comprise the third operand DWORD PTR [EAX+0X4], and its read-write mode is r. The read-write mode of the last one EBX is w.

5) Depth of nesting function coding

deep represents the function which the assembly instruction belongs to is invoked in deep layer. Even for the same instructions, the meanings of different nested depths are very different. Such as PUSH SP, the value in SP in functions of different nested depths differs greatly.

6) Timestamp coding

time indicates the exact point of execution. When $\mathrm{time} = 10$, it means that it is the 10th assembly instruction when the program executes.

4 Similarity measurement algorithm of assembly instructions

After quantifying each element in the sextuple model, we can deduce the calculating formula that describes the smilarity between two assembly instructions.

$$\mathrm{SimIn} = \mathrm{SimC} * (\mu_b * \mathrm{SimB} + \mu_s * \mathrm{SimS}) \qquad (4\text{-}1)$$

Among them, SimC represents the semantic similarity of instruction, referring to the similarity of $< \mathrm{op_{code}}, \mathrm{op_{num}} >$. SimB represents the behavior similarity of instruction, referring to the similarity of $< psw, \mathrm{op_{wr}} >$. SimS represents the structural similarity of instruction, referring to the similarity of $< deep, time >$. μ_b、μ_s represents the weights to SimB、SimS. Since $< \mathrm{op_{code}}, \mathrm{op_{num}} >$ plays an important role in the sextuple, SimC itself is a multiplication factor. These variables need to fulfill the below conditions:

$$\mu_b + \mu_s = 1，0 \le \mathrm{SimC} \le 1，0 \le \mathrm{SimB} \le 1，0 \le \mathrm{SimS} \le 1 \qquad (4\text{-}2)$$

The computational method above, we take into full account the semantic information, the behavior information and the structural information. It can be more accurate and comprehensive to display the similar situation between two instructions.

5 Similarity measurement algorithm of assembly instruction sequence

After having calculated the similarity between two assembly instructions , we start to calculate the similarity betweeen two assembly instruction sequences. Here we adopt the idea of the longest common sequence(LCS) problem which is a dynamic programming.

$\mathrm{InS}(1, \mathrm{n})$ represents a assembly instruction sequence that consists of n assembly instructions, and it is numbered from 1. So the similarity between two assembly instruction sequences $\mathrm{SimInS}\ (\mathrm{n_1}, \mathrm{n_2})$ is defined below:

$$\mathrm{SimInS}(\mathrm{i}, 0) = 0 \ \ (1 \le \mathrm{i} \le \mathrm{n_1}) \qquad (5\text{-}1)$$

$$\mathrm{SimInS}(0,\mathrm{j}) = 0 \ \ (1 \le \mathrm{j} \le \mathrm{n}_2) \qquad (5\text{-}2)$$

$$\mathrm{SimInS}(\mathrm{i},\mathrm{j}) = \max\begin{cases} \mathrm{SimInS}(\mathrm{i}-1,\mathrm{j}) \\ \mathrm{SimInS}(\mathrm{i},\mathrm{j}-1) \\ \mathrm{SimInS}(\mathrm{i}-1,\mathrm{j}-1) + \mathrm{SimIn}(\mathrm{i},\mathrm{j}) \end{cases} \quad (1 \le \mathrm{i} \le \mathrm{n}_1, 1 \le \mathrm{j} \le \mathrm{n}_2) \qquad (5\text{-}3)$$

Pseudo-code as follows:

```
Input:  Assembly instruction sequence  InS1;
        Assembly instruction sequence  InS2;
Output: The similarity of this two sequences;
function Sim(InS1,InS2)
  for i = 1 to end do
    for j = 1 to end do
      t=max(SimInS[i-1][j],SimInS[i][j-1])
      SimInS[i][j]=max(t,SimIns[i-1][j-1]+SimIn[i][j])
    end for
  end for
 return SimIns[n1][n2]
 end function
```

$\mathrm{SimInS}(\mathrm{i},\mathrm{j})$ represents the global similarity of two assembly instruction sequence and it has not been normalized. To normalize this result, we define $\mathrm{SimInS}(\mathrm{i1},\mathrm{i2},\mathrm{j1},\mathrm{j2})$ as the partial similarity of two assembly instruction sequence $\mathrm{InS}_{1(\mathrm{i}_1,\mathrm{i}_2)} =< I\mathrm{n}_{\mathrm{i1}}, \mathrm{In}_{\mathrm{i1+1}}, \mathrm{In}_{\mathrm{i1+2}} \ldots \mathrm{In}_{\mathrm{i2}} >$, $\mathrm{InS}_{2(\mathrm{j}_1,\mathrm{j}_2)} =< I\mathrm{n}_{\mathrm{j1}}, \mathrm{In}_{\mathrm{j1+1}}, \mathrm{In}_{\mathrm{j1+2}} \ldots \mathrm{In}_{\mathrm{j2}} >$.We get the following equation:

$$\mathrm{SimInS}(\mathrm{i1},\mathrm{i2},\mathrm{j1},\mathrm{j2}) = \mathrm{SimInS}(\mathrm{i2},\mathrm{j2}) - \mathrm{SimInS}(\mathrm{i1}-1,\mathrm{j1}-1) \qquad (5\text{-}4)$$

So the normalized similarity of two assembly instruction sequence are as follows:

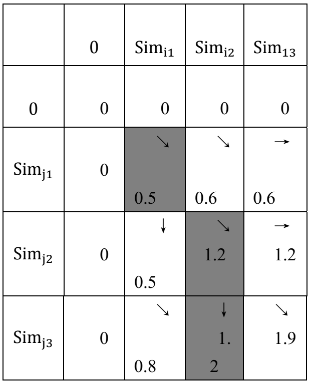
$$\mathrm{len}(\mathrm{i1},\mathrm{i2},\mathrm{j1},\mathrm{j2}) = \max\,(\mathrm{i2}-\mathrm{i1}+1, \mathrm{j2}-\mathrm{j1}+1) \qquad (5\text{-}5)$$

$$\mathrm{NorSimIns}(\mathrm{i1},\mathrm{i2},\mathrm{j1},\mathrm{j2}) = \frac{\mathrm{SimInS}(\mathrm{i1},\mathrm{i2},\mathrm{j1},\mathrm{j2})}{\mathrm{len}(\mathrm{i1},\mathrm{i2},\mathrm{j1},\mathrm{j2})} \qquad (5\text{-}6)$$

$\mathrm{len}(\mathrm{i1},\mathrm{i2},\mathrm{j1},\mathrm{j2})$ indicates the length of the instruction sequence interval. The assembler instruction sequence similarity normalized result $\mathrm{NorSimIns}(\mathrm{i1},\mathrm{i2},\mathrm{j1},\mathrm{j2})$ can be get by the partial similarity divide the interval length.

According to the above transfer equation,we can draw a sketch map describing the algorithm.

Each grid represents the similarity between two assembly instructions. The arrows represent the value of the source in the position. Choosing dark boxes indicate the path traversed when calculating ParSimInS (i1, i2, j1, j3).



	0	$\mathrm{Sim}_{\mathrm{i1}}$	$\mathrm{Sim}_{\mathrm{i2}}$	Sim_{13}
0	0	0	0	0
$\mathrm{Sim}_{\mathrm{j1}}$	0	↘ 0.5	↘ 0.6	→ 0.6
$\mathrm{Sim}_{\mathrm{j2}}$	0	↓ 0.5	↘ 1.2	→ 1.2
$\mathrm{Sim}_{\mathrm{j3}}$	0	↘ 0.8	↓ 1. 2	↘ 1.9

Figure 5-1 The algorithm sketch map

So we can get the result from the figure above:

$$\mathrm{ParSimInS}(i1,i2,j1,j3)=1.2,\ \mathrm{NorSimIns}(i1,i2,j1,j3)=\frac{\mathrm{ParSimInS}(i1,i2,j1,j3)}{\mathrm{len}(i1,i2,j1,j3)}=\frac{1.2}{3}=0.4$$

6 Optimum of algorithm

6.1 Reduce of time complexity

The current average time complexity of the algorithm is $\mathrm{O}(\mathrm{n}^2)$. n is the number of instructions in the program. In practical applications, a program contains a large number of assembly instructions, so now the time complexity of the algorithm is high, efficiency is low. So we find a Instruction - Basic Block mode to reduce the time complexity of the algorithm.

The basic block is a sequence of instructions executed in the program. During the running time, it maybe execute the same basic block. First we calculate the similarity between assembly instructions for the basic element in the static state. Then we calculate the similarity between basic blocks for the basic element, while establishing the similarity between the two-dimensional table of basic blocks. We treat $< index(\mathrm{BBL}_1), \mathrm{index}(\mathrm{BBL}_2) >$ as key value, thus the number between two basic blocks is only once. When we find these two basic blocks has been calculated, the query time is only O(1), eliminating many repetitive calculations. Time complexity has been reduced. In summary, the similarity of two basic blocks $\mathrm{BBLS}_1(1,\mathrm{n})$ and $\mathrm{BBLS}_2(1,\mathrm{n})$ may be expressed as:

$$\mathrm{SimBBL}(\mathrm{BBL}_1,\mathrm{BBL}_2)=\mathrm{SimInS}(i1,i2,j1,j2) \quad (6\text{-}1)$$

$\mathrm{BBLS}(1,\mathrm{n})$ represents that there are n basic block sequence $< BB\mathrm{L}_1, \mathrm{BBL}_2, \mathrm{BBL}_3 \ldots \mathrm{BBL}_\mathrm{n} >$, we can get the equation from above.

$$\mathrm{SimBBLS}(\mathrm{i},0)=0\ \ (1\le \mathrm{i}\le \mathrm{n}_1) \quad (6\text{-}2)$$

$$\mathrm{SimBBLS}(0,\mathrm{j})=0\ \ (1\le \mathrm{j}\le \mathrm{n}_2) \quad (6\text{-}3)$$

$$\mathrm{SimBBLS}(\mathrm{i},\mathrm{j})=\max\begin{cases}\mathrm{SimBBLS}(\mathrm{i}-1,\mathrm{j})\\ \mathrm{SimBBLS}(\mathrm{i},\mathrm{j}-1)\\ \mathrm{SimBBLS}(\mathrm{i}-1,\mathrm{j}-1)+\mathrm{SimBBL}(\mathrm{i},\mathrm{j})\end{cases} \quad (1\le \mathrm{i}\le \mathrm{n}_1, 1\le \mathrm{j}\le \mathrm{n}_2) \quad (6\text{-}4)$$

The partial similarity $\mathrm{SimBBLS}(i1,i2,j1,j2)$ between two basic block sequence can be expressed as:

$$\mathrm{SimBBLS}(i1,i2,j1,j2)=\mathrm{SimBBLS}(i2,j2)-\mathrm{SimBBL}S(i1-1,j1-1) \quad (6\text{-}5)$$

The normalized representation is:

$$\mathrm{len}(i1,i2,j1,j2)=\max\,(i2-i1+1,j2-j1+1) \quad (6\text{-}6)$$

$$\mathrm{NorSimBBLs}(i1,i2,j1,j2)=\frac{\mathrm{SimBBLS}(i1,i2,j1,j2)}{\mathrm{len}(i1,i2,j1,j2)} \quad (6\text{-}7)$$

6.2 Reduce of space complexity

The current space complexity of the algorithm is $\mathrm{O}(\mathrm{n}^2)$. n is the number of instructions in the program. The memory which the program needs is relatively large. As can be seen in Figure 5-1, During the calculation, when we use the dynamic programming algorithm to calculate $\mathrm{SimInS}(\mathrm{i},\mathrm{j})$, the information we need is only the last layer and this layer. That means the information before has been useless, so a two-dimensional array is sufficient to achieve the purpose of the calculation of each state. In this case,we reduce the space complexity and improve efficiency in the use of memory.

To sum up, after optimizing the time complexity and space complexity, the program can calculate the degree of similarity between the basic blocks, and thus from the bottom to up, we can also calculate the similarity between functions. Ultimately the similarity between any two applications can be calculated, so we achieve the purpose of identifying the behavior of the program.

7 Experimental results and analysis

To verify the accuracy, time complex and space complex of this algorithm, we adopt the programs form the program online judge system of Beijing University of Posts and Telecommunications. Users can submit a variety of source code on the same topic on it. We select 2000 samples of 20 kinds.

First we gather and classify all the samples. For each type of the programs, we select the most representative program. We don t describe the clustering algorithm detailedly here. Then we discriminate and classify many unknown programs. The result is correct when the original source code belongs to the same category, otherwise the result is wrong.

Final results are as follows: (Ins represents the average number of assembly instructions in all programs in that category, BBL represents the average number of basic blocks in all programs of this category, ETBO represents the elapsed time before optimization, ETAO represents the elapsed time after optimization, MBO represents the memory before optimization, MAO represents the memory after optimization).

Program behavior	Accuracy	Ins	BBL	ETBO	ETAO	MBO	MAO
Program behavior	100%	1567	305	10.000s	0.078s	1896kb	1888kb
Find the max	96%	108	20	0.093s	0.016s	1892kb	1888kb
Encryption	99%	555	152	1.077s	0.062s	1896kb	1888kb
Simple selection	98%	26	11	0.015s	0.001s	1888kb	1888kb
Sort the structure	94%	1953	288	14.118s	0.639s	1896kb	1888kb
Reverse a string	93%	647	89	2.792s	0.296s	1896kb	1888kb
Length of string	99%	244	64	0.390s	0.031s	1892kb	1888kb
Matrix operations	96%	858	177	4.275s	0.078s	1896kb	1888kb
Average	98%	744	138	4.095s	0.150s	1894kb	1888kb

Table 2 Experimental results

Experimental results show that the average accuracy rate is 98%.When we use the Instruction - Basic Block mode and scroll array, the time complexity and space complexity is reduced by several tens of times.

8 Conclusion

This paper puts forward a program behavior recognition algorithm based on the similarity of assembly instructions. We treat the assembly instructions as our basic size and generalise the model of assembly instruction sextuple. Based on this model, we design a matching algorithm for the assembly instruction sequence to detect the program behavior. Meanwhile, the algorithms are optimized for time complexity and space complexity, the experimental results show that this method can detect the basic behavior of the program, with the advantage of high accuracy. Next we hope to detect instruction semantics for further, constantly optimize the time complexity and maximize efficiency.

Acknowledgments. This work is supported by the National Natural Science Foundation of China (No. 61272493，61502536，U1536122).

References

1.http://www.bsa.org/~/media/Files/StudiesDownload/BSA_GSS_A4.pdf

2. Gröbert F, Willems C, Holz T. Automated Identification of Cryptographic Primitives in Binary Programs[J]. Lecture Notes in Computer Science, 2011:41-60.

3.Jingwei Zhang. Research on Public Key Cryptographic Algorithm Recognition Technology [D]. The PLA Information Engineering University, 2011.

4. LI Xiang, KANG Fei, SHU Hui. Cryptographic Algorithm Recognition Based on Dynamic Binary Analysis. Computer Engineering, 2012, 38(17): 106-109,115.

5.Caballero J, Yin H, Liang Z, et al. Polyglot: automatic extraction of protocol message format using dynamic binary analysis[J]. Ccs 07 Proceedings of Acm Conference on Computer & Communications Security Acm, 2007:317--329.

Energy-aware Migration of Virtual Machines in a Cluster

Dilawaer Duolikun, Shigenari Nakamura, Ryo Watanabe, Tomoya Enokido, and Makoto Takizawa

Abstract In order to realize eco-society, we have to reduce the electric energy consumed by servers. Virtual machines are now widely used to support applications with virtual computation service in server clusters. Here, a virtual machine can migrate to a guest server while processes are being performed. In the EAMV algorithm we previously proposed, the termination time of each process on each virtual machine has to be estimated. However, it is not easy to obtain the state of each process and takes time to calculate the expected termination time. In this paper, we newly propose a virtual machine migration (VMM) algorithm where termination time of each virtual machine is estimated without considering each process. We evaluate the VMM algorithm and show the total electric energy consumption and active time of servers and the average execution time of processes can be reduced in the VMM algorithm compared with non-migration algorithms. The VMM algorithm is simpler than the EAMV algorithm

Dilawaer Duolikun
Hosei University, Tokyo, Japan, e-mail: dilewerdolkun@gmail.com

Shigenari Nakamura
Hosei University, Tokyo, Japan, e-mail: nakamura.shigenari@gmail.com

Ryo Watanabe
Hosei University, Tokyo, Japan, e-mail: ryo.watanabe.4h@stu.hosei.ac.jp

Tomoya Enokido
Rissho University, Tokyo, Japan, e-mail: eno@ris.ac.jp

Makoto Takizawa
Hosei University, Tokyo, Japan, e-mail: makoto.takizawa@computer.org

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_3

1 Introduction

We have to reduce the electric energy consumed in information systems, especially server clusters [27] to realize eco society [26]. In order to discuss how to reduce the electric energy consumption of servers in a cluster, we first need a power consumption model which shows how much electric power a server consumes to perform application processes. Types of power consumption models are proposed in our previous studies [12, 13, 14, 20, 21, 22]. The power consumption models are also proposed for communication [16] and storage [17] types of application processes. In order to reduce the electric energy consumption, types of server selection algorithms [7, 13, 14, 17, 20, 21, 22] are proposed. Here, a server to perform a request process is selected so that the expected total electric energy consumption of the servers can be reduced. A process migration approach is also discussed where processes migrate to more energy-efficient servers [4, 5, 6, 7, 11]. However, it is not easy to migrate types of processes to servers with various architectures and operating systems.

A server cluster provides applications with virtual computation service by using virtual machines like KVM [25] and VMware [27]. Applications processes can be performed on a virtual machine without being conscious of what servers are included in a cluster. A virtual machine on a host server can migrate to a guest server while processes are being performed on the virtual machine [25]. The EAMV (Energy-Aware Migration of Virtual machines) algorithm [10] is proposed to select a virtual machine for a request process and migrate a virtual machine to another guest server. Here, the termination time of every current process on each virtual machine has to be estimated to obtain the expected electric energy consumption of the servers. In this paper, we newly propose a virtual machine migration (VMM) algorithm which is simpler than the EAMV algorithm. As discussed in paper [28], the average execution time of processes depends on the total number of current processes on a server and is independent of the number of virtual machines. A server is selected for a request process which is expected to consume the minimum electric energy to perform the process and every current process. Then, a virtual machine where the minimum number of processes are performed is selected in the selected server. If a server is expected to consume more electric energy to perform processes, one virtual machine is selected in the server, where the maximum number of processes are performed. Then, the selected virtual machine migrates to another guest server which is expected to consume smaller electric energy. We evaluate the VMM algorithm compared with non-migration algorithms. In the evaluation, we show the total electric energy consumption and active time of servers and the average execution time of processes are reduced in the VMM algorithm.

In section 2, we present a model of virtual machines. In section 3, we discuss power consumption and computation models of a server with virtual machines. In section 4, we propose the VMM algorithm. In section 5, we evaluate the VMM algorithm.

2 System Model

A cluster S is composed servers s_1, …, s_m ($m \geq 1$). A server s_t is equipped with a set CP_t of np_t (≥ 1) homogeneous CPUs, cp_{t0}, …, cp_{t,np_t-1}. Each CPU cp_{tk} is composed of nc_{tk} (≥ 1) cores c_{tk0}, …, $c_{tk,nc_{tk}-1}$. Each core c_{tki} supports a set $\{th_{tki0}$, …, $th_{tki,ct_{tki}-1}\}$ of threads ($ct_{tki} \geq 1$). Here, $nc_{tk} = nc_t$ and $ct_{tki} = ct_t$ for each core ct_{tki}. A server s_t supports processes with the total number nt_t of threads, where $nt_t = np_t \cdot ct_{tk} \cdot nc_t$.

A server s_t is modeled to support processes with vt_t (≥ 1) virtual processors vt_{t0}, …, vt_{t,vt_t-1}. In this paper, every virtual processor is homogeneous in each server s_t. Applications can use virtual processors to perform processes without being conscious of which thread in which core of which CPU is supported. One virtual processor is at a time allocated to a process p_i [24]. In this paper, we assume each virtual processor is in a one-to-one correspondent relation with one thread. Hence, $vt_t = nt_t$. A virtual processor is *active* if at least one process is performed, otherwise *idle*. A server is *active* if and only if (iff) at least one virtual processor is active, otherwise *idle*. In this paper, a *process* means an application process to be performed on a server, which uses CPU.

A cluster S supports applications with a set VM of virtual machines $\{VM_1$, …, $VM_v\}$ ($v \geq 0$). Each virtual machine VM_h is supported with virtual processors of a server s_t. Here, s_t is a *host* server of the virtual machine VM_h and VM_h is a *resident* virtual machine of the server s_t. $SVM_t(\tau)$ shows a set of resident virtual machines on a host server s_t and $HS_h(\tau)$ denotes a host server of a virtual machine VM_h at time τ. $VP_h(\tau)$ ($\subseteq VP_t$) shows a subset of virtual processors on a host server s_t, which are allocated to a virtual machine VM_h at time τ. One virtual machine VM_h ($\in VM$) on a host server s_t is selected for a process p_i issued by a client. Then, the process p_i is performed on the virtual machine VM_h. Here, the process p_i is a *resident* process of the virtual machine VM_h. $VCP_h(\tau)$ shows a set of resident processes of a virtual machine VM_h at time τ. A virtual machine VM_h is *active* at time τ if $|VCP_t(\tau)| > 0$, i.e. at least one process is performed, otherwise *idle*. $CP_t(\tau)$ is a set of all the resident processes performed on virtual machines of a server s_t at time τ, i.e. $CP_t(\tau) = \cup_{VM_h \in SVM_t(\tau)} VCP_h(\tau)$.

A virtual machine VM_h on a host server s_t can migrate to a guest server s_u. First, a copy of memory of a virtual machine VM_h is created on a guest server s_u. On issuing a migration command [25] on the host server s_t, the memory state of VM_h is first transferred to the server s_u while processes are being performed. On termination of the state transfer to the host server s_u, the processes are resumed on the virtual machine VM_h and the state of VM_h changed after the state transfer is transfered to the server s_u. Then, the processes on the virtual machine VM_h are restarted on the server s_u.

3 Power Consumption and Computation Models

3.1 MLPCM and MLC Models

The electric power consumption $E_t(\tau)$ [W] of a server s_t with multiple CPUs to perform computation processes at time τ is given as follows [23]:

[Multi-Level Power Consumption with Multiple CPUs (MLPCM) model]

$$E_t(\tau) = minE_t + \sum_{k=0}^{np_t-1}\{\gamma_{tk}(\tau)\ [bE_t + \sum_{i=0}^{nc_t-1} \alpha_{tki}(\tau)(cE_t + \beta_{tki}(\tau)\ tE_t)]. \quad (1)$$

Here, $\gamma_{tk}(\tau) = 1$ if a CPU cp_{tk} is active. Otherwise, $\gamma_{tk}(\tau) = 0$. That is, $E_t(\tau) = minE_t$ [W] in an idle server. $\alpha_{tki}(\tau) = 1$ if a core c_{tki} is active on a CPU cp_{tk}. Otherwise, $\alpha_{tki}(\tau) = 0$. $\beta_{tki}(\tau)$ ($\leq ct_t$) is the number of active threads on a core c_{tki}.

In Linux operating systems, processes are allocated to nt_t ($\geq$ 1) virtual processors, in the round-robin (RR) algorithm [24]. The, electric power consumption $CE_t(n)$ [W] of a server s_t to concurrently perform n ($\geq$ 1) processes at time τ is given in the MLPCM model [23] as follows:

[MLPCM model]

$$CE_t(n) = \begin{cases} minE_t\ if\ n = 0. \\ minE_t + n \cdot (bE_t + cE_t + tE_t)\ if\ 1 \leq n \leq np_t. \\ minE_t + np_t \cdot bE_t + n(cE_t + tE_t)\ if\ np_t < n \leq nc_t \cdot np_t. \\ minE_t + np_t \cdot (bE_t + nc_t \cdot cE_t) + nt_t \cdot tE_t\ if\ nc_t \cdot np_t < n < nt_t. \\ maxE_t\ if\ n \geq nt_t. \end{cases} \quad (2)$$

In this paper, we assume $E_t(\tau) = CE_t(|CP_t(\tau)|)$ for each server s_t. The total electric energy $TE_t(st, et)$ [J] consumed by a server s_t from time st to time et is TE_t $(st, et) = \sum_{\tau=st}^{et} E_t(\tau)$.

It takes T_{ti} [sec] to perform a process p_i on a thread in a server s_t. If only a process p_i is exclusively performed on a server s_t without any other process, the execution time T_{ti} of the process p_i is minimum, i.e. T_{ti} = $minT_{ti}$. In a cluster S of servers s_1, $\ldots$, s_m ($m \geq 1$), $minT_i$ shows a minimum one of $minT_{1i}, \ldots, minT_{mi}$. That is, $minT_i$ = $minT_{fi}$ on the fastest thread which is on a server s_f in the cluster S. Here, the server s_f is referred to as $fastest$. We assume one virtual computation step [vs] is performed on the fastest server s_f for one time unit [tu]. This assumption means, the maximum computation rate $maxCRT_f$ of a fastest server s_f is assumed to be one [vs/sec]. Here, $maxCRT = maxCRT_f$. On another slower server s_t, $maxCRT_t \leq maxCRT_f$ (= 1). The total number VC_i of virtual computation steps to be performed in a process p_i is defined to be $minT_i$ [sec] $\cdot$ $maxCRT_f$ [vs/sec] = $minT_i$ [vs] where a server s_f is the fastest. The maximum computation rate $maxCR_{ti}$ of a process p_i on a server s_t is VC_i / $minT_{ti}$ [vs/sec] ($\leq$ 1). On a fastest server s_f, $maxCR_{fi}$

= $maxCRT$ $=$ 1. For every pair of processes p_i and p_j on a server s_t, $maxCR_{ti}$ = $maxCR_{tj}$ = $maxCRT_t$ ($\leq$ 1). The maximum computation rate $maxCR_t$ of a server s_t is $nt_t \cdot maxCRT_t$.

[Multi-level computation (MLC) model] [20, 21, 22] The computation rate $CR_{ti}(\tau)$ [vs/sec] of a process p_i on a server s_t at time τ is given as follows:

$$CR_{ti}(\tau) = \begin{cases} maxCR_t \;/\; |CP_t(\tau)| & if\; |CP_t(\tau)| > nt_t. \\ maxCRT_t & if\; |CP_t(\tau)| \leq nt_t. \end{cases} \tag{3}$$

Suppose a process p_i on a server s_t starts at time st and ends at time et. Here, $\sum_{\tau=st}^{et} CR_{ti}(\tau) = VC_i$ [vs]. At time τ a process p_i starts on a server s_t, the computation laxity $plc_{ti}(\tau)$ of a process p_i is VC_i. At each time τ, $plc_{ti}(\tau)$ is decremented by the computation rate $CR_{ti}(\tau)$.

[Computation of a process p_i]

1. At initial time τ the process p_i starts, $plc_{ti}(\tau) = VC_i$;
2. At each time τ, $plc_{ti}(\tau+1) = plc_{ti}(\tau)$ - $CR_{ti}(\tau)$;
3. Then, if $plc_{ti}(\tau+1) \leq 0$, p_i terminates at time τ;

3.2 Computation Model of a Virtual Machine

Let p_{hi} show a process p_i performed on a virtual machine VM_h of a server s_t. $plc_{hi}(\tau)$ is the computation laxity $plc_{ti}(\tau)$ of a process p_{hi} on the server s_t at time τ. The *virtua*l *machine* (*VM*) *laxity* $vlc_h(\tau)$ [vs] of a virtual machine VM_h at time τ is defined to be the summation of computation laxities of the resident processes of VM_h:

- $vlc_h(\tau) = \sum_{p_i \in VCP_h(\tau)} plc_{hi}(\tau)$.

The *server laxity* $slc_t(\tau)$ [vs] of a server s_t is the summation of VM laxities of virtual machines hosted by the server s_t at time τ:

- $slc_t(\tau) = \sum_{VM_h \in SVM_t(\tau)} vlc_h(\tau)$.

The *VM* computation rate $VCR_h(\tau)$ [vs/sec] of a virtual machine VM_h is defined as follows:

[Virtual machine (VM) computation ratio] The *VM* computation rate $VCR_h(\tau)$ of a virtual machine VM_h on a server s_t at time τ is given as follows:

$$VCR_h(\tau) = \begin{cases} maxCR_t \cdot |VCP_h(\tau)| \;/\; |CP_t(\tau)| & if\; |CP_t(\tau)| > nt_t. \\ |VCP_h(\tau)| \cdot maxCRT_t & if\; |CP_t(\tau)| \leq nt_t. \end{cases} \tag{4}$$

Here, $VCR_h(\tau) \leq VCR_k(\tau)$ if $|VCP_h(\tau)| \leq |VCP_k(\tau)|$ for every pair of different virtual machines VM_h and VM_k on a same server s_t. $VCR_h(\tau)$ / $VCR_k(\tau)$ =

$|VCP_h(\tau)|$ / $|VCP_k(\tau)|$. The computation rate $CR_{ti}(\tau)$ of each process p_i depends on the total number $|CP_t(\tau)|$ of processes but is independent of the number $|SVM_t(\tau)|$ of virtual machines of a host server s_t [28].

The VM laxity $vlc_h(\tau)$ of a virtual machine VM_h and the server laxity $slc_t(\tau)$ of a server s_t which hosts VM_h are manipulated as follows:

[VM computation (VMC) model]

```
    $VCP_h = VCP_h(\tau)$;
    **while** ($VCP_h \neq \phi$) {
    **for** each process $p_i$ on a virtual machine $VM_h$ in $SVM_t(\tau)$, i.e. $p_i \in VCP_h$,
        $plc_{hi}(\tau+1) = plc_{hi}(\tau) - VCR_h(\tau)$ / $|VCP_h|$;
    **if** $plc_{hi}$ $(\tau + 1) \leq 0$, $p_i$ terminates at time $\tau$ and $VCP_h = VCP_h - \{p_i\}$;
    $vlc_h(\tau+1) = vlc_h(\tau) - VCR_h(\tau)$;
    **if** $vlc_h(\tau+1) \leq 0$, every process on $VM_h$ terminates, i.e. $VM_h$ gets idle;
    $\tau = \tau + 1$;
};  /* **while** end */
```

A virtual machine VM_h is referred to as *terminate* if VM_h gets idle, i.e. no process is performed on VM_h In this paper, we estimate the termination time ET_t and electric energy consumption EE_t of a server s_t to perform every process by considering active virtual machines, not each process as follows:

[Virtual machine computation (VMC) model]

```
**VMEST** $(s_t, \tau; EE_t, ET_t)$
**input**  $s_t$;  $\tau$;
**output** $EE_t$;  $ET_t$;
  { $ncp = |CP_t(\tau)|$; /*number of processes on $s_t$*/
    $vlc = 0$;
    $SVM = SVM_t(\tau)$; /* set of virtual machines on $s_t$ */
    $x = \tau$;
    $EE_t = 0$;
  /* obtain laxity $vlc$ of the server $s_t$ */
    **for** each virtual machine $VM_h$ **in** $SVM$, /* $VM$ laxity of $VM_h$ */
      $vlc_h = vlc_h(\tau)$ $(= \sum_{p_i \in VCP_h(\tau)} plc_i(\tau))$;
      $ncp_h = |VCP_h(\tau)|$; /*number of processes on $VM_h$*/
      $vlc = vlc + vlc_h$; /* server laxity of $s_t$ */
    }; /* **for** end */
    **while** ($SVM \neq \phi$) {
      $EE_t = EE_t + CE_t(ncp)$; /* electric energy */
      **for** each virtual machine $VM_h$ **in** $SVM$, {
        $vlc_h = vlc_h - VCR_h(\tau)$; /* VM laxity is decremented */
        **if** $vlc_h \leq 0$, /*$VM_h$ gets idle, i.e. terminates */ {
            $SVM = SVM - \{VM_h\}$;
            $ncp = ncp - ncp_h$;
        } **else** $vlc = vlc - vlc_h$; /*decrement server laxity*/
      }; /* **for** end */
      $x = x + 1$; /* time advances */
    }; /* **while** end */
    $ET_t = x - 1$; /* every $VM$ terminates, i.e. gets idle on $s_t$*/
};
```

Here, the VM computation rate $VCR_h(\tau)$ of a virtual machine VM_h depends on how many number of processes are totally performed on VM_h. The more number of processes are performed on a virtual machine VM_h, the larger VM computation rate $VCR_h(\tau)$. Here, it is noted we do not consider the termination time of each process p_i and only consider each virtual machine.

4 A Virtual Machine Migration (VMM) Algorithm

A client issues a process p_i to virtual machines $VM_1, \ldots, VM_v$ ($v \geq 1$) in a cluster S. The expected electric energy consumption EE_t and expected termination time ET_t of a server s_t to perform every current process on the virtual machines are obtained by the procedure **VMEST** $(s_t, \tau; ET_t, EE_t)$. Then, one virtual machine VM_h on a server s_t is selected to perform a process p_i as follows:

[VM selection]

```
  for each server $s_u$ in a cluster $S$, VMEST $(s_u, \tau; EE_u, ET_u)$;
  $MS = \{s_u \mid EE_u$ is minimum in $S\}$;
  select $s_t$ in $MS$ where $|CP_t(\tau)|$ is minimum;
  select a virtual machine $VM_h$ in $s_t$ where $|VCP_h(\tau)|$ is minimum;
```

Then, the process p_i is performed on the selected virtual machine VM_h in the selected host server s_t.

A server s_t is *overloaded* at time τ iff $|CP_t(\tau)| > maxNCP_t$. For example, the computation rate $CR_{ti}(\tau)$ of each process p_i should be larger than $\alpha \cdot maxCR_t$. Since $CR_{ti}(\tau) < \alpha \cdot maxCR_t$, $CR_{ti}(\tau) = nt_t \cdot maxCR_t \,/\, |CP_t(\tau)|$, $nt_t \cdot maxCR_t \,/\, maxNCP_t = \alpha \cdot maxCR_t$. Hence, $maxNCP_t = nt_t \,/\, \alpha$. A server s_t more overloaded than a server s_u if $|CP_t(\tau)| \,/\, maxNCP_t > |CP_u(\tau)| \,/\, maxNCP_u$.

First, an *overloaded* server s_t is selected whose expected electric energy EE_t is the largest in a cluster S. Then, a virtual machine VM_h is selected in the selected server s_t, $VM_h \in SVM_t(\tau)$, where the number $|VCP_h(\tau)|$ of processes performed on VM_h is minimum.

[VM selection in s_t]

```
  for each server $s_u$ in a cluster $S$, VMEST $(s_u, \tau; EE_u, ET_u)$;
  $OS = \{s_u \mid s_u$ is overloaded and $SVM_u(\tau) \neq \phi$ in $S\}$;
  while ($OS \neq \phi$)
  {
    select $s_t$ whose $EE_t$ is maximum in $OS$;
      while ($s_t$ is overloaded)
      {
            select a virtual machine $VM_h$ in $SVM_t(\tau)$ where $|VCP_h(\tau)|$ is minimum;
            select a server $s_u$ where $|CP_u(\tau)|$ is minimum and which is not overloaded;
            if not found, break;
            migrate $VM_h$ from $s_t$ to $s_u$;
        }; /* while end */
    OS = OS − $\{s_t\}$;
}; /*while end */
```

5 Evaluation

We evaluate the VMM algorithm in terms of the total electric energy consumption TEE [J] and total active time TAT [sec] of servers and the average execution time AET [sec] of processes compared with the random (RD), round robin (RR), and NVM (non-migration of virtual machines). In the NVM algorithm, a virtual machine is selected in the same VM selection algorithm as the VMM algorithm but no virtual machine migrates. In the RD algorithm, one virtual machine VM_h is randomly selected. In the RR algorithm, a virtual machine VM_h is selected after a virtual machine VM_{h-1}. In the RD, RR, and NVM algorithms, every virtual machine VM_h does not migrate. In the VMM algorithm, each virtual machine VM_h migrates to a guest server.

There are m heterogeneous servers $s_1, \ldots, s_m$ in a cluster S. The power consumption parameters like $minE_t$ and $maxE_t$ [W] and the performance parameters like $maxCRT_t$ and $maxCR_t$ of a server s_t are randomly taken as shown in Table 1. There are a set VM of v ($\geq$ 1) virtual machines $VM = \{VM_1, \ldots, VM_v\}$. In the evaluation, $m = 6$ and $v = 8$.

The number n ($\geq$ 1) of processes $p_1, \ldots, p_n$ are randomly issued to the cluster S. In the simulation, one time unit [tu] is assumed to be 100 [msec]. In each process configuration PF_{ng}, the minimum execution time $minT_i$ of each process p_i is randomly taken from 5 to 10 [tu], i.e. 0.5 to 1.0 [sec]. The amount VS_i [vs] of virtual computation steps of each process p_i is $minT_i$ as discussed in this paper. The start time $stime_i$ of each process p_i is randomly taken from 0 to $xtime$ - 1. The simulation time $xtime$ is 200 [tu] (= 20 [sec]). The simulation is time-based. We randomly generate four process configurations $PF_{n1}, \ldots, PF_{n4}$ of the processes $p_1, \ldots, p_n$.

We randomly generate four server configurations $SF_1, \ldots, SF_4$ of the servers $s_1, \ldots, s_m$ ($m = 6$). In each server configuration SF_k, the parameters of each server s_t are randomly taken. We also generate four VM configurations $VF_1, \ldots, VF_4$ of the virtual machines $VM_1, \ldots, VM_v$ ($v = 8$). In each VM configuration VF_l, initially each virtual machine VM_h is randomly deployed on a server. For each combination of the configurations SF_k, VF_l, and PF_{ng}, the electric energy consumption EE_t and active time AT_t of each server s_t and the execution time ET_i of each process p_i are obtained.

Figure 1 shows the total electric energy consumption (TEE) [J] of six servers $s_1, \ldots, s_6$ ($m = 6$) with eight virtual machines $VM_1, \ldots, VM_8$ ($v = 8$) for number n of

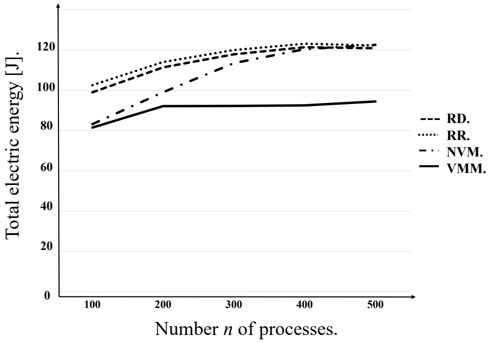


Fig. 1 Total electric energy consumption.

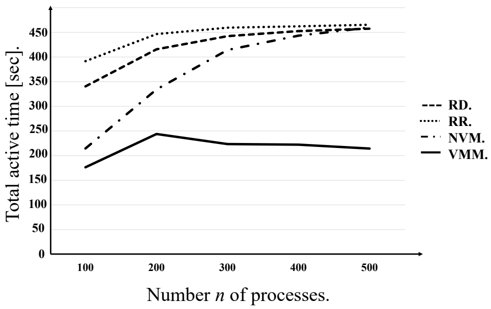


Fig. 2 Total active time of servers.

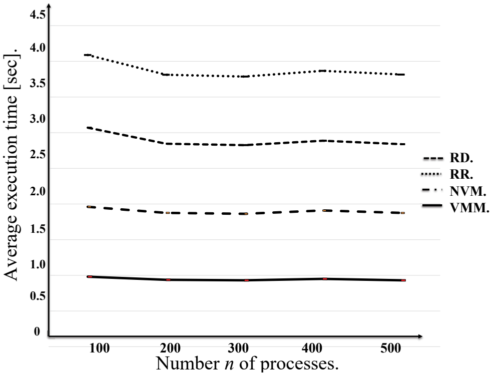


Fig. 3 Average execution time of processes.

processes. TEE is the summation $EE_1 + \ldots + EE_m$. As shown in Figure 1, TEE of the VMM algorithm is smaller than the other non-migration algorithms. Thus, TEE can be reduced in the VMM algorithm.

Figure 2 shows the total active time (TAT) [sec] of six servers ($m = 6$) for the number n of processes. TAT is $AT_1 + \ldots + AT_m$. TAT in the VMM algorithm is shorter than the other algorithms. This means, the servers are more lightly loaded in the VMM algorithm than the other algorithms.

Figure 3 shows the average execution time (AET) [sec] of the number n of processes. AET is $(ET_1 + \ldots + ET_n) / n$. AET of the VMM algorithm is shorter than the other algorithms.

As shown here, the total electric energy consumption and active time of servers and average execution time of processes can be reduced in the VMM algorithm.

Table 1 Parameters.

parameters	values
m	number of servers $s_1, \ldots, s_m$ (≥ 1).
np_t	number of CPUs (≤ 2).
nc_t	number of cores (1, 2, 4, 6)/ CPU.
ct_t	threads/core (≤ 2)
nt_t	number of threads (= $ct_t \cdot np_t \cdot nc_t$).
$maxCRT_t$ [vs/tu]	0.5 $\sim$ 1.
$maxCR_t$ [vs/tu]	$nt_t \cdot maxCRT_t$.
$minE_t$ [W]	80 $\sim$ 100.
$maxE_t$ [W]	$minE_t$ $+E$ ($50 \leq E \leq 150$).
bE_t [W]	$(maxE_t$ - $minE_t)$ / $(4 \cdot np_t)$.
cE_t [W]	$5 \cdot (maxE_t$ - $minE_t)$ / $(8 \cdot np_t \cdot nc_t)$.
tE_t [W]	$(maxE_t$ - $minE_t)$ / $(8 \cdot nt_t)$.
n	number of processes $p_1, \ldots, p_n$ (≥ 1).
$minT_i$ [tu]	minimum computation time of p_i (0.5 $\sim$ 1.0).
VS_i [vs]	$VS_i = minT_i$.
$stime_i$	starting time of p_i ($0 \leq st_i < xtime$ - 1).
$xtime$	simulation time (= 200 [tu] = 20 [sec]).
v	number of virtual machines $VM_1, \ldots, VM_v$.

6 Concluding Remarks

In this papers, we proposed the VMM algorithm to reduce the electric energy consumption of servers in a cluster. A virtual machine migrates from a host server to a guest server if the guest server is expected to consume smaller electric energy than the host server. The termination time of every current process is estimated for each virtual machine without considering each process The computation time of the VMM algorithm is smaller than the other algorithms. In the evaluation, we showed the total electric energy consumption and active time of servers and the average execution time of processes can be reduced in the VMM algorithm compared with the non-migration algorithms.

Acknowledgements This work was supported by JSPS KAKENHI grant number 15H02695.

References

1. Aikebaier, A., Enokido, T., and Takizawa, M.: Energy-Efficient Computation Models for Distributed Systems, Proc. of the 12th International Conference on Network-Based Information Systems (NBiS-2009), pp.424-431, (2009).
2. Coulouris, G., Dollimore, J., Kindberg, T., and Blair, G.: Distributed Systems Concepts and Design, 4th ed., Addison-Wesley, (2012).
3. Duolikun, D., Aikebaier, A., Enokido, T., and Takizawa, M.: Energy-aware Passive Replication of Processes, Journal of Mobile Multimedia, **9**(1&2), pp.53–65, (2013).
4. Duolikun, D., Aikebaier, A., Enokido, T., and Takizawa, M.: Power Consumption Models for Redundantly Performing Mobile-Agents, Proc. of the 8th International Conference on Complex, Intelligent and Software Intensive Systems (CISIS-2014), pp.185-190, (2014).
5. Duolikun, D., Aikebaier, A., Enokido, T., and Takizawa, M.: Power Consumption Models for Migrating Processes in a Server Cluster Proc. of the 17th International Conference on Network-Based Information Systems (NBiS-2014), pp.155-162, (2014).
6. Duolikun, D., Enokido, T., and Takizawa, T.: Asynchronous Migration of Process Replicas in a Cluster, Proc. of IEEE the 29th International Conference on Advanced Information Networking and Applications (AINA-2015), pp.271-278, (2015).
7. Duolikun, D., Enokido, T., and Takizawa, T.: Energy-Efficient Replication and Migration of Processes in a Cluster, Proc. of the 9th International Conference on Complex, Intelligent and Software Intensive Systems (CISIS-2015), pp.118–125, (2015).
8. Duolikun, D., Aikebaier, A., Enokido, T., and Takizawa, M.: Energy-Efficient Dynamic Clusters of Servers, Journal of Supercomputing, **71**(5), pp.1642–1656, (2015).
9. Duolikun, D., Watanabe, R., Enokido, T., and Takizawa, T.: A Model for Migration of Virtual Machines to Reduce Electric Energy Consumption, Proc. of the 10th International Conference on Complex, Intelligent and Software Intensive Systems (CISIS-2016), pp.160-166, (2016).
10. Duolikun, D., Watanabe, R., Enokido, T., and Takizawa, T.: A Model for Migration of Virtual Machines to Reduce Electric Energy Consumption, Proc. of the 19th International Conference on Network-based Information Systems (NBiS-2016), CD-ROM, (2016).
11. Duolikun, D., Nakamura, S., Enokido, T., and Takizawa, M.: An Energy-efficient Process Migration Approach to Reducing Electric Energy Consumption in a Cluster of Servers, International Journal of Communication Networks and Distributed Systems, **15**(4), pp.400-420, (2015).
12. Enokido, T., Aikebaier, A., Deen, S, M., and Takizawa, M.: Power Consumption-based Server Selection Algorithms for Communication-based Systems, Proc. of the 13th International Conference on Network-based Information Systems (NBiS-2010), pp.201-208, (2010).
13. Enokido, T., Aikebaier, A., and Takizawa, M.: A Model for Reducing Power Consumption in Peer-to-Peer Systems, IEEE Systems Journal, **4**(2), pp.221-229, (2010).
14. Enokido, T., Aikebaier, A., and Takizawa, M.: Process Allocation Algorithms for Saving Power Consumption in Peer-to-Peer Systems, IEEE Transactions on Industrial Electronics, **58**(6), pp.2097-2105, (2011).
15. Enokido, T. and Takizawa, M.: An Extended Power Consumption Model for Distributed Applications, Proc. of IEEE the 26th International Conference on Advanced Information Networking and Applications (AINA-2012), pp.912-919, (2012).
16. Enokido, T. and Takizawa, M.: An Integrated Power Consumption Model for Distributed Systems, IEEE Transactions on Industrial Electronics, **60**(2), pp.824-836, (2013).
17. Enokido, T., Aikebaier, A., and Takizawa, M.: An Extended Simple Power Consumption Model for Selecting a Server to Perform Computation Type Processes in Digital Ecosystems, IEEE Transactions on Industrial Informatics, **10**(2), pp.1627-1636, (2014).
18. Enokido, T. and Takizawa, M.: An Energy-Efficient Load Balancing Algorithm for Virtual Machine Environment to Perform Communication Type Application Processes, Proc. of IEEE the 30th International Conference on Advanced Information Networking and Applications (AINA-2016), pp. 392-399, (2016).

19. Ghemawat, S., Gobioff, H., and Leung, S, T.: The Google File System, Proc. of ACM the 19th Symposium on Operating System Principle (SOPI 03), pp.29-43, (2003).
20. Kataoka, H., Duolikun, D., Enokido, T., and Takizawa, T.: Power Consumption and Computation Models of a Server with a multi-core CPU and Experiments, Proc. of IEEE the 29th International Conference on Advanced Information Networking and Applications (AINA-2015), pp.217-222, (2015).
21. Kataoka, H., Duolikun, D., Enokido, T., and Takizawa, T.: Evaluation of Energy-Aware Server Selection Algorithms, Proc. of the 9th International Conference on Complex, Intelligent and Software Intensive Systems (CISIS-2015), pp.318–325, (2015).
22. Kataoka, H., Duolikun, D., Enokido, T., and Takizawa, T.: Multi-level Computation and Power Consumption Models, Proc. of the 18th International Conference on Network-based Information Systems (NBiS-2015), pp.40–47, (2015).
23. Kataoka, H., Sawada, A., Duolikun, D., Enokido, T., and Takizawa, T,: Energy-aware Server Selection Algorithms in a Scalable Cluster, Proc. of IEEE the 30th International Conference on Advanced Information Networking and Applications (AINA-2016), pp. 565-572, (2016).
24. Negus, C. and Boronczyk, T.: CentOS Bible, ISBN: 978-0-470-48165-3, (2009).
25. Rosa, J., D., la.: KVM Virtualization in RHEL 6 Made Easy, Dell Linux Engineering, (2011).
26. 2015 United Nations Climate Change Conference (COP21).
https://en.wikipedia.org/wiki/2015 United Nations Climate Change Conference.
27. VMware Virtualization, http://www.vmware.com/jp.html
28. Watanabe, R., Duolikun, D., Enokido, T., and Takizawa, M.: An Eco Model of Process Migration with Virtual Machines, Proc. of the 19th International Conference on Network-based Information Systems (NBiS-2016), pp.292–297, (2016).

An Energy-efficient Migration Model of Processes with Virtual Machines in a Server Cluster

Ryo Watanabe, Dilawaer Duolikun, Tomoya Enokido, and Makoto Takizawa

Abstract In cloud computing systems, computation resources like CPU and storages are virtualized. Virtual machines are now widely used to support applications with virtual computation service and to perform application processes. Furthermore, virtual machines can migrate from a host server to a guest server while processes are being performed on the virtual machines. We have to reduce electric energy consumption of servers in server clusters. In this paper, we take advantage of the migration technologies of virtual machines to reduce the electric energy consumed by servers. We propose a simple virtual machine migration (SVM) algorithm to migrate a virtual machine to another energy-efficient server in order to reduce the electric energy consumption. We show the total electric energy consumption of the servers can be reduced in the SVM algorithm.

1 Introduction

We have to reduce electric energy consumption of clusters like cloud computing systems in order to realize eco society [14, 22, 23]. There are approaches to reducing the electric energy consumption of servers [1, 2, 3, 7, 8, 10, 11, 13, 16, 18, 19]. In one approach, if a client issues a request process to a cluster, one energy-efficient

Ryo Watanabe
Hosei University, Tokyo, Japan e-mail: ryo.watanabe.4h@stu.hosei.ac.jp

Dilawaer Duolikun
Hosei University, Tokyo, Japan e-mail: dilewerdolkun@gmail.com

Tomoya Enokido
Rissho University, Tokyo, Japan e-mail: eno@ris.ac.jp

Makoto Takizawa
Hosei University, Tokyo, Japan e-mail: makoto.takizawa@computer.org

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_4

server is selected to perform the process. Types of algorithms to select an energy-efficient server to perform a process are proposed [9, 12, 17, 20].

In another migration approach, a process on a server migrates to another server. For example, a process migrates from a server to another server which is expected to consume smaller electric energy [4, 5, 6]. However, it is not easy to migrate types of processes among servers with various architectures and operating systems. Cloud computing systems support applications with virtual computation services by using virtual machines [24, 25]. Furthermore, a virtual machine can migrate from a host server to another guest server while processes are being performed on the virtual machine. Thus, it is easy to realize the process migration by using virtual machine technologies. In our previous studies [5, 6], a virtual machine to perform a process and to migrate to a guest server is selected by estimating time when each process terminates and electric energy to be consumed by each server. However, it takes time and is also difficult to collect plenty of information of processes and virtual machines on each server to do the estimation.

In this paper, we newly propose a *simple virtual machine migration* (*SVM*) algorithm where we do not estimate the termination time of each process. A virtual machine is idle if no process is performed. Time when a virtual machine gets idle is termination time. Without considering each process, the termination time of each virtual machine is estimated. We can reduce time to estimate the termination time of processes and the electric energy consumption of servers in a cluster by considering virtual machines as units of computation. We evaluate the SVM algorithm compared with non-migration algorithms. In the evaluation, we show the total electric energy consumption of a cluster can be reduced in the SVM algorithm.

In section 2, we present a system model. In section 3, we present the power consumption model and computation model of a virtual machine. In section 4, we propose the SVM algorithm to select a server in a cluster. In section 5, we evaluate the SVM algorithm.

2 System Model

A cluster S is composed of physical servers $s_1, \dots, s_m$ $(m \geq 1)$. A cluster S supports applications on clients with virtual computation service by using virtual machines [24, 25]. Here, applications can use computation resources like CPU in a cluster without being conscious of which servers support the computation resources. Thus, a cluster S supports applications with a set VM of virtual machines $VM_1, \dots, VM_v$ $(v \geq 1)$ like KVM (Kernel-based virtual machine) [24] and VMware [25]. If a client issues a request to the cluster S, one virtual machine VM_h is selected in the cluster S. A process p_i to handle the request is created and performed on a virtual machine VM_h. Let p_{hi} denote a process p_i performed on a virtual machine VM_h. $HS_h(\tau)$ shows a host server s_t of a virtual machine VM_h at time τ. $VCP_h(\tau)$ indicates a set of processes performed on VM_h at time τ. $SVM_t(\tau)$ denotes a set of virtual machines on a server s_t. Furthermore, a virtual machine VM_h can migrate from a host server

s_t to another guest server s_u while processes are being performed on the virtual machine VM_h without suspending the processes. For example, a current host server s_t of VM_h is heavily loaded and another server s_u is less loaded. The virtual machine VM_h migrates from the host server s_t to the guest server s_u. Then, processes on VM_h can be more efficiently performed on the server s_u.

For each virtual machine VM_h on a host server s_t, memory area to store the virtual machine VM_h is first allocated in each server. On issuing a migration command [24] to the host server s_t, the memory state of the virtual machine VM_h is transfered to a guest server s_u. Here, processes on VM_h are performed on the host server s_t. When the memory state of VM_h is transfered to the guest server s_u, the processes are suspended and a part of the memory state which is changed after the memory state transmission starts is transrated to the guest server s_u. Then, the processes on VM_h are restarted on the server s_u.

We consider a pair of virtual machines VM_1 and VM_2 which are realized in KVM [24] on a server s_t with one CPU (Intel corei7-6700k) and memory (16 GB). Totally n (≥ 1) processes $p_1, \ldots, p_n$ are concurrently performed on the server s_t. Figure 1 shows the average execution time of the n (n = 120, 180, 240) processes on v (≤ 2) virtual machines $VM_1, \ldots, VM_v$. "$v = 0$" means the processes are directly performed on the server s_t. "$v = 1$" shows that every process is performed on a virtual machine, say VM_1 of the server s_t. "$v = 2$" indicates $n/2$ processes are performed on each of the virtual machines VM_1 and VM_2. As shown in Figure 1, the average execution time of the processes depends on the number n of the processes performed on the server s_t but is independent of the number v of virtual machines on the server s_t.

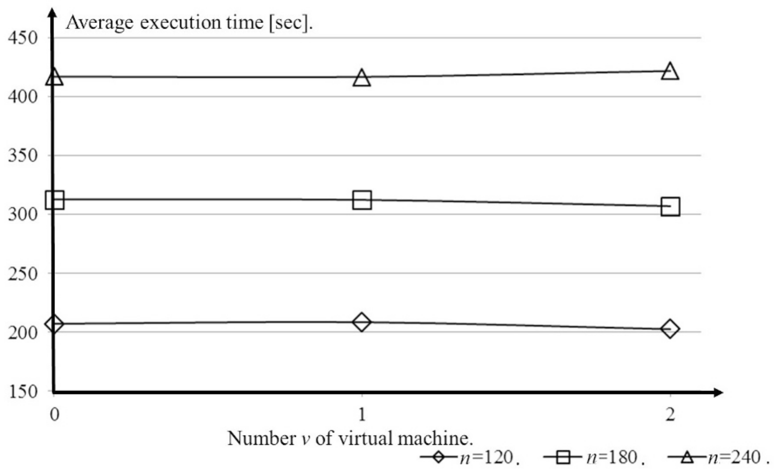


Fig. 1 Average execution time of processes on VM_1 and VM_2.

3 Power Consumption and Computation Models

3.1 MLPCM Model

In this paper, a process means a computation type of application process where CPU resource is used. A server s_t is composed of np_t (≥ 1) homogeneous CPUs $cp_{t0}, \ldots, cp_{t,np_t-1}$. Each CPU cp_{tk} is composed of nc_t (≥ 1) homogeneous cores $c_{tk0}, \ldots, c_{tk,nc_t-1}$. Each core c_{tkh} supports the same number ct_t (≤ 2) of threads. The total number nt_t of threads on a server s_t is $np_t \cdot nc_t \cdot ct_t$. An *active* thread and core are ones where at least one process is performed. An *active* server is a server where at least one thread is active. A server which is not active is *idle*.

Let $CP_t(\tau)$ be a set of processes performed on a server s_t at time τ. The electric power consumption $E_t(\tau)$ [W] of a server s_t to perform processes at time τ is given in the MLPCM (Multi-Level Power Consumption with Multiple CPUs) model as follows [17, 18, 19, 20]:

$$E_t(\tau) = minE_t + \textstyle\sum_{k=0}^{np_t-1} \{\gamma_{tk}(\tau)\ [bE_t + \displaystyle\sum_{i=0}^{nc_t-1} \alpha_{tki}(\tau)(cE_t + \beta_{tki}(\tau)\, tE_t)]\}. \tag{1}$$

Here, $\gamma_{tk}(\tau)$ = 1 if at least one core is active on a CPU cp_{tk} at time τ (k = 0, 1, $\ldots$, $np_t - 1$). Otherwise, $\gamma_{tk}(\tau)$ = 0. Even an idel server s_t consumes the minimum electric power $E_t(\tau)$ = $minE_t$ [W] at time τ. $\alpha_{tki}(\tau)$ = 1 if a core c_{tki} is active on a CPU cp_{tk}. Otherwise, $\alpha_{tki}(\tau)$ = 0. $\beta_{tki}(\tau)$ ($\leq ct_t$) is the number of active threads on a core c_{tki}. If $\alpha_{tki}(\tau) = 0$, $\beta_{tki}(\tau) = 0$. If $\alpha_{tki}(\tau) = 1$, $1 \leq \beta_{tki}(\tau) \leq ct_t$.

In Linux operating systems [15], each process is allocated to one of nt_t (≥ 1) threads $tr_{tk0}, \ldots, tr_{tk,nt_t-1}$ of a server s_t in the round-robin (RR) algorithm. The electric power consumption $CE_t(n)$ [W] of a server s_t to concurrently perform n processes at time τ is given in the MLPCM model as follows [17, 18, 19, 20]:

[MLPCM model for n **processes]** [Figure 2]

$$CE_t(n) = \begin{cases} minE_t \text{ if } n = 0. \\ minE_t + n \cdot (bE_t + cE_t + tE_t) \text{ if } 1 \leq n \leq np_t. \\ minE_t + np_t \cdot bE_t + n \cdot (cE_t + tE_t) \text{ if } np_t < n \leq nc_t \cdot np_t. \\ minE_t + np_t \cdot (bE_t + nc_t \cdot cE_t) + nt_t \cdot tE_t \text{ if } nc_t \cdot np_t < n < nt_t. \\ maxE_t \text{ if } n \geq nt_t. \end{cases} \tag{2}$$

The electric power consumption $E_t(\tau)$ [W] is assumed to be $CE_t(|\ CP_t(\tau)\ |)$ in this paper. The total electric energy $TE_t(st, et)$ [J] consumed by a server s_t from time st to time et is $TE_t(st, et) = \sum_{\tau=st}^{et} E_t(\tau)$.

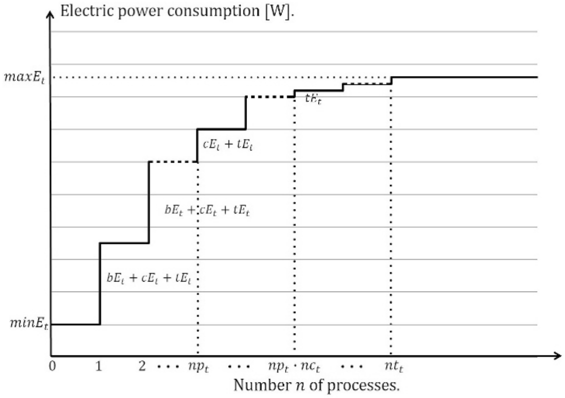


Fig. 2 MLPCM model.

3.2 MLCM Model

Each process is at a time performed on a thread of a server. It takes T_{ti} [sec] to perform a process p_i on a thread of a server s_t. A server s_f with fastest threads is referred to as *fastest* in a cluster S. If only a process p_i is performed on a server s_t without any other process, the execution time T_{ti} of the process p_i is minimum, i.e. T_{ti} = $minT_{ti}$. In a cluster S of servers $s_1, \ldots, s_m$ ($m \geq 1$), $minT_i$ shows a minimum one in a set $\{minT_{1i}, \ldots, minT_{mi}\}$ of minimum execution time. This means, $minT_i$ = $minT_{fi}$ on the fastest server s_f.

We assume one virtual computation step [vs] is performed on a fastest server s_f for one time unit [tu]. The maximum computation rate $maxCRT_f$ of the fastest server s_f is assumed to be one, $maxCRT_f$ = 1[vs/sec]. The number VS_i of total virtual computation steps of a process p_i is defined to be $minT_i$ [sec] $\cdot$ $maxCRT_f$ [vs/sec] = $minT_i$ [vs]. The maximum computation rate $maxCR_{ti}$ of a process p_i on a server s_t is VS_i / $minT_{ti}$ [vs/sec] ($\leq$ 1). On a thread of a server s_t, $maxCR_{ti}$ = $maxCRT_t$[vs/sec] for every process p_i.

The maximum computation rate $maxCR_t$ [vs/sec] ($\leq$ 1) of a server s_t is $np_t \cdot nc_t \cdot ct_t \cdot maxCRT_t = nt_t \cdot maxCRT_t$ where nt_t is the total number of threads. $CR_t(\tau)$ indicates the computation rate of a server s_t at time τ where $CR_t(\tau) \leq maxCR_t$. Suppose a process p_i is performed on a server s_t at time τ. The computation rate $CR_{ti}(\tau)$ of every process p_i on a server s_t is $CR_t(\tau)$ / $|CP_t(\tau)|$. The more number of processes are concurrently performed on a server s_t at time τ, the smaller computation rate $CR_{ti}(\tau)$ of each process p_i. The computation rate $CR_{ti}(\tau)$ of a process p_i on a server s_t at time τ is given in the MLCM (Multi-Level Computation with Multiple CPUs) model as follows [17, 18, 19, 20]:

[MLCM model] The computation rate $CR_{ti}(\tau)$ [vs/sec] of a process p_i on a server s_t at time τ is given as follows:

$$CR_{ti}(\tau) = \begin{cases} maxCR_t \, / \, |CP_t(\tau)| & \text{if } |CP_t(\tau)| > nt_t. \\ maxCRT_t & \text{if } |CP_t(\tau)| \leq nt_t. \end{cases} \tag{3}$$

Figure 3 shows the computation rate $CR_{ti}(\tau)$ of a process p_i on a server s_t which supports the number nt_t of threads. The computation rate $CR_{ti}(\tau)$ is constant (= $maxCRT_t$) if n (= $|CP_t(\tau)|$) $\leq nt_t$. For $n > nt_t$, $CR_{ti}(\tau)$ is $maxCR_t$ (= $nt_t \cdot maxCRT_t$) / n. The computation rate $maxCR_t$ is equally allocated to each current process p_i. Here, $\sum_{\tau=st}^{et} CR_{ti}(\tau)$ = VS_i [vs] = $maxCRT_f \cdot minT_i$ which shows the total amount of computation to be performed by a process p_i. At each time τ, a process p_i is performed at computation rate $CR_{ti}(\tau)$ on a server s_t. The computation laxity $lc_{ti}(\tau)$ [vs] is the number of virtual computation steps [vs] to be performed in the process p_i on a server s_t after time τ. At time τ a process p_i starts, $lc_{ti}(\tau)$ = VS_i. Then, at each time τ, the laxity $lc_{ti}(\tau)$ is decremented by $CR_{ti}(\tau)$, i.e. $lc_{ti}(\tau + 1)$ = $lc_{ti}(\tau)$ − $CR_{ti}(\tau)$. If $lc_{ti}(\tau+1) \leq 0$, the process p_i terminates at time τ.

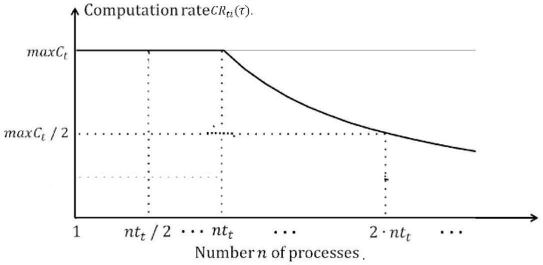


Fig. 3 Computation rate.

4 Energy-efficient Migration of Virtual Machines

4.1 Estimation Model

A client issues a request process p_i to a cluster S with a set VM = $\{VM_1, \dots, VM_v\}$ ($v \geq 1$) of virtual machines. Each virtual machine VM_h is on a host server s_t. First, one virtual machine VM_h is selected for performing the process p_i in the set VM. Then, the process p_i is performed on VM_h. Furthermore, the virtual machine VM_h can migrate from a host server s_t to a guest server s_u. Thus, we have to discuss how to select a virtual machine where a process p_i is to be performed and which virtual machine to migrate to which server in a cluster.

It is not easy to estimate the termination time of each current process p_i in every virtual machine VM_h on a server s_t by using the computation laxity $lc_{ti}(\tau)$ and the computation rate $CR_{ti}(\tau)$ of each process p_i as discussed [16, 17, 18, 19, 20, 21]. In this paper, a virtual machine is considered to be a unit of estimation of termination time of processes without considering each current process.

For a virtual machine VM_h on a server s_t, the *virtual machine* (VM) *laxity* $vlc_h(\tau)$ [vs] at time τ is defined as follows:

$$vlc_h(\tau) = \sum\nolimits_{p_i \in VCP_t(\tau)} VS_i \ /2. \tag{4}$$

In this paper, we assume each current process p_i ($\in VCP_t(\tau)$) on a virtual machine VM_h finishes the half of the total computation VS_i by time τ.

The *VM computation rate* $VCR_{ht}(\tau)$ [vs/sec] of a virtual machine VM_h on a host server s_t at time τ is defined as follows:

$$VCR_{ht}(\tau) = \begin{cases} maxCR_t \cdot \mid VCP_h(\tau) \mid / |CP_t(\tau)| & \text{if } |CP_t(\tau)| > nt_t. \\ maxCRT_t \cdot \mid VCP_h(\tau) \mid & \text{if } |CP_t(\tau)| \leq nt_t. \end{cases} \tag{5}$$

The larger number of processes are performed on a virtual machine VM_h, the larger VM computation rate $VCR_{ht}(\tau)$ is given to the virtual machine VM_h in a server s_t. The summation of VM_h computation rates of virtual machines on a server s_t is the computation rate $CR_t(\tau)$ of the server s_t, i.e. $\sum_{VM_h \in SVM_t(\tau)} vlc_h(\tau) = CR_t(\tau)$ at time τ.

The *server laxity* $slc_t(\tau)$ [vs] of a server s_t at time τ is the summation of the VM laxites of the virtual machines in the server s_t as follows:

$$slc_t(\tau) = \sum\nolimits_{VM_h \in SVM_t(\tau)} vlc_h(\tau). \tag{6}$$

This means, a server s_t is assumed to have the total amount $slc_t(\tau)$ of virtual computation steps to be performed at time τ.

We introduce the following function $ef_t\ (vl, n)$ for VM laxity vl and number n of processes on a server s_t:

$$ef_t(vl,\ n) = \begin{cases} vl/maxCR_t & \text{if } n > nt_t. \\ vl/(n \cdot maxCRT_t) & \text{if } n \leq nt_t. \end{cases} \tag{7}$$

The expected termination time ET_t of a server s_t is given at time τ by using the factor ef_t as follows:

$$ET_t = ef_t(slc_t(\tau), |CP_t(\tau)|). \tag{8}$$

The expected electric energy consumption EE_t [J] of a server s_t is given at time τ as follows:

$$EE_t = ET_t \cdot E_t(|CP_t(\tau)|). \tag{9}$$

4.2 VM Selection Algorithm

First, suppose a client issues a process p_i to a cluster S. A virtual machine VM_h is first selected for the process p_i by the following algorithm:

[VM selection]

1. Select a server s_t for a process p_i where the expected electric energy consumption $EE_t = ef_t(slc_t(\tau) + VS_i, |CP_t(\tau)| + 1) \cdot maxE_t$ is minimum at current time τ if a client issues the process p_i.

2. Select a virtual machine VM_h on the server s_t where the number $|VCP_h(\tau)|$ of processes is minimum.
3. Issue the process p_i to the selected virtual machine VM_h on the host server s_t.

If a process p_i is issued to a virtual machine on a server s_t, the server laxity $slc_t(\tau)$ is incremented by the amount VS_i of computation of the process p_i, i.e. $slc_t(\tau)$ + VS_i. The expected termination time ET_t to finish the process p_i and every current process in the set $CP_t(\tau)$ is given as $ef_t(slc_t(\tau) + VS_i, |CP_t(\tau)| + 1)$. Here, the expected electric energy EE_t consumed by a server s_t to perform the new process p_i in addition to every current process is $ef_t(slc_t(\tau) + VS_i, |CP_t(\tau)| + 1) \cdot CE_t(|CP_t(\tau)|$ + 1) [J]. A server s_t whose EE_t is minimum is first selected in the cluster S. Then, a virtual machine VM_h where the minimum number $|VCP_h(\tau)|$ of processes are performed is selected in the selected server s_t. The process p_i is issued to the selected virtual machine VM_h.

A server s_t is *overloaded* at time τ if $|CP_t(\tau)| \geq maxCN_t$. If a host server s_t is overloaded at time τ, one virtual machine VM_h is selected and migrates to another guest server s_u in the following VM migration algorithm:

[VM migration]

1. Let X be a set of overloaded servers in a cluster S.
2. If $X = \phi$, terminate; Select a server s_t whose expected electric energy consumption EE_t is largest in the set X.
3. Let V be a set of virtual machines in the server s_t.
4. Select a virtual machine VM_h where $|VCP_h(\tau)|$ is maximum in the set V.
5. Select a server s_u which is not overloaded even if VM_h migrates to the server s_u, i.e. $|CP_u(\tau)| + |VCP_h(\tau)| \leq maxCN_u$ and the expected electric energy consumption $EE_t = ef_u(slc_u(\tau) + vlc_h(\tau), |CP_u(\tau)| + |VCP_h(\tau)|) \cdot cE_u(|CP_u(\tau)| + |VCP_h(\tau)|)$ is minimum.
6. If not found, $V = V - \{VM_h\}$; If $V \neq \phi$, go to 4, else $X = X - \{s_u\}$ and go to 2.
7. Migrates the virtual machine VM_h from the server s_t to the selected server s_u.
8. If the server s_t is still overloaded, go to 4.

5 Evaluation

We consider a cluster S composed of four real servers DSLab4, DSLab, Sunny, and Atria (m = 4) and four virtual machines $VM_1, \ldots, VM_4$ (v = 4). The servers DSLab4 and DSLab are equipped with four and two Intel Xeon E5-2667 v2 CPUs, respectively. Sunny and Atria are equipped with an Intel Xeon E5-2620 CPU and an Intel Corei7-6700K CPU, respectively. The performance parameters like $maxCRT_t$ and electric energy parameters like $minE_t$ of each server s_t are shown in Table 1. Initially, one virtual machine is deployed on each server. There are n (> 0) processes $p_1, \ldots, p_n$. The starting time $stime_i$ of each process p_i is randomly taken from time

0 to *xtime* - 1. Here, *xtime* is 200 time unit [tu]. In fact, one time unit [tu] shows 100 [msec] [16]. The minimum execution time $minT_i$ is randomly taken from 5 to 10 [tu]. The parameters of each process p_i are shown in Table 2. We consider the random (RD), round robin (RR), and SVM algorithm. In the RD and RR algorithms, virtual machines do not migrate. The total electric energy consumption of the servers is smaller in the SVM algorithm than the others.

Figure 4 shows the total electric energy consumption of the servers in the cluster *S* for number *n* of processes. Total electric energy consumption of the servers can be reduced in the SVM algorithm.

Figure 5 shows the total active time (TAT) of the servers. The total active time (TAT) of the servers in the SVM algorithm is shorter than the RD and RR algorithms. This means, servers are less loaded in the SVM algorithm than the other algorithms.

Figure 6 shows the average execution time of the processes. The average execution time of the SVM algorithm is shorter than the other algorithms. Each process can be more efficiently performed by migrating virtual machines.

Table 1 Parameters of servers.

parameters	DSLab4	DSLab	Sunny	Atira
np_t	4	2	1	1
nc_t	8	8	6	4
nt_t	64	32	12	8
$maxCRT_t$ [vs/tu]	1	1.0	0.5	0.7
$maxCR_t$ [vs/tu]	64	32	6	5.6
$minE_t$ [W]	126.1	126.1	87.2	41.3
$maxE_t$ [W]	454.4	301.1	131.2	89.5
bE_t [W]	30	30	16	15
cE_t [W]	5.6	5.6	3.6	4.7
tE_t [W]	0.8	0.8	0.9	1.1

Table 2 Parameters of processes.

parameters	values
n	number of processes $p_1, \ldots, p_n$ (≥ 0)
$minT_i$ [tu]	minimum computation time of a process p_i
VS_i [vs]	$0.5 \sim 1.0$ $(VS_i = minT_i)$
st_i [tu]	starting time of p_i ($0 \leq st_i < xtime$ - 1)
$xtime$ [tu]	simulation time (= 200 (= 20[sec]))

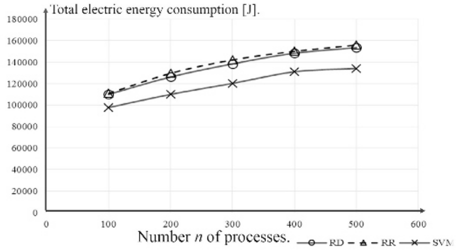


Fig. 4 Total electric energy consumption.

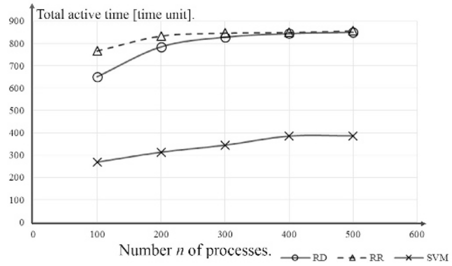


Fig. 5 Total active time.

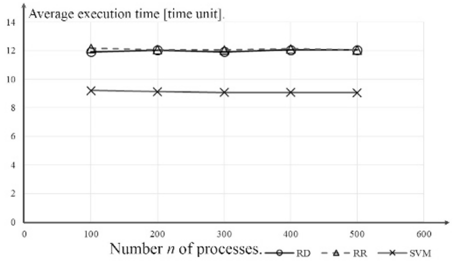


Fig. 6 Average execution time.

6 Concluding Remarks

It is critical to discuss how to reduce the electric energy consumption of servers to realize eco society. Here, virtual machines on a host server migrates to a guest server which is expected to consume smaller electric energy while processes are being performed on the virtual machines. In this paper, we newly proposed the SVM algorithm where each virtual machine is considered to be a unit of computation. It is easier to estimate the termination time of each virtual machine than the termination time of each process. In the evaluation, we showed the total electric energy consumption and active time of servers and the average execution time of processes can be reduced in the SVM algorithm compared with the random (RD) and round-robin (RR) algorithms. We are now evaluating the SVM algorithm in a scalable cluster where more number of processes are performed on more number of servers.

Acknowledgment

This work was supported by JSPS KAKENHI grant number 15H0295.

References

1. Duolikun, D., Aikebaier, A., Enokido, T., and Takizawa, M.: Energy-aware Passive Replication of Processes. Journal of Mobile Multimedia, **9** (1&2), pp. 53–65. (2013)
2. Duolikun, D., Aikebaier, A., Enokido, T., and Takizawa, M.: Power Consumption Models for Migrating Processes in a cluster, Proc. of International Conference on Complex, Intelligent, and Software Intensive Systems (NBiS-2014), pp.15–22. (2014).
3. Duolikun, D., Aikebaier, A., Enokido, T., and Takizawa, M.: Energy-efficient Dynamic Cluster of Servers, Journal of Supercomputing, **71**(5), pp.1647–1656. (2015).
4. Duolikun, D., Aikebaier, A., Enokido, T., and Takizawa, M.: Power Consumption Model for Redumdantly Performing Mobile-Agents, Proc. of the 8th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2014), pp.185–190. (2014).
5. Duolikun, D., Enokido, T., and Takizawa, M.: Asynchronous Migration of Process Replica in a Cluster, Proc. of IEEE the 29th International Conference on Advanced Information Networking and Applications (AINA-2015), pp.271–278. (2015).
6. Duolikun, D., Enokido, T., and Takizawa, M.: Asynchronous Migration of Process Replica in a Cluster, Proc. of the 9th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2015), pp.118–125. (2015).
7. Enokido, T., Aikebaier, A., and Takizawa, M.: A Model for Reducing Power Consumption in Peer-to-Peer Systems, IEEE Systems Journal, **4**(2), pp.221–229. (2010).
8. Enokido, T., Aikebaier, A., and Takizawa, M.: An Integrated Power Consumption Model for Communication and Transaction Based Applications, Proc. of IEEE the 25th International Conference on Advanced Information Networking and Applications (AINA-2011), pp.627–636. (2011).
9. Enokido, T., Aikebaier, A., and Takizawa, M.: Process Allocation Algorithms for Saving Power Consumption in Peer-to-Peer Systems, IEEE Transactions on Industrial Electronics, **58**(6), pp.2097–2105. (2011).
10. Enokido. T., Aikebaier. A., and Takizawa. M.: An Extended Simple Power Consumption Model for Selecting a Server to Perform Computation Type Processes in Digital Ecosystems, IEEE Transactions on Industrial Informatics, **10**(2), pp.1627–1636. (2014).
11. Enokido, T., Aikebaier, A., and Takizawa, M.: Evaluation of the Extended Improved Redundant Power Consumption Laxity-Based (EIRPCLB) Algorithm, Proc. of IEEE the 28th International Conference on Advanced Information Networking and Applications (AINA-2014), pp.940–947. (2014).
12. Enokido, T. and Takizawa, M.: Energy-Efficient Delay Time-Based Process Allocation Algorithm for Heterogeneous Server Clusters, Proc. of IEEE the 29th International Conference on Advanced Information Networking and Applications (AINA-2015), pp.279–286. (2015).
13. Enokido, T. and Takizawa, M.: Power Consumption and Computation Models of Virtual Machines to Perform Computation Type Application Processes, Proc. of the 9th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2015), pp.126–133. (2015).
14. Google, Google Green,http://www.google.com/green/, 2015.
15. Job Scheduling Algorithms in Linux Virtual Server,http://www.linuxvirtualserver.org/docs /scheduling.html, 2010.
16. Kataoka, H., Duolikun, D., Enokido, T., and Takizawa, M.: Power Consumption and Computation Models of a Server with a Multi-core CPU and Experiments, Proc. of IEEE the 29th

International Conference on Advanced Information Networking and Applications Workshops (AINA-2015), pp.217–223. (2015).

17. Kataoka, H., Duolikun, D., Enokido, T., and Takizawa, M.: Evaluation of Energy-aware Server Selection Algorithm, Proc. of the 9th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2015), pp.318–325. (2015).
18. Kataoka, H., Duolikun, D., Enokido, T., and Takizawa, M.: Multi-level Computation and Power Consumption Models, Proc. of the 18th International Conference on Network-Based Information Systems (NBiS -2015), pp.40–47. (2015).
19. Kataoka, H., Duolikun, D., Enokido, T., and Takizawa, M.: Energy-efficient Virtualisation of Threads in a Server Cluster, Proc. of the 10th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA-2015), pp.288–295. (2015).
20. Kataoka, H., Duolikun, D., Enokido, T., and Takizawa, M.: Energy-aware Server Selection Algorithm in a Scalable Cluster, Proc. of IEEE the 30th International Conference on Advanced Information Networking and Applications (AINA-2016), pp.565–572. (2016).
21. Sawada, A., Kataoka, H., Duolikun, D., Enokido, T., and Takizawa, M.: Energy-aware Clusters of Servers for Storage and Computation Applications, Proc. of IEEE the 30th International Conference on Advanced Information Networking and Applications (AINA-2016), pp.400–407. (2016).
22. United Nations Framework Convention on Climate Change (UNFCCC), https://en.wikipedia.org/wiki/Kyoto Protocol, 1992.
23. United Nations Climate Change Conference (COP21), https://en.wikipedia.org/wiki/2015, 2015.
24. A virtualization infrastructure for the Linux kernel (Kernel-based virtual machine), https://en.wikipedia.org/wiki/Kernel-based Virtual Machine.
25. An American company that provides cloud and virtualization software and services (VMware, Inc.), https://en.wikipedia.org/wiki/VMware.

Energy-aware Server Selection Algorithms for Storage and Computation Processes

Atsuhiro Sawada, Hiroki Kataoka, Dilawaer Duolikun, Tomoya Enokido, and Makoto Takizawa

Abstract Application processes like Web applications use not only CPU but also storages like HDD. In our previous studies, the algorithms to select a server in a cluster are proposed to energy-efficiently perform processes which use either CPU or storages. In this paper, we consider a more general type of process which does both the computation and accesses to storages. In this paper, we newly propose LEAG and GEAG algorithms to select servers to perform general processes in a cluster so that the total electric energy consumption of the servers can be reduced. We evaluate the LEAG and GEAG algorithms in terms of total electric energy consumption of the servers and average execution time of the processes. We show the electric energy consumed by servers can be reduced in the LEAG and GEAG algorithms.

1 Introduction

Information systems are now getting scalable like cloud computing systems [9] and Internet of Things (IoT) [10] and consume a plenty of electric energy. There are hardware-oriented approaches to realizing energy-efficient servers [1], [9], [13],

Atsuhiro Sawada
Hosei University, Tokyo, Japan e-mail: atsuhiro.sawada.7n@stu.hosei.ac.jp

Hiroki Kataoka
Hosei University, Tokyo, Japan e-mail: hiroki.kataoka.6v@stu.hosei.ac.jp

Dilawaer Duolikun
Hosei University, Tokyo, Japan e-mail: dilewerdolkun@gmail.com

Tomoya Enokido
Rissho University, Tokyo, Japan e-mail: eno@ris.ac.jp

Makoto Takizawa
Hosei University, Tokyo, Japan e-mail: makoto.takizawa@computer.org

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_5

[22]. On the other hand, in our macro-level approach [5], [6], [7], we aim at reducing the total amount of electric energy [J] consumed by a whole server to perform application processes without synthesizing the power consumption of each hardware component like CPU and memory. In this paper, a *process* means an application process to be performed on a server. Types of macro-level power consumption models [3]-[8], [12], [15]-[18] of a server are proposed. The MLPCM (Multi-Level Power Consumption with Multiple CPUs) model [19] is also proposed where the power consumption [W] of a server to concurrently perform computation processes depends on the number of active CPUs, cores, and threads where at least one process is performed. The MLPCMS (MLPCM for computation and Storage processes) model and the MLCS (Multi-Level Computation for computation and Storage processes) model are proposed to show the electric power to be consumed by a server and the execution time of each process to perform both computation and storage processes, respectively [23], [24]. In this paper, we propose an MLCMG (MLCM for General process) model which shows the execution time of a general process on a server. Based on the MLPCMS and MLCMG models, the LEAS (locally energy-aware server selection) [16] and GEAS (globally energy-aware server selection) [25] algorithm are proposed. In the LEAS algorithm, a server s_t is selected to perform a process issued by a client, whose expected electric energy to perform every current process and the process is minimum in a cluster. In the GEAS algorithm, a server s_t is selected where the total electric energy consumed by not only the selected server s_t but also all the other servers is minimum. However, each process is either a computation or storage type. In reality, processes like Web application processes use not only CPU but also access to storages like HDD. In this paper, we consider a more general type of process which both does the computation and accesses to storages. In this paper, we newly propose LEAG (LEA for General processes) and GEAG (GEA for General processes) algorithms by extending the LEAS and GEAS algorithms so that general processes can be handled. We evaluate the LEAG and GEAG algorithms in terms of the total electric energy consumption and active time of the servers compared with the round-robin (RR) and random (RD) algorithms. We show the total electric energy consumption and active time of the servers in a cluster can be reduced in the LEAG and GEAG algorithms.

In section 2, we discuss the power consumption and computation models of general processes. In section 3, we propose the LEAG and GEAG algorithms. In section 4, we evaluate the LEAG and GEAG algorithms.

2 Multi-level Power Consumption and Computation Models

2.1 MLPCMG Model

A cluster S is composed of m ($\geq$ 1) servers $s_1, \ldots, s_m$. Each server s_t is composed of np_t ($\geq$ 1) homogeneous CPUs $cp_{t0}, \ldots, cp_{t,np_t-1}$ [13]. Each CPU cp_{tk} is composed of the same number nc_t (≥ 1) of homogeneous cores $c_{tk0}, \ldots, c_{tk,nc_t-1}$. Each core c_{tkh} supports the same number ct_t (≤ 2) of threads. Let nt_t be the total number of threads supported by a server s_t, i.e. $nt_t = np_t \cdot nc_t \cdot ct_t$. A thread is *active* iff at least one process is performed. On an active server s_t, at least one thread is active. On an idle server, no thread is active.

First, we consider processes which use only CPU resource. Let $CP_t(\tau)$ be a set of processes which use CPU and are performed on a server s_t at time τ.

The electric power consumption $E_t(\tau)$ [W] of a server s_t to perform computation processes at time τ is given in the MLPCM (Multi-Level Power Consumption with Multiple CPUs) model [17] [18] [19] as follows:

[MLPCM model for computation processes]

$$E_t(\tau) = minE_t + \sum_{k=0}^{np_t-1} \{\gamma_{tk}(\tau)\ [bE_t + \sum_{h=0}^{nc_t-1} \alpha_{tkh}(\tau)(cE_t + \beta_{tkh}(\tau)\ tE_t)], \qquad (1)$$

where
$\gamma_{tk}(\tau)$ = 1 if at least one core is active on a CPU cp_{tk} at time τ, else 0,
$\alpha_{tkh}(\tau)$ = 1 if a core c_{tkh} is active on a CPU cp_{tk} at time τ, else 0,
$\beta_{tkh}(\tau)$ = number ($\leq ct_t$) of active threads on a core c_{tkh} at time τ.

In Linux operating systems, processes are allocated to threads of a server s_t in the round-robin (RR) algorithm [14]. After a process is allocated with a thread th_{ti}, a next coming process is allocated with a thread th_{tj} where $j = (i+1)$ modulo nt_t. $maxCE_t$ is maximum value of $E_t(\tau)$, $minE_t + np_t \cdot (bE_t + nC_t \cdot cE_t + ct_t \cdot tE_t)$. The electric power consumption $CE_t(n)$ [W] of a server s_t to concurrently perform n computation processes is given as follows [17] [18] [19]:

$$CE_t(n) = \begin{cases} minE_t\ if\ n = 0. \\ minE_t + n \cdot (bE_t + cE_t + tE_t) \text{ if } 1 \leq n \leq np_t. \\ minE_t + np_t \cdot bE_t + n \cdot (cE_t + tE_t) \text{ if } np_t < n \leq nc_t \cdot np_t. \\ minE_t + np_t \cdot (bE_t + nc_t\, cE_t) + nt_t \cdot tE_t \text{ if } nc_t \cdot np_t < n < nt_t. \\ maxCE_t \text{ if } n \geq nt_t. \end{cases} \qquad (2)$$

The electric power consumption $E_t(\tau)$ [W] of a server s_t for the number $|CP_t(\tau)|$ of processes at time τ is assumed to be $E_t(\tau) = CE_t(|\ CP_t(\tau)\ |)$ in this paper.

Next, we consider general processes which read data in a file while doing the computation. $RP_t(\tau)$ is a set of processes which read data in storages on a server s_t at time τ. $PP_t(\tau)$ is a set of processes performed on a server s_t at time τ, $PP_t(\tau) = CP_t(\tau) \cup RP_t(\tau)$. It is noted $CP_t(\tau) \cap RP_t(\tau)$ might be not empty. The electric power computation $E_t(\tau)$ of a server s_t to perform general processes at time τ is given in an MLPCMG (MPCM for General processes) model as follows;

[MLPCMG model] The power consumption $E_t(\tau)$ [W] of a server s_t to perform general processes at time τ is given as follows:

$$E_t(\tau) = CE_t(|CP_t(\tau)|) + \delta_t(\tau) \cdot RE_t \text{ where } \delta_t(\tau) = 1 \text{ if } |RP_t(\tau)| > 0, \text{else } 0. \quad (3)$$

The maximum electric power of a server s_t is $maxCE_t$ + RE_t. The electric energy [J] consumed by a server s_t from time st and et is $\sum_{x=st}^{et} E_t(\tau)$.

2.2 MLCMG Model

We newly propose an MLCMG (MLCM for General processes) model based on the MLCMS model [24], [25]. A computation process p_i is at a time performed on a thread of some server s_t in a cluster S. $minCT_{ti}$ shows the minimum execution time of a process p_i on a thread of a server s_t. That is, it takes $minCT_{ti}$ [sec] to perform a process p_i without any other process on a thread of a server s_t. Since each server s_t is assumed to support homogeneous CPUs, $minCT_{ti}$ is the same on any thread of a server s_t. $minCT_i$ indicates the minimum one of $minCT_{1i}$, $\cdots$, $minCT_{mi}$. A thread of a server s_f is fastest in a cluster S if $minCT_i = minCT_{fi}$. Here, the server s_f is referred to as $fastest$. That is, $minCT_i = minCT_{fi}$ for a process p_i. We assume one virtual computation step is performed on a fastest server s_f for one time unit. That is, $maxCRT$ = $maxCRT_f$ = 1 [vs/sec] for a fastest server s_f. A thread of a server s_t supports the computation rate $maxCRT_t = (minCT_i / minCT_{ti}) \cdot maxCRT = minCT_i / minCT_{ti}$ [vs/sec] $(\leq maxCRT)$. Then, VC_i shows the total number of virtual computation steps of a process p_i. VC_i is defined to be $minCT_i$ [sec] $\cdot$ $maxCRT$ [vs/ec] = $minCT_i$ [vs]. As discussed in papers [15]-[19], the maximum computation rate $maxCR_t$ of a server s_t is $nt_t \cdot maxCRT_t$ where nt_t is the number of threads supported by the server s_t.

We consider a general process which manipulates data in storages while doing the computation. For simplicity, we assume each process p_i only reads data in a file f_i different from every other process p_j $(f_i \neq f_j)$. Here, b_i shows the size [B] of a file f_i. $maxRR_t$ indicates the maximum read rate [B/sec] of a server s_t. The minimum read time $minRT_{ti}$ of a process p_i on a server s_t is b_i / $maxRR_t$ [sec].

In this paper, we take a fastest server s_f as a canonical server in a cluster S. The canonical read time cRT_i [sec] of a process p_i is defined to be $minRT_{fi}$ on the fastest server s_t. We assume one virtual computation step is performed on the fastest server s_f to read $maxRR_f$ [B] for one time unit [sec]. The canonical read rate cRR is defined to be one [vs/sec]. Hence, cRR [vs/sec] $\cdot$ cRT_i [sec] $(= cRT_i)$ virtual computation steps [vs] are performed to read data in a file f_i on the fastest server s_f. The amount VR_i of virtual computation steps of a process p_i to read data in a file f_i is cRR [vs/sec] $\cdot$ cRT_i [sec] $= minRT_{fi}$ [vs]. The maximum virtual read rate $maxVRR_t$ [vs/sec] of a server s_t is $cRR \cdot maxRR_t$ / $maxRR_f$ $= maxRR_t$ / $maxRR_f$.

$maxVRR_f = cRR$ $(= 1)$ for the fastest server s_f. Even if a server s_t is slower than the fastest server s_f, i.e. $maxCRT_t < maxCRT_f$, the maximum read rate $maxRR_t$ might be larger than $maxRR_f$. Here, $maxVRR_t > maxVRR_f$ $(= 1)$. Thus, it is noted $maxVRR_t$ might be larger than one.

We define the minimum execution time $minT_i$ of a process p_i to be $minCT_i + cRT_i$ [sec]. The amount VP_i of virtual computation of a process p_i is defined to be $VC_i + VR_i$ $(= minCT_i + minRT_{fi})$ [vs]. The computation ratio cr_i of a process p_i is VC_i / VP_i $(= VC_i + VR_i)$ and the read ratio rr_i is VR_i / VP_i. Here, $cr_i + rr_i = 1$.

[Virtual computation (VC) rates] The virtual computation rate $CR_{ti}(\tau)$ [vs/sec] and virtual read rate $RR_{ti}(\tau)$ [vs/sec] of a process p_i on a server s_t at time τ are given as follows :

$$CR_{ti}(\tau) = \begin{cases} maxCRT_t \cdot cr_i \text{ if } |CP_t(\tau)| \le nt_t. \\ maxCR_t \cdot cr_i / |CP_t(\tau)| \text{ if } |CP_t(\tau)| > nt_t. \end{cases} \tag{4}$$

$$RR_{ti}(\tau) = maxVRR_t \cdot rr_i / |RP_t(\tau)|. \tag{5}$$

The variables $vc_i(\tau)$ and $vr_i(\tau)$ show the computation laxity and read laxity of a process p_i, respectively, at time τ. At time τ a process p_i starts on a server s_t, $vc_i(\tau)$ $= VC_i$ and $vr_i(\tau) = VR_i$ [vs]. At each time τ, the variables $vc_i(\tau)$ and $vr_i(\tau)$ are decremented by the computation rate $CR_{ti}(\tau)$ and read rate$RR_{ti}(\tau)$ [vs/sec], respectively, as follows:

[Virtual computation (VC) model]

```
At each time $\tau$,
  for each server $s_t$ in a cluster $S$, {
    if there is a new process $p_i$ to be performed on $s_t$, {
        $PP_t(\tau) = PP_t(\tau) \cup \{p_i\}$;
        $vc_i(\tau) = VC_i$;   $vr_i(\tau) = VR_i$;
    }; /* if end */
    for each process $p_i$ in $PP_t(\tau)$, { /* laxity is decremented */
      $vc_i(\tau+1) = vc_i(\tau) - CR_{ti}(\tau)$;    $vr_i(\tau+1) = vr_i(\tau) - RR_{ti}(\tau)$;
      if $vc_i(\tau+1) \le 0$ and $vr_i(\tau+1) \le 0$, /* $p_i$ terminates */
        $PP_t(\tau+1) = PP_t(\tau) - \{p_i\}$;
      }; /* if end */
    }; /* for process $p_i$ end */
  }; /* for server $s_t$ end */
```

2.3 Estimation Model

We discuss how to estimate the electric energy consumption EE_t [J] of a server s_t to perform all the current processes and the termination time ET_t of every current process in the current process set $PP_t(\tau)$. We assume no process additionally starts on a server s_t after time τ. We calculate the expected electric energy consumption

EE_t [J] and expected termination time ET_t [sec] of a server s_t at time τ by the following estimation procedure **EST** (s_t, τ, $PP_t(\tau)$; EE_t, ET_t):

```
[Estimation algorithm]
EST ($s_t$, $\tau$, $PP$ ; $EE$, $ET$) {
input $s_t$ /* server */;
      $\tau$ /* current time */;
      $PP$ /* set of current processes on $s_t$ */;
output $EE$ /* electric energy to be consumed */;
       $ET$ /* termination time of all the processes */;
  $EE = 0$;  $x = \tau$;
  while ($PP \neq \phi$) {
  /* electric energy $EE$ */
    if $| VC_t(x) | \geq 1$ and $| VR_t(x) | = 0$ { /* only computation */
        $EE = EE + CE_t(| PP |)$;
    } else
    if $| VC_t(x) | = 0$ and $| VR_t(x) | > 0$ { /* only read */
        $EE = EE + RE_t$;
    } else
    if $| VC_t(x) | \geq 1$ and $| VR_t(x) | > 1$ { /* computation and read */
        $EE = EE + CE_t(| PP |) + RE_t$;
    }; /* if end */
    for each process $p_i$ in $PP$, {
        $vc_i(x+1) = vc_i(x) - CR_{ti}(x)$;    $vr_i(x+1) = vr_i(x) - RR_{ti}(x)$;
        if $vc_i(x+1) \leq 0$ and $vr_i(x+1) \leq 0$, /* $p_i$ terminates at time $x$ */
            $PP = PP - \{p_i\}$;
    }; /* for each process $p_i$ end */
    $x = x + 1$; /* time advances */
  }; /* while end */
  $ET = x - \tau$; /* every process terminates at $x - \tau$ */
};
```

Initially, x is current time τ and PP is a set $PP_t(\tau)$ of current processes on a server s_t. At each time x, the computation laxity $vc_i(\tau)$ and read laxity $vr_i(\tau)$ of each process p_i in the set PP are decremented by the computation rate $CR_{ti}(x)$ and read rate $RR_{ti}(x)$, which are given in formulas (4) and (5), respectively, as discussed here. If the computation laxity $vc_i(x+1)$ and read laxity $vr_i(x+1)$ are equal to or smaller than 0, a processes p_i terminates at time x and p_i is removed in the set PP. The electric energy consumption EE is incremented by the electric power consumption $E_t(x)$. If PP = ϕ, the estimation procedure **EST** terminates. Here, every current process of time τ terminates until time x - 1.

3 Server Selection Algorithms

3.1 LEAG Algorithm

One server s_t is selected for performing a process p_i in a cluster S of servers s_1, ..., s_m. If a process p_i is performed on a server s_t, the expected electric energy consumption EE_t and expected termination time ET_t of the server s_t, are obtained by the estimation procedure **EST** (s_t, τ, $PP_t(\tau) \cup \{p_i\}$; EE_t, ET_t). In an LEAG (Locally Energy-Aware for General processes) selection algorithm, a server s_t is selected in the cluster S, whose expected electric energy consumption EE_t to perform not only the new process p_i but also every current process in the current process set $PP_t(\tau)$ is minimum as follows:

[LEAG algorithm]

```
$EE = \infty$;
for each server $s_u$ in $S$ {
        EST($s_u$, $\tau$, $PP_u(\tau) \cup \{p_i\}$ ; $EE_u$, $ET_u$);
        if $EE > EE_u$, { $EE = EE_u$;    $s_t = s_u$; };
}; /* for end */
```

The process p_i is issued to the server s_t selected in the LEAG algorithm. Every server is checked for each process in a cluster $S = \{s_1, \ldots, s_m\}$. Hence, the computation complexity of the LEAG algorithm is $O(m)$.

3.2 GEAG Algorithm

In the LEAG algorithm, a server s_t is selected to perform a process p_i in a cluster S, whose expected electric energy to be consumed to perform the new process p_i and every current process is minimum. As discussed in paper [17], another server s_u where the process p_i is not performed also consumes the electric energy. For example, even an idle server s_u just consumes the minimum electric power $minE_u$. We propose a *GEAG* (*Globally Energy-Aware for General processes*) server selection algorithm where we take into account the electric energy consumption of not only a server s_t where a process p_i is to be performed but also the other servers. For a process p_i, a server s_t is selected in the GEAG algorithm as follows:

[GEAG algorithm]

```
    for   each server $s_t$ in $S$, {
/* $NE_t$ is the electric energy consumption to perform not only every current process
but also $p_i$ on $s_t$ */
        EST ($s_t$, $\tau$, $PP_t(\tau) \cup \{p_i\}$ ; $NE_t$, $NET_t$);
/* $EE_t$ is the electric energy consumption to perform every current process on $s_t$*/
        EST ($s_t$, $\tau$, $PP_t(\tau)$ ; $EE_t$, $ET_t$);
    }; /* for end */
    $XET = \mathbf{max}\{NET_1, \ldots, NET_m\}$; /* maximum termination time */
    for each server $s_t$ in $S$, {
        if $ET_t = 0$, $EE_t = minE_t \cdot XET$; /* $s_t$ is idle */
        else if $ET_t < XET$, $EE_t = EE_t + minE_t \cdot (XET - ET_t)$;
    } /* for end */
    $E = \infty$;
    for   each server $s_u$ in $S$ {
/* $GE_u$ is the total electric energy of $s_u$ to perform both $p_i$ and the other processes */
        $GE_u = NE_u + \sum_{s_v \in S - \{s_u\}} EE_v$;
        if $GE_u < E$, $\{E = GE_u;\quad s_t = s_u;\}$
    }; /* for end */
```

First, we obtain the expected electric energy NE_u of each server s_u to perform not only every current process in the process set $PP_u(\tau)$ but also a new process p_i. We also obtain the expected termination time NET_u. XET is the longest one of NET_1, …, NET_m. In addition, we obtain the expected electric energy EE_u of each server s_u to perform only every current process. In an idle server s_t, i.e. $ET_t = 0$, EE_t is $minE_t \cdot XET$. If every current process terminates on a server s_t before XET, i.e. $ET_t < XET$, the server s_t consumes the electric energy $minE_t \cdot (XET - ET_t)$ after every current process terminates. Hence, $EE_t = EE_t + minE_t \cdot (XET - ET_t)$. If a process p_i is to be performed on a server s_t, the cluster S is expected to totally consume the electric energy $GE_t = EE_1 + \cdots + EE_{t-1} + NE_t + EE_{t+1} + \cdots + EE_m$. In the GEAG algorithm, a server s_t where the expected total electric energy consumption GE_t is minimum is selected for a process p_i. The process p_i is performed on the selected server s_t.

In the GEAG algorithm, m servers s_1, …, s_m are twice searched in a cluster S. Hence, the computation complexity of the GEAG algorithm is larger than the LEAG algorithm but $O(m)$ for number m of servers in a cluster S.

4 Evaluation

Each server s_t is characterized by the parameters shown in Tables 1 and 2. The parameters of servers and processes are randomly taken for each server s_t. In the evaluation, a cluster S is composed of four servers ($m = 4$). The maximum computation rate $maxCRT_t$ of a thread of a server s_t is randomly taken from 0.5 to 1 [vs / time unit (tu)]. In reality, one time unit [tu] shows 100 [msec] in the evaluation. The maximum computation rate $maxCR_t$ of a server s_t is $nt_t \cdot maxCRT_t$ where nt_t is the total number of threads. The maximum virtual read rate $maxVRR_t$ of a server s_t is randomly taken from mR (= 0.5) to xR (= 1.5) [vs / tu]. $maxVRR_f = 1$ for a fastest server s_f. The computation rate $CR_{ti}(\tau)$ and read rate $RR_{ti}(\tau)$ of a process p_i on a server s_t are calculated depending on the numbers $|CP_t(\tau)|$ and $|RP_t(\tau)|$ of processes performed at timeτ, respectively, as discussed in this paper.

The number n of processes p_1, …, p_n are performed on the servers s_1, …, s_m. A process p_i starts at time st_i with the computation laxity $vc_i = VC_i$ and storage laxity $vr_i = VR_i$. The start time st_i is randomly taken from 0 to *xtime* - 1 [tu]. *xtime* is simulation time. In the evaluation, *xtime* is 200 [tu], i.e. 20 [sec]. The minimum computation time $minCT_i$ of each process p_i is randomly taken from mT (= 5) to xT (= 10) [tu], i.e. mT = 0.5 and xT = 1 [sec]. The total number VC_i [vs] of virtual computation steps of each process p_i is $minCT_i$. The canonical read time cRT_i is $\alpha \cdot minCT_i$ where $\alpha \geq 0$. In the evaluation, α = 0.5. The minimum execution time $minT_i$ of a process p_i is $minCT_i + cRT_i = (1 + \alpha) \cdot minCT_i = 1.5 \cdot minCT_i$ [tu]. The total amount VR_i of virtual read steps is $cRT_i \cdot cRR = cRT_i$. When a process p_i starts at time τ, the computation laxity vc_i is VS_i (= $minCT_i$) and the read laxity vr_i is VR_i (= cRT_i). At each time τ, the computation laxity vc_i and storage laxity vr_i are decremented by the computation rate $CR_{ti}(\tau)$ and read rate $RR_{ti}(\tau)$, respectively, as discussed in the preceding section.

Four algorithms, random (RD), round-robin (RR), LEAG, and GEAG algorithms are performed on the same pair of server and process configurations. Then, the electric energy consumption ET_t and active time AT_t of each server s_t and the execution time ET_i of each process p_i are obtained. The total active time AT_t of a server s_t shows time when the server is active. In the RD algorithm, a server s_t is randomly selected for each process p_i in a cluster S. In the RR algorithm, after a server s_t is selected for a previous process, a server s_{t+1} is selected for a next process. In the LEAG algorithm, one server s_t whose expected electric energy to perform a process p_i is minimum. In the GEAG algorithm, one server s_t where the total electric energy of all the servers $s_1, \ldots, s_m$ is minimum is selected. et_i is termination time of each process p_i, when the computation and read laxities get $vc_i \leq 0$ and $vr_i \leq 0$. *etime* is a maximum one of $et_1, \cdots, et_n$. The simulation ends at time *etime*.

Figure 1 shows the ratio of the total electric energy (TEE) consumption [J] of the m (= 4) servers $s_1, \ldots, s_m$ for number n of processes in each algorithm. TEE is $EE_1 + \cdots + ET_m$. TEE of the LEAG algorithm is minimum in the algorithms. TEEs of the LEAG and GEAG algorithms are almost invariant for the number n ($\leq$ 200) of processes. TEE of the GEAG algorithm is 30% to 40% smaller than the RD and RR algorithms and 10% larger than the LEAG algorithm.

The AT ratio is $\sum_{t=1}^{m} AT_t$ / (*etime* $\cdot$ m). Figure 2 shows the AT ratio of the servers. Here, the AT ratio of the GEAG algorithm is shown to be shorter than the RR and RD algorithms. This means, the servers are less loaded in the GEAG algorithm than the other algorithms.

5 Concluding Remarks

In this paper, we considered a general type of a process which both does the computation and accesses to storages like Web application process. Based on the MLPCMS and MLCMS models [24], [25], we newly proposed the MLPCMG and MLCMG models to perform general processes on a server. By using the MLPCMG and MLCMG models, we proposed the LEAG and GEAG algorithms by extending the LEAS [16] and GEAS [25] algorithms to select a server for a general process, which

Table 1 Parameters of a server s_t.

parameters	values
m	number of servers $s_1, \ldots, s_m$ (≥ 1)
np_t	number of CPUs (≤ 2)
nc_t	number of cores (1,2,4,6,8) in a server s_t
ct_t	threads/core (≤ 2)
nt_t	number of threads ($= ct_t \cdot np_t \cdot nc_t$)
$maxCRT_t$ [vs/tu]	$0.5 \sim 1$ [vs/tu]
$maxCR_t$ [vs/tu]	$nt_t \cdot maxCR_t$ [vs/tu]
$maxVRR_t$	$0.5 \sim 1.5$ [vs/tu]
$minE_t$ [W]	$40 \sim 90$ [W]
$maxCE_t$ [W]	$100 \sim 160$ [W]
bE_t [W]	$(maxCE_t - minE_t) / (4 \cdot np_t)$ [W]
cE_t [W]	$5 \cdot (maxCE_t - minE_t) / (8 \cdot np_t \cdot nc_t)$ [W]
tE_t [W]	$(maxCE_t - minE_t) / (8 \cdot nt_t)$ [W]
RE_t [W]	$(maxCE_t - minE_t)$ [W]
$maxE_t[W]$	$maxCE_t + RE_t$

Table 2 Parameters of a process p_i.

parameters	values
n	number of processes $p_1, \ldots, p_n$
α	computation - read ratio factor ($\alpha = 0.5$)
$minCT_i$ [tu]	minimum computation time ($5 \sim 10$)
cRT_i [tu]	canonical read time ($\alpha \cdot minCT_i$)
$minT_i$	$minCT_i + cRT_i = (1 + \alpha) \cdot minCT_i$
VC_i [vs]	$minCT_i$ [vs]
VR_i [vs]	$\alpha \cdot minCT_i$ [vs]
$st_i[tu]$	starting time of p_i ($0 \leq st_i < xtime - 1$) [tu]
$xtime[tu]$	simulation time (= 200) [tu]

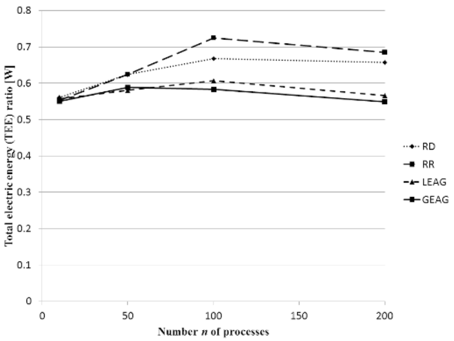


Figure 1 Total electric energy (TEE) (α = 0.5, m = 4).

is expected to consume the minimum electric energy. We evaluated the LEAG and GEAG algorithms in terms of the total electric energy consumption and active time of servers and the average execution time of processes compared with the RD and RR algorithms. We showed the total electric energy consumption and active time of

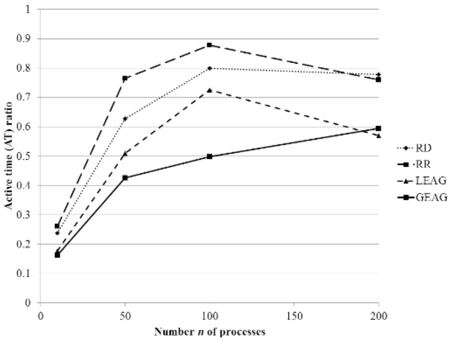


Figure 2 Active time (AT) ratio ($\alpha = 0.5$, $m = 4$).

servers and the average execution time of processes can be reduced in the LEAG and GEAG algorithms compared with the RD and RR algorithms. The computation complexity of the LEAG and GEAG algorithms is $O(m)$ for number m of servers in a cluster.

Acknowledgment

This work was supported by JSPS KAKENHI grant number 15H0295.

References

1. Bianchini, R. and Rajamony, R.: Power and Energy Management for Server Systems, IEEE Computer, **37**(11), pp. 68-74, (2004).
2. Duolikun, D., Aikebaier, A., Enokido, T., and Takizawa, M.: Energy-aware Passive Replication of Processes, Journal of Mobile Multimedia, **9**(1&2), pp.53-65, (2013).
3. Duolikun, D., Aikebaier, A., Enokido, T., and Takizawa, M.: Energy-Efficient Dynamic Clusters of Servers, Journal. of Supercomputing, **71**(5), pp.1642–1656, (2015).
4. Enokido, T., Aikebaier, A., and Takizawa, M.: A Model for Reducing Power Consumption in Peer-to-Peer Systems, IEEE Systems Journal, **4**(2), pp.221-229, (2010).
5. Enokido, T., Aikebaier, A., and Takizawa, M.: Process Allocation Algorithms for Saving Power Consumption in Peer-to-Peer Systems, IEEE Transactions on Industrial Electronics, **58**(6), pp.2097–2105, (2011).
6. Enokido, T., Aikebaier, A., and Takizawa, M.: An Extended Simple Power Consumption Model for Selecting a Server to Perform Computation Type Processes in Digital Ecosystems, IEEE Transactions on Industrial Informatics, **10**(2), pp.1627-1636, (2014).
7. Enokido, T., Aikebaier, A., and Takizawa, M.: An Integrated Power Consumption Model for Communication and Transaction Based Applications, Proc. of IEEE the 25th International Conference on Advanced Information Networking and Applications (AINA-2011), pp. 627-636, (2011).
8. Enokido, T., Aikebaier, A., and Takizawa, M.: Evaluation of the Extended Improved Redundant Power Consumption Laxity-Based (EIRPCLB) Algorithm, Proc. of IEEE the 28th Inter-

national Conference on Advanced Information Networking and Applications (AINA-2014), pp.940–947, (2014).

9. Ghemawat, S., Gobioff, H., and Leung, S.-T.: The Google Files System, Proc. of the 19th ACM Symposium on Operating Systems Principles, (SOSP'03), pp.29-43, (2003).
10. Hoeller, J., Tsiatsis, V., Mulligan, C., Karnouskos, S., Avesand, S., and Boyle, D.: From Machine-to-Machine to the Internet of Things: Introduction to a New Age of Intelligence. Elsevier, 352 pages, (2014).
11. HP server: http://h50146.www5.hp.com/products/servers /proliant/system_pdf/dl360pgen8.pdf.
12. Inoue, T., Aikebaier, A., Enokido, T., and Takizawa, M.: Power Consumption and Processing Models of Servers in Computation and Storage Based Applications, Journal of Mathematical and Computer Modeling, **58**(5&6), pp.1475-1488, (2013).
13. Intel Xeon Processor 5600 Series: The Next Generation of Intelligent Server Processors, white paper [online]. Available: http://www.intel.com/content/www/us/en/processors/xeon/xeon-5600-brief.html, (2010).
14. Job Scheduling Algorithms in Linux Virtual Server, http://www.linuxvirtualserver.org/docs/scheduling.html, (2010).
15. Kataoka, H., Duolikun, D., Enokido, T., and Takizawa, M.: Power Consumption and Computation Models of a Server with a Multi-core CPU and Experiments, Proc. of IEEE the 29th International Conference on Advanced Information Networking and Applications Workshops (WAINA-2015), pp.318-325, (2015).
16. Kataoka, H., Duolikun, D., Enokido, T., and Takizawa, M.: Evaluation of Energy-Aware Server Selection Algorithms, Proc. of the 9th international Conference on International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2015), pp.318-326, (2015).
17. Kataoka, H., Duolikun, D., Enokido, T., and Takizawa, M.: Multi-level Computation and Power Consumption Models. Proc. of the 18th International Conference on Network-Based Information Systems (NBiS-2015), pp.40-47, (2015).
18. Kataoka, H., Duolikun, D., Enokido, T., and Takizawa, M.: Energy-efficient Virtualisation of Threads in a Server Cluster, Proc. of the 10th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA-2015), pp.288-295, (2015).
19. Kataoka, H., Duolikun, D., Enokido, T., and Takizawa, M.: Energy-aware Server Selection Algorithms in a Scalable Cluster, Proc. of IEEE the 30th International Conference on Advanced Information Networking and Applications (AINA-2016), pp.565-572, (2016).
20. Linux distribution. Available: http://itpro.nikkeibp.co.jp/article/COLUMN/20120223/382669/.
21. Metaprotocol Corp: "UWmeter" Online'. Available: http://www.metaprotocol.com/UWmeter/Feautures.html, 2011.
22. Natural Resources Defense Council (NRDS): Data Center Efficiency Assessment - Scaling up Energy Efficiency across the Data Center Industry: Evaluating Key Drivers and Barriers. http://www.nrdc.org/energy/files/data-center-efficiency-assessment-IP.pdf, (2014).
23. Sawada, A., Nakamura, S., Enokido, T., and Takizawa, M.: Eco Model of Storage-based Servers. Proc. of the NBiS Workshops, pp.407-411, (2015).
24. Sawada, A., Kataoka, H., Duolikun, D., Enokido, T., and Takizawa, M.: Energy-aware Clusters of Servers for Storage and Computation Applications. Proc. of IEEE the 30th International Conference on Advanced Information Networking and Applications (AINA-2016), pp.400-407, (2016).
25. Sawada, A., Kataoka, H., Duolikun, D., Enokido, T., and Takizawa, M.: Energy-aware Clusters of Servers for Storage and Computation Applications. Proc. of the 10th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS-2016), pp.162-169, (2016).

Topic-based Synchronization (TBS) Protocols to Prevent Illegal Information Flow in Peer-to-Peer Publish/Subscribe Systems

Shigenari Nakamura, Tomoya Enokido, and Makoto Takizawa

Abstract In a peer-to-peer type of topic-based publish/subscribe (P2PPS) model, each peer (process) can publish an event message and receive an event message in which the peer is interested. Subscription of a peer and publication of an event message are specified in terms of topics. In the topic-based access control (TBAC) model proposed in our previous studies, only a peer granted publication and subscription rights is allowed to publish event messages with publication topics and to subscribe events, respectively. In our previous studies, the subscription-based synchronization (SBS) and subscription initialization SBS (SI-SBS) protocols are proposed where notifications which may cause illegal information flow are banned to prevent illegal information flow. It is checked whether or not an illegal information flow to occur in terms of subscription and publication rights granted to each peer. However, even some legal notifications are banned while no illegal event message is notified. In this paper, we newly propose a topic-based synchronization (TBS) and subscription initialization TBS (SI-TBS) protocols where only topics which each peer manipulates are considered. We show the number of notifications banned is reduced in the TBS and SI-TBS protocols compared with the SBS and SI-SBS protocols in the evaluation.

1 Introduction

A distributed system is composed of processes which are cooperating with one another by exchanging messages in networks. In distributed systems, informa-

Shigenari Nakamura
Hosei University, Tokyo, Japan, e-mail: nakamura.shigenari@gmail.com

Tomoya Enokido
Rissho University, Tokyo, Japan, e-mail: eno@ris.ac.jp

Makoto Takizawa
Hosei University, Tokyo, Japan, e-mail: makoto.takizawa@computer.org

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_6

tion in objects flow to other objects by transactions' manipulating the objects. In order to prevent illegal information flow, types of synchronization protocols [6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17] are discussed based on the role-based access control (RBAC) model [4]. On the other hand, context-based systems like publish/subscribe (PS) systems [1, 3, 5, 23, 24] are getting important in various applications. In this paper, we consider a peer-to-peer (P2P) model of topic-based PS system [22] (P2PPS model) [20, 21] where each peer (process) can both publish and subscribe event messages.

The *topic-based access control* (TBAC) model in PS systems is proposed [18]. Here, a peer manipulates topics, not objects, in publish (*pb*) and subscribe (*sb*) operations. An access rule $\langle p_i, op, t \rangle$ means that a peer p_i is allowed to manipulate a topic t in an operation op. $p_i.P$ shows the publication of a peer p_i, which is a subset of topics granted to a peer p_i. $p_i.S$ indicates the subscription of a peer p_i which is also a subset of topics which the peer p_i is allowed to subscribe. An event message e published by a peer p_i is notified to a target peer p_j if the subscription $p_j.S$ includes at least one common topic with the publication $e.P$. Here, the event message e is related with *forgotten* topics which are in the publication $e.P$ but not in the subscription $p_j.S$. In addition, the peer p_i may publish the event message e_2 after receiving another event message e_1. The event message e_1 may be related with *hidden* topics in the subscription $p_i.S$ but not in the publication of e_2. This means, the event message e_1 may bring events related with the hidden topics to the target peer p_j. Hidden or forgotten topics of an event message for a peer are *implicit* topics. Thus, the target peer p_j receives an event message which is related with topics which the peer p_j is not allowed to subscribe. Here, the peer p_i *illegally flows* to the peer p_j. The legal information flow relation among the peers is defined based on the TBAC model [18].

In our previous studies [18, 19], the subscription-based synchronization (SBS) and subscription initialization SBS (SI-SBS) protocols are proposed based on the TBAC model to prevent illegal information flow. Here, the notification of an event message which may cause illegal information flow are banned at each target peer. It is checked whether or not the notification may cause illegal information flow in terms of subscription and publication rights of each peer. In reality, each peer manipulates only some, not necessarily all topics in the access rights of each peer. Here, some notifications which are not illegal may be banned.

In this paper, we newly propose *topic-based synchronization* (*TBS*) and *subscription initialization TBS* (*SI-TBS*) protocols based on the TBAC model to reduce the number of banned notifications of event messages. Here, notifications which may cause illegal information flow are banned as well as the SBS and SI-SBS protocols. However, it is checked whether or not the notification may cause illegal information flow in terms of only topics which each peer manipulates. We evaluate the TBS and SI-TBS protocols in terms of number of notifications banned compared with the SBS and SI-SBS protocols. We show the number of notifications banned is reduced in the TBS and SI-TBS protocols.

In section 2, we present the information flow relation among peers in the TBAC model. In section 3, we newly propose the TBS and SI-TBS protocols to prevent illegal information flow to occur. In section 4, we evaluate the TBS and SI-TBS protocols.

2 Information Flow in TBAC Model

2.1 TBAC Model

In this paper, we consider a peer-to-peer (P2P) model of a publish/subscribe (PS) system [1, 3, 5, 23, 24] (P2PPS model) [20, 21] which includes a set P of peer processes (peers) $p_1, \ldots, p_{pn}$ ($pn \geq 1$). Each peer p_i can play both publisher and subscriber roles in the P2PPS model while only a publisher process publishes event messages and a subscriber process just receives event messages in the PS model. In this paper, we consider a topic-based PS system [22]. Let T be a set $\{t_1, \ldots, t_{tn}\}$ ($tn \geq 1$) of all topics in a system. A peer p_i publishes an event message e with publication $e.P$ ($\subseteq T$). A peer p_i specifies the subscription $p_i.S$ in a subset of topics ($p_i.S \subseteq T$). An event message e is notified to a peer p_i if the publication $e.P$ and the subscription $p_i.S$ include at least one common topic, i.e. $e.P \cap p_i.S \neq \phi$.

In the *topic-based access control* (TBAC) model [18], an access rule $\langle p_i, t, op\rangle$ means that a peer p_i is allowed to manipulate a topic t in an operation op. Here, an operation op is a subscribe (sb) or publish (pb). Let A be a set of access rules authorized in the system. An access right is specified in a pair $\langle t, op\rangle$ of a topic t and an operation op. A peer p_i is granted an access right $\langle t, op\rangle$ where t is a topic ($t \in T$) and op is an operation ($op \in \{pb$ (publish), sb (subscribe)$\}$). A peer p_i is allowed to publish an event message e with publication $e.P$ ($\subseteq T$) only if the peer p_i is granted a publication right $\langle t, pb\rangle$ for every topic t in the publication $e.P$. The publication $p_i.P$ ($\subseteq T$) of a peer p_i is a subset $\{t \mid \langle p_i, t, pb\rangle \in A\}$ of topics on which the peer p_i is allowed to publish an event message. A peer p_i publishes an event message e with publication $e.P$ which is a subset of the publication $p_i.P$. Here, topics in the publication $p_i.P$ which are not in the publication $e.P$, i.e. $\{t \mid t \in p_i.P$ but $t \notin e.P\}$ are *hidden* topics $e.H$ of the event message e. A hidden topic t is a topic which may be related with an event message e but which is not specified in the event message e. Here, even if a peer p_i receives an event message e, the peer p_i does not recognize the event message e to be related with hidden topics of the event message e.

A peer p_i is allowed to subscribe a topic t only if a subscription right $\langle t, sb\rangle$ is granted to the peer p_i. The subscription $p_i.S$ ($\subseteq T$) of a peer p_i is a subset of topics on which a peer p_i is allowed to receive event messages, i.e. $\{t \mid \langle p_i, t, sb\rangle \in A\}$.

2.2 Information Flow Relations

Suppose a peer p_i publishes an event message e with a publication $e.P$ ($\subseteq p_i.P$). The subscription $p_j.S$ of a peer p_j shows topics in which the peer p_j is interested. That is, a peer p_j can receive an event message e if $p_j.S \cap e.P \neq \phi$. A peer p_j is a *target* peer of an event message e if and only if (iff) $e.P \cap p_j.S \neq \phi$, i.e. the subscription $p_j.S$ of a peer p_j has a common topic with the publication $e.P$ of an event message e. Let $e.H$ be a set of *hidden* topics of an event message e with respect to the target peer p_j. Hidden topics may be related with an event message e but are not included in the publication $e.P$, i.e. $\{t \mid t \in p_i.S$ but $t \notin e.P\}$. Here, even if a target peer p_j receives an event message e, the peer p_j does not recognize the event message e may be related with the hidden topics. Topics which are in the publication $e.P$ but not in the subscription $p_j.S$, i.e. $\{t \mid t \in e.P$ but $t \notin p_j.S\}$, are *forgotten* topics $e.F$ ($= e.P - p_j.S$) of the event message e with respect to the target peer p_j. A target peer p_j recognizes an event message e to be related with topics in $e.P \cap p_j.S$ but forgets the event message e is related with the topics in $e.F$.

If an event message e is notified to a target peer p_i, the event message e is related with not only topics $p_i.S$ which the peer p_i subscribes and forgotten topics $e.F$ in the publication $e.P$ but also hidden topics $e.H$ which the event message e does not bring to the peer p_i. *Implicit* topics of a peer p_i are hidden or forgotten topics of event messages which the peer p_i receives. Let $p_i.I$ indicate a set of implicit topics of a peer p_i. $p_i.I$ is manipulated as follows:

1. $p_i.I$ is initially ϕ.
2. Each time a peer p_i receives an event message e, $p_i.I = p_i.I \cup (e.H - p_i.S) \cup e.F$;

First, the information flow relation ($\rightarrow$) among peers is defined as follows [18, 19]:

[Definition] A peer p_i *flows* to a peer p_j ($p_i \rightarrow p_j$) iff (if and only if) $p_i.P \cap p_j.S \neq \phi$.

The information flow relation $p_i \rightarrow p_j$ means an event message published by a peer p_i is allowed to be notified to a peer p_j. A pair of different peers p_i and p_j are *equivalent* ($p_i \leftrightarrow p_j$) iff $p_i \rightarrow p_j$ and $p_j \rightarrow p_i$. A peer p_i is *compatible* with a peer p_j ($p_i \rightharpoonup p_j$) iff the peer p_i does not flow to the peer p_j, i.e. $p_i \not\rightarrow p_j$. There is no information flow relation among the peers p_i and p_j if $p_i \rightharpoonup p_j$. A pair of peers p_i and p_j are compatible with each other ($p_i \rightleftharpoons p_j$) iff $p_i \rightharpoonup p_j$ and $p_j \rightharpoonup p_i$.

[Definition] [18, 19]

1. A peer p_i *legally flows* to a peer p_j ($p_i \Rightarrow p_j$) iff one of the following conditions holds:
   a. $p_i.S \neq \phi$, $p_i \rightarrow p_j$, and $p_i.S \subseteq p_j.S$.
   b. For some peer p_k, $p_i \Rightarrow p_k$ and $p_k \Rightarrow p_j$.
2. A pair of peers p_i and p_j are *legally equivalent* with one another ($p_i \Leftrightarrow p_j$) iff $p_i \Rightarrow p_j$ and $p_j \Rightarrow p_i$.

3. A peer p_i *illegally flows* to a peer p_j ($p_i \mapsto p_j$) iff $p_i \rightarrow p_j$ but $p_i \not\Rightarrow p_j$.

The legal information flow relation $\Rightarrow$ is transitive but not symmetric. If a peer p_i flows to a peer p_j ($p_i \rightarrow p_j$), i.e. $p_i.P \cap p_j.S \neq \phi$, an event message published by the peer p_i can be notified to the peer p_j. Otherwise, no information from the peer p_i flow into the peer p_j. The condition $p_i.S \subseteq p_j.S$ means that every topic in the subscription $p_i.S$ is also in the subscription $p_j.S$. This means, an event message e from the peer p_i to the peer p_j has no hidden topic for the peer p_j, i.e. $e.H = \phi$. It is noted $p_i.S = p_j.S$ if $p_i \Leftrightarrow p_j$. For a pair of peers p_i and p_j, if $p_i \Rightarrow p_j$ and $p_i.S \neq p_j.S$, $p_j \Rightarrow p_i$ does not hold. This means, the legal information flow relation $\Rightarrow$ is acyclic.

On the other hand, the illegal flow relation $\mapsto$ is not transitive, differently from the transitive legal flow relation $\Rightarrow$. Even if $p_i \mapsto p_j$ and $p_j \mapsto p_k$, $p_i \Rightarrow p_k$ may hold.

Suppose there are three peers p_i, p_j, and p_k in a system. We also suppose a peer p_i is granted a pair of access rights $\langle y, pb \rangle$ and $\langle x, sb \rangle$, i.e. the publication $p_i.P$ ($= \{y\}$) and subscription $p_i.S$ ($= \{x\}$), another peer p_j is granted access rights $\langle x, pb \rangle$, $\langle x, sb \rangle$, and $\langle y, sb \rangle$, i.e. $p_j.P$ ($= \{x\}$) and $p_j.S$ ($= \{x, y\}$), and the other peer p_k is granted access rights $\langle z, pb \rangle$, $\langle x, sb \rangle$, and $\langle z, sb \rangle$, i.e. $p_k.P$ ($= \{z\}$) and $p_k.S$ ($= \{x, z\}$). First, the peer p_i publishes an event message e_1 with publication $e_1.P = \{y\}$ ($\subseteq p_i.P$). Here, the peer p_i flows to the peer p_j ($p_i \rightarrow p_j$) since $p_i.P$ ($= \{y\}$) $\cap$ $p_j.S$ ($= \{x, y\}$) $\neq \phi$. $p_i \Rightarrow p_j$ since $p_i.S \neq \phi$, $p_i \rightarrow p_j$, and $p_i.S$ ($= \{x\}$) $\subseteq$ $p_j.S$ ($= \{x, y\}$). Hence, the event message e_1 is notified to the peer p_j.

Next, suppose a peer p_j publishes an event message e_2 with publication $e_2.P = \{x\}$ ($\subseteq p_j.P$). Here, $p_j \rightarrow p_k$ since $p_j.P$ ($= \{x\}$) $\cap$ $p_j.S$ ($= \{x, z\}$) $\neq \phi$. However, the peer p_j illegally flows to the peer p_k ($p_j \mapsto p_k$) since $p_j.S$ ($= \{x, y\}$) $\not\subseteq$ $p_k.S$ ($= \{x, z\}$). This means, an event message on the topic y which the peer p_k is not allowed to subscribe can be notified to the peer p_k. Here, event information illegally flow to the peer p_k from the peer p_j.

3 Synchronization Protocols

3.1 Subscription-based Synchronization (SBS) Protocol

We discuss how to prevent illegal information flow to occur among peers by publication and notification of event messages based on the TBAC model. A peer p_i is granted topics in the publication $p_i.P$ and subscription $p_i.S$. A peer p_i is associated with subsets $p_i.PP$ ($\subseteq p_i.P$) and $p_i.PS$ ($\subseteq p_i.S$) of the topics granted for publication and subscription, respectively, to the peer p_i. $p_i.PP$ and $p_i.PS$ are referred to as publication and subscription *purposes* [2] of a peer p_i, respectively. A peer p_i is allowed to issue a publication operation pb on a topic t only if $t \in p_i.PP$ and to issue a subscription operation sb on a topic t only if $t \in p_i.PS$. We first overview the subscription-based synchronization (SBS) protocol [18].

[Subscription-based synchronization (SBS) protocol] A peer p_i publishes an event message e with the publication $e.P$ including a topic t to a peer p_j.

1. If $p_i.S \Rightarrow p_j.PS$, the event message e is notified to the peer p_j and $p_j.S = p_i.PS \cup p_j.S$;
2. If $p_i.S \Rightarrow p_j.PS$ and $p_i.PS = \phi$, the event message e is notified to a peer p_j and $p_j.S = \{t\} \cup p_j.S$;
3. Otherwise, the notification of the event message e is banned at the peer p_j.

Suppose there are three peers p_i, p_j, and p_k as shown in Figure 1. Here, the publication purposes of the peers are $p_i.PP = \{y\}$, $p_j.PP = \{y\}$, and $p_k.PP = \{z\}$. The subscription purposes are $p_i.PS = \{x, y\}$, $p_j.PS = \{x, y\}$, and $p_k.PS = \{y, z\}$. The subscription $p_i.S$ is $\{y\}$. This means, the peer p_i may already have some event information on a topic y. $p_j.S$ and $p_k.S$ in the other peers p_j and p_k are ϕ. First, the peer p_i publishes an event message e_1 with publication $\{y\}$ ($\subseteq p_i.PP$). Here, the event message e_1 is notified to the peer p_j since $p_i \Rightarrow p_j$ ($p_i.PS \neq \phi$, $p_i \rightarrow p_j$, and $p_i.S \subseteq p_j.PS$). The subscription $p_j.S$ of the peer p_j is changed with $\{x, y\}$ since $p_j.S = p_i.PS$ $(= \{x, y\}) \cup p_j.S$ $(= \phi) = \{x, y\}$. Here, the subscription $p_j.S$ including the topic x means that the peer p_j may get some event information on the topic x from the peer p_i. However, the peer p_j does not get the information on the topic x since the peer p_i whose subscription $p_i.S$ does not include the topic x cannot event information messages the topic x in reality. Then, the peer p_j publishes an event message e_2 with publication $\{y\}$ ($\subseteq p_j.PP$). Here, $p_j \mapsto p_k$ ($p_j.PS \neq \phi$, $p_j \rightarrow p_k$, but $p_j.S$ $(= \{x, y\}) \not\subseteq p_k.PS$ $(= \{y, z\})$). Hence, the notification of e_2 is banned at the peer p_k.

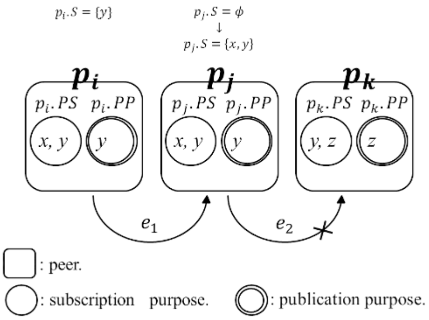


Fig. 1 SBS protocol.

In the SBS protocol, the notifications which may cause illegal information flow are banned. Each time an event message e is notified to a peer p_i, the implicit topics $e.I$ are accumulated in the peer p_i. This means, the more number of event messages are notified, the more number of event messages are banned [18].

3.2 Subscription Initialization SBS (SI-SBS) Protocol

Next, we overview the subscription initialization (SI) protocol [19]. The SI protocol is here renamed a subscription initialization SBS (SI-SBS) protocol. In the SI-SBS protocol, a notification which may cause illegal information flow is banned as well as the SBS protocol. Furthermore, if an event message e issued by a peer p_i is not notified, i.e. banned at some target peer p_j, the topics accumulated in $p_i.S$ of the peer p_i are initialized, i.e. removed in order to reduce the number of notifications to be banned.

[Subscription initialization SBS (SI-SBS) protocol] A peer p_i publishes an event message e with the publication $e.P$ including a topic t to a peer p_j.

1. If $p_i.S \Rightarrow p_j.PS$, the event message e is notified to the peer p_j and $p_j.S = p_i.PS \cup p_j.S$;
2. If $p_i.S \Rightarrow p_j.PS$ and $p_i.PS = \phi$, the event message e is notified to a peer p_j and $p_j.S = \{t\} \cup p_j.S$;
3. Otherwise, if the ratio of notifications banned to the total number of notifications which the peer p_i publishes is equal to or more than α, the subscription $p_i.S$ of the peer p_i is initialized, i.e. $p_i.S = \phi$;

In the SI-SBS protocol, we consider an initialization parameter α to allow a peer p_i to initialize its subscription $p_i.S$ if the notification of an event message published by the peer p_i is banned at some peer. If the ratio of number of notifications banned to the total number of notifications which the peer p_i publishes is equal to or more than the initialization parameter α, the subscription $p_i.S$ of the peer p_i is initialized.

We consider an example as shown in Figure 1. Suppose the initialization parameter $alpha$ is 0.5. The notification from the peer p_j to the peer p_k is banned since $p_j \not\mapsto p_k$. In this case, the ratio of the number of notifications banned to the total number of notifications in a publication of the peer p_j is 1. Here, $1 > \alpha$ (= 0.5). Hence, the subscription $p_j.S$ of the peer p_j is initialized.

3.3 Topic-based Synchronization (TBS) Protocol

In the SBS and SI-SBS protocol, it is checked if illegal information flow is to occur in terms of subscription purpose $p_i.PS$ of each peer p_i. In fact, each peer p_i manipulates only some, not necessarily all topics in the purpose $p_i.PS$. In order to reduce the number of notifications banned, we newly propose a topic-based synchronization (TBS) protocol where only topics which each peer manipulates are considered to check if illegal information flow to occur.

[Topic-based synchronization (TBS) protocol] A peer p_i publishes an event message e with the publication $e.P$ including a topic t to a peer p_j.

1. If $p_i.S \Rightarrow p_j.PS$, the event message e is notified to a peer p_j and $p_j.S = p_i.S \cup p_j.S$;

2. If $p_i.S \Rightarrow p_j.PS$ and $p_i.S = \phi$, the event message e is notified to a peer p_j and $p_j.S = \{t\} \cup p_j.S$;
3. Otherwise, the notification is banned at the peer p_j.

We consider the example of Figure 1. In the TBS protocol, the subscription $p_j.S$ is changed with $\{y\}$ when the event message e_1 published by the peer p_i is notified to the peer p_j since $p_j.S = p_i.S\ (= \{y\}) \cup p_j.S\ (= \phi) = \{y\}$. An illegal information flow relation $p_j \mapsto p_k$ does not hold since the subscription $p_j.S$ does not include the topic x. Then, the peer p_j publishes the event message e_2 to the peer p_k. Here, the notification of the event message e_2 is not banned at the peer p_k differently from the SBS protocol as shown in Figure 2.

In the SBS protocol, some event information on topics which are included in the subscription purpose of the publisher peer are considered to flow to the target peer even if the publisher peer does not have some event information on the topics. Hence, notifications more highly cause illegal information flow than the TBS protocol.

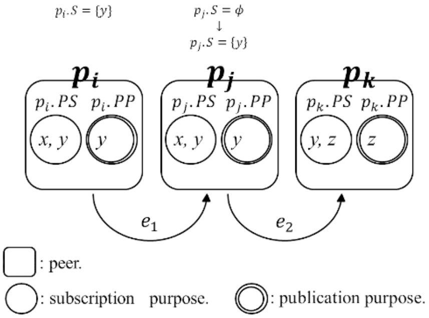


Fig. 2 TBS protocol.

3.4 Subscription Initialization TBS (SI-TBS) Protocol

In this paper, we also newly propose a subscription initialization TBS (SI-TBS) protocol.

[Subscription initialization TBS (SI-TBS) protocol] A peer p_i publishes an event message e with the publication $e.P$ including a topic t to a peer p_j.

1. If $p_i.S \Rightarrow p_j.PS$, the event message e is notified to the peer p_j and $p_j.S = p_i.S \cup p_j.S$;
2. If $p_i.S \Rightarrow p_j.PS$ and $p_i.S = \phi$, the event message e is notified to a peer p_j and $p_j.S = \{t\} \cup p_j.S$;

3. Otherwise, if the ratio of notifications banned to the total number of notifications which the peer p_i publishes is equal to or more than α, the subscription $p_i.S$ of the peer p_i is initialized, i.e. $p_i.S = \phi$;

We consider an example as shown in Figure 2. Suppose the initialization parameter α is 0.5. The event message e_2 published by the peer p_j is notified to the peer p_k since $p_j \Rightarrow p_k$. In this case, the ratio of the number of notifications banned to the total number of notifications published by the peer p_j is 0. Here, $0 < \alpha$ (= 0.5). Hence, the subscription $p_j.S$ of the peer p_j is not initialized.

4 Evaluation

We evaluate the TBS and SI-TBS protocols compared with the SBS and SI-SBS protocols on a topic set $T = \{t_1, \ldots, t_{tn}\}$ ($tn \geq 1$) and a peer set $P = \{p_1, \ldots, p_{pn}\}$ ($pn \geq 1$) in terms of number of notifications banned. If an event message is illegally notified to a peer p_i, the notification of the event message is banned in every protocol. In every protocol, the subscription $p_i.S$ is updated each time an event message is notified to the peer p_i. On illegally notifying an event message, the notification of the event message may be banned in every protocol. We assume an event message can be reliably broadcast to every target peer.

Publish (*pb*) and subscribe (*sb*) operations are supported on each topic t_k. Each peer p_i is granted publication purpose $p_i.PP$ and subscription purpose $p_i.PS$. Topics in the subsets $p_i.PP$ and $p_i.PS$ are randomly selected from the topic set T, i.e. access rights are randomly granted to each peer p_i. Publication purpose $p_i.PP$ of each peer p_i is composed of ptn_i ($\leq tn$) topics. In the evaluation, the number ptn_i of topics for each peer p_i is randomly selected out of numbers 0, 1, ..., tn. The subscription purpose of each peer p_i is composed of stn_i ($\leq mpstn$) topics. In the evaluation, the number stn_i of topics for each peer p_i is randomly selected out of numbers 0, 1, ..., $mpstn$. The publication $p_i.P$ and subscription $p_i.S$ of a peer p_i are initially empty ϕ.

First, a peer p_i is randomly selected in the peer set P. Then, the selected peer p_i publishes an event message with a topic t in the publication purpose $p_i.PP$ to every peer p_j whose subscription purpose $p_j.PS$ includes the topic t. In the SBS and SI-SBS protocols, each time an event message is notified to a target peer p_j, it is checked if illegal information flow to occur in terms of access rights of each peer. If illegal information flow might occur, the notification is banned. In the TBS and SI-TBS protocols, each time an event message is notified to a target peer p_j, it is checked if illegal information flow is to occur in terms of topics which each peer manipulates. If illegal information flow might occur, the notification of the event message is banned. Thus, no illegal information flow from the peer p_i to the peer p_j occur but some notifications may be banned in every protocol. The numbers of notifications banned in every protocol are measured.

In the evaluation, we consider twenty topics ($tn = 20$) and fifty peers ($pn = 50$). The initialization parameter α is 0.5. First, a collection P of fifty peers $p_1, \ldots, p_{50}$ are randomly generated on twenty topics $t_1, \ldots, t_{20}$, i.e. $P = \{p_1, \ldots, p_{50}\}$ and T

$= \{t_1, \dots t_{20}\}$. *en* shows the number of event messages exchanged by publication and subscription. Here, $0 \leq en \leq 500$. The number *en* of event messages exchanged between peers are performed on the topic set T in every protocol. We randomly create a peer set P on the topic set T seven hundred times for each *en*. Here, *mpstn* = 8. This means, the number of topics in the subscription purpose of each peer is randomly selected out of numbers 0, 1, ..., 8. One peer p_i is randomly selected in the peer set P and one topic t is randomly selected in the purpose $p_i.PP$. Then, the peer p_i publishes an event message e with the topic t. The event message e is notified to a target peer p_j. Then, it is checked if the legal information flow condition $p_i \rightarrow p_j$ holds. If not satisfied, the notification of the event message e is banned. In the SI-SBS and SI-TBS protocols, if the ratio of the notifications banned to the total number of notifications is equal to or more than α, the subscription $p_i.S$ of the peer p_i is initialized. These steps are iterated *en* times. For a given peer set P and each of every protocol, *en* event messages are published seven hundreds times. Then, we calculate the average ratio of the numbers of notifications banned in every protocol.

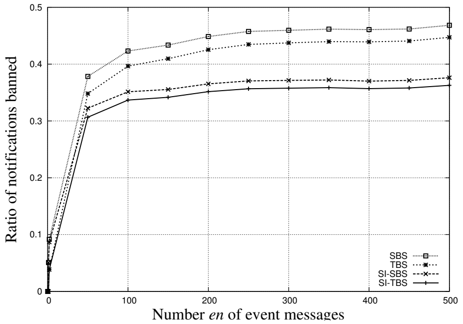


Fig. 3 Ratio of the notifications banned.

Figure 3 shows the ratios of number of illegal notifications to the total number of notifications in the every protocol. In the TBS protocol, the fewer number of notifications are banned than the SBS protocol. For example, about 47% of notifications are banned in the SBS protocol, but about 44% are banned in the TBS protocol for three hundred event messages ($en = 300$). In the SI-TBS protocol, the fewer number of notifications are banned than the SI-SBS protocol. For example, about 38% of notifications are banned in the SI-SBS protocol, but about 35% are banned in the SI-TBS protocol for three hundred event messages ($en = 300$). In the SI-TBS protocol, the fewest number of notifications are banned.

5 Concluding Remarks

In this paper, we newly proposed the topic-based synchronization (TBS) and subscription initialization TBS (SI-TBS) protocols to prevent illegal information flow among peers in a P2PPS model based on the topic-based access control (TBAC) model. The legal information flow relation $p_i \Rightarrow p_j$ among a pair of peers p_i and p_j means, no illegal information flow occur if an event message published by a peer p_i is notified to a peer p_j. In the SBS [18] and SI-SBS [19] protocols, it is checked whether or not the notification may cause illegal information flow in terms of subscription and publication rights of each peer. The notifications which may cause illegal information flow are banned to prevent illegal information flow. Here, some notifications which are not illegal are banned. In this paper, the TBS and SI-TBS protocols are proposed to reduce the number of banned notifications of event messages. In the TBS and SI-TBS protocols, it is checked whether or not the notification may cause illegal information flow in terms of only topics which each peer manipulates unlike the SBS and SI-SBS protocols. We evaluated the TBS and SI-TBS protocols in terms of number of notifications banned compared with the SBS and SI-SBS protocols. In the evaluation, we showed the number of notifications banned in the TBS and SI-TBS protocols is fewer than the SBS and SI-SBS protocols, respectively.

Acknowledgment

This work was supported by JSPS KAKENHI grant number 15H0295.

References

1. Blanco. R. and Alencar. P.: Event Models in Distributed Event based systems. Principles and Applications of Distributed Event-Based Systems, pp. 19–42, (2010).
2. Enokido, T. and Takizawa, M.: Purpose-based Information Flow Control for Cyber Engineering. IEEE Trans. on Industrial Electronics, **58**(6), pp. 2216–2225, (2011).
3. Eugster, P. T., Felber, P. A., Guerraoui, R., and Kermarrec, A. M.: The Many Faces of Publish/Subscribe. ACM Computing Surveys, **35**(2), pp. 114–131, (2003).
4. Ferraiolo, D. F., Kuhn, D. R., and Chandramouli, R. Role-based Access Control (2nd ed.), Artech, (2007).
5. Google alert, http://www.google.com/alerts.
6. Nakamura, S., Duolikun, D., Aikebaier, A., Enokido, T., and Takizawa, M.: Role-based Information Flow Control Models. Proc. of IEEE the 28th International Conference on Advanced Information Networking and Applications (AINA-2014), pp. 1140–1147, (2014).
7. Nakamura, S., Duolikun, D., Aikebaier, A., Enokido, T., and Takizawa, M.: Synchronization Protocols to Prevent Illegal Information Flow in Role-based Access Control Systems. Proc. of the 8th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2014), pp. 279–286, (2014).
8. Nakamura, S., Duolikun, D., Aikebaier, A., Enokido, T., and Takizawa, M.: Read-Write Abortion (RWA) Based Synchronization Protocols to Prevent Illegal Information Flow. Proc. of the 17th International Conference on Network-Based Information Systems (NBiS-2014), pp. 120–127, (2014).

9. Nakamura, S., Duolikun, D., Enokido, T., and Takizawa, M.: A Flexible Read-Write Abortion Protocol to Prevent Illegal Information Flow. Proc. of IEEE the 29th International Conference on Advanced Information Networking and Applications (AINA-2015), pp. 155–162, (2015).
10. Nakamura, S., Duolikun, D., Enokido, T., and Takizawa, M.: A Flexible Read-Write Abortion Protocol with Sensitivity of Objects to Prevent Illegal Information Flow. Proc. of the 9th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2015), pp. 289–296, (2015).
11. Nakamura, S., Duolikun, D., Enokido, T., and Takizawa, M.: A Flexible Read-Write Abortion Protocol with Sensitivity of Roles. Proc. of the 18th International Conference on Network-Based Information Systems (NBiS-2015), pp. 132–139, (2015).
12. Nakamura, S., Duolikun, D., Enokido, T., and Takizawa, M.: Role Safety in a Flexible Read-Write Abortion Protocol. Proc. of the 10th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA-2015), pp. 333–340, (2015).
13. Nakamura, S., Duolikun, D., Enokido, T., and Takizawa, M.: A Write Abortion-based Protocol in Role-based Access Control systems. International Journal of Adaptive and Innovative Systems, **2**(2), pp. 142–160, (2015).
14. Nakamura, S., Duolikun, D., Enokido, T., and Takizawa, M.: A Flexible Read-Write Abortion Protocol to Prevent Illegal Information Flow among Objects. Journal of Mobile Multimedia, **11**(3&4), pp. 263–280, (2015).
15. Nakamura, S., Duolikun, D., and Takizawa, M.: Read-abortion (RA) Based Synchronization Protocols to Prevent Illegal Information Flow. Journal of Computer and System Sciences, **81**(8), pp. 1441–1451, (2015).
16. Nakamura, S., Duolikun, D., Enokido, T., and Takizawa, M.: A Read-Write Abortion (RWA) Protocol to Prevent Illegal Information Flow in Role-based Access Control Systems. International Journal of Space-Based and Situated Computing, **6**(1), pp. 43–53, (2016).
17. Nakamura, S., Duolikun, D., Enokido, T., and Takizawa, M.: Influential Abortion Probability in a Flexible Read-Write Abortion Protocol. Proc. of IEEE the 30th International Conference on Advanced Information Networking and Applications (AINA-2016), pp. 1–8, (2016).
18. Nakamura, S., Enokido, T., and Takizawa, M.: Information Flow Control Models in Peer-to-Peer Publish/Subscribe Systems. Proc. of the 10th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2016), pp. 167–174, (2016).
19. Nakamura, S., Enokido, T., and Takizawa, M.: Subscription Initialization (SI) Protocol to Prevent Illegal Information Flow in Peer-to-Peer Publish/Subscribe Systems. accepted for publication at Proc. of the 19th International Conference on Network-Based Information Systems (NBiS-2016), (2016).
20. Nakayama, H., Duolikun, D., Enokido, T., and Takizawa, M.: Selective Delivery of Event Messages in Peer-to-peer Topic-based Publish/Subscribe Systems. Proc. of the 18th International Conference on Network-Based Information Systems (NBiS-2015), pp. 379–386, (2015).
21. Nakayama, H., Duolikun, D., Enokido, T., and Takizawa, M.: Reduction of Unnecessarily Ordered Event Messages in Peer-to-peer Model of Topic-based Publish/Subscribe Systems. Proc. of IEEE the 30th International Conference on Advanced Information Networking and Applications (AINA-2016), pp. 1160–1167, (2016).
22. Setty, V., Steen, M. V., Vitenberg, R., and Voulgaris, S.: PolderCast: Fast, Robust, and Scalable Architecture for P2P Topic-based Pub/Sub. Proc. of ACM/IFIP/USENIX 13th International Conference on Middleware (Middleware 2012), pp. 271–291, (2012).
23. Tarkoma, S.: Publish/Subscribe System : Design and Principles (First Edition). John Wiley and Sons, Ltd, (2012).
24. Tarkoma, S., Ain, M., and Visala, K.: The Publish/Subscribe Internet Routing Paradigm (PSIRP) : Designing the Future Internet Architecture. Future Internet Assembly, pp. 102–111, (2009).

Load-aware ACB Scheme for M2M Traffic in LTE-A Networks

Lijun Song, Wen'an Zhou, Yanjun Hou, Mengyu Gao
School of computer science, Beijing University of Posts and Telecommunications
Beijing, China
songlj1992@gmail.com, zhouwa@bupt.edu.cn, 605575422@qq.com, itsmengyu@163.com

Abstract. It's a primary challenge to support massive machine-type devices to access in LTE-A networks. Surging random access attempts will result in severe congestion to the network. Access Class Barring (ACB) scheme is a critical barring scheme proposed by 3GPP to control access attempts to alleviate the overload for the LTE-A networks. Most existing ACB schemes just consider the ideal case and ignore the effects of radio channels while setting the ACB factors. In this paper, we propose a load-aware Access Class Barring (ACB) scheme to predict access load and adjust the barring factor dynamically, which take into account the effects of radio channels. We improve two load prediction methods. Based on the prediction, we propose a method to adjust the barring factor as the predicted traffic load varies. The simulations results demonstrate that our proposed load-aware scheme efficiently outperforms the traditional ACB scheme in access success performance.

1 Introduction

The machine-to-machine (M2M) communication technology enables a large number of machine-type communication (MTC) devices to communicate with each other or remote servers without human interventions. Meanwhile, it enables the implementation of the Internet of things, in which ubiquitous connections can be established either on demand or in a periodic manner [1]. It is expected that there will be 12.5 billion M2M devices by 2020 [2].

The Long Term Evolution Advanced (LTE-A) network has great potentials to support M2M based on its ubiquitous coverage and mobility support. However, in general, massive devices may be triggered almost simultaneously and attempt to access the base station through the Random Access Channel (RACH). Thus M2M devices will pose a critical challenge for the network duo to the access of massive devices, which will cause severe congestion, intolerable delays, packet losses and even service unavailability [3].

To mitigate the congestion, many schemes have been proposed for solving the problem [4]–[6]. In [7], a scheme named Access Class Barring (ACB) is proposed by 3GPP, which is a scheme proposed for Radio Access Network (RAN) overload control.

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_7

In [6], the ACB scheme is adopted, However, the disadvantage is that the base station can't change the barring factor in time. In addition, some other mechanisms are proposed to alleviate the congestion caused by M2M communications. The authors In [8] propose an algorithm based on adaptive multiple Access Class Barring MACB factors, according to M2M traffic category. The authors in [9] propose two dynamic access class barring algorithms to determine the ACB factors. In [10]-[12], the authors propose some load-estimation ACB schemes to deal with the congestion. However, they just consider the ideal case without thinking about the preamble detection probability which results from channel fading and path-loss.

Generally, the network status depends on the number of device arrivals in current slot. So it is necessary to obtain the number of device arrivals to solve the congestion problem in M2M communications. However, the schemes mentioned above only consider the ideal condition and ignore the effects of radio channels, for example path-loss, fading, inter-cell interference, etc. In our work, as described in [6], we take into account the effects of radio channels and the preamble detection probability is assumed as $1-\frac{1}{e^i}$, where i indicates the i^{th} preamble transmission. Thus the proposed scheme can improve the access success performance based on the case.

Our main contribution is that we proposed a method to predict the numbers of the M2M devices in the next slot which considers the effects of radio channel while transmitting the preamble. Therefore, we can set the ACB parameters dynamically to guarantee the high access success probability. The simulation results demonstrate that our proposed scheme can provide the high access success probability.

The rest of the paper is structured as follows. The system model is presented in Section 2. The proposed ACB scheme is described elaborately in Section 3, including the predicting methods and the approach that eNB adjusting the barring factor dynamically. In Section 4, we display the simulation results and evaluate the performance of the proposed strategy. Finally, the paper is concluded in Section 5.

2 System Model

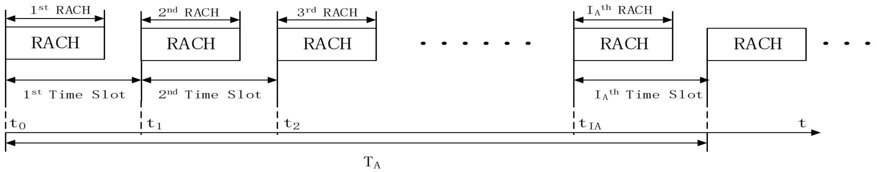


Fig. 1. Random access channel during T_A

We consider the M2M communication in an LTE-A system. The system consists of an eNB and massive M2M devices which are distributed in the cell uniformly. Furthermore, we consider the cell with N active M2M devices during T_A. We denote I_A as the number of random access channels within T_A. As shown in Fig. 1, we divide T_A into I_A discrete slots, each of which is set as 5ms. The i^{th} time slot starts at time t_{i-1} and ends at time t_i. We consider that new access within time slot i will only take place at the beginning of this time slot and choose the random access channel in this time slot for their random access attempts. Moreover, we consider the access procedure collided if two or more devices select the same preamble in the same time slot and these devices can proceed to the access procedure after the back-off time. Meanwhile, the access of the device fails if the retransmission times exceeds the maximum retransmission times. Additionally, the access of the device succeeds if no other devices select the same preamble as the device in the same time slot. There are M_i devices which transmit the preambles after ACB procedure in the i^{th} slot and the number of available preambles are K in each time slot. As described in [13], the new arrival density probability of the devices denoted as T_A with probability $g(t)$ and follows a beta distribution with parameters $x = 3, y = 4$

$$g(t) = \frac{t^{x-1}(T_A - t)^{y-1}}{T_A^{x+y-1}\beta(x,y)}\ , 0 \le t \le T_A\ . \quad \textbf{(1)}$$

where $\beta(x, y)$ is the beta function [14].

2.1 Access Class Barring

In each time slot, the eNB broadcasts an ACB factor p ($0 \le p \le 1$) and the M2M devices need to pass through the ACB procedure before transmitting the preamble. Each M2M device generates a random number q ($0 \le q \le 1$). If q is less than the ACB factor, the device proceeds to the random access procedure. Otherwise, it is barred for the duration of the ACB time, which is calculated as (2) [15].

$$T1 = (\beta + \gamma \times rand) \times T. \quad \textbf{(2)}$$

In this equation, *rand* is a random number uniformly drawn from interval [0,1). Meanwhile, T is the barring time factor which is also broadcasted by the eNB and we set it as 20ms in this paper, $\beta = 0.7$, $\gamma = 0.6$. After the duration of the ACB time, the device can repeat the ACB procedure again. Fig. 2.depicts the ACB scheme.

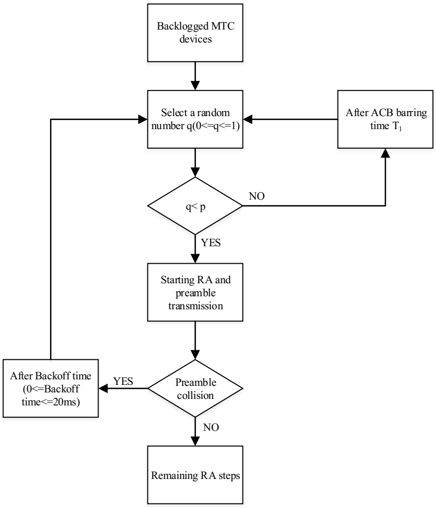

Backlogged MTC devices
Select a random number q(0<=q<=1)
After ACB barring time T1
q< p
NO
YES
Starting RA and preamble transmission
Preamble collision
YES
After Backoff time (0<=Backoff time<=20ms)
NO
Remaining RA steps

Fig. 2. ACB procedure and RA diagram

We denote N_i as the number of device arrivals in slot i. The actual number of the devices which passed through the ACB check is denoted as M_i and the estimation value of M_i is denoted as M_i’. Because M_i is from N_i which pass through the ACB procedure, it can estimate N_i based on the value of M_i’. Meanwhile, we consider the number of device arrivals in the next slot N_{i+1} is approximate to the N_i.

We consider M_i= j devices which pass through the ACB check successfully among $Ni = n$ backlogged devices. We further consider each of these devices that passed through the ACB check selects a random access preamble from K available preambles with an equal probability which is equal to $\frac{1}{K}$. For a specific preamble m transmitted to the eNB, let $D_m = 1$ denote the case that the preamble m is selected by exactly one user, which means the device can access the network successfully. The probability that only one user selects preamble m is

$$\mathrm{P}\,(D_m = 1 \mid M_i = j) \ = \sum_{q=1}^{j} \binom{j}{q} \binom{q}{1} \frac{1}{e^{q-1}} \frac{1}{K^q} (1-\frac{1}{e})(1-\frac{1}{K})^{j-q} \ . \tag{3}$$

In (3), for simplicity, we assume the preamble detection probability is$1-\frac{1}{e}$, however, the preamble detection probability is related to the transmission times of the preamble. Thus we will adjust to the result in the below analysis.

The expected number of successful preamble transmissions in time slot i can be obtained by

$$\mathrm{E}(D_i \mid M_i = j) = \sum_{m=1}^{K} \mathrm{P}\,(D_m = 1 \mid X_i = j) \ = K \sum_{q=1}^{j} \binom{j}{q} \binom{q}{1} \frac{1}{e^{q-1}} \frac{1}{K^q} \left(1-\frac{1}{e}\right) \left(1-\frac{1}{K}\right)^{j-q} \ . \tag{4}$$

Therefore,

$$\begin{aligned} \mathrm{E}(D_i \mid N_i = n) &= \sum_{j=1}^{n} P(M_i = j \mid N_i = n) \times \sum_{m=1}^{K} \mathrm{P}\,(D_m = 1 \mid M_i = j) \\ &= \sum_{j=1}^{n} \binom{n}{j} p^j (1-p)^{n-j} \times K \sum_{q=1}^{j} \binom{j}{q} \binom{q}{1} \frac{1}{e^{q-1}} \frac{1}{K^q} \left(1-\frac{1}{e}\right) \left(1-\frac{1}{K}\right)^{j-q} \\ &= np \left(1-\frac{p}{Ke}\right)^{n-1} (1-\frac{1}{e}) \quad . \end{aligned} \tag{5}$$

2.2 The Random Access Procedure

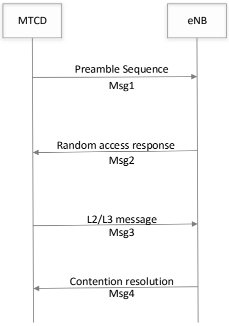


Fig. 3. RA four-steps procedure

The devices which pass through the ACB check will attempt the contention-based random access (RA) procedure, which consists of a four-message handshake with the eNB[16], as shown in Fig. 3.In this process, the preamble sequences are generated from cyclic shifts of root Zadoff-Chu sequences [17]. The eNB can get the number of successful preamble and collisional preamble, but it does not know the actual value of *Mi* and has to obtain its estimation value M_i*'*, if it requires to figure out the status of the network.

3 Load-aware ACB Scheme

In this section, we firstly propose two schemes to predict the number of devices which will arrive in the next slot based on the estimation number of devices transmitting the preambles in the current slot, so we can estimate the number of device arrivals N_i according to the estimation value of M_i. Then, we propose an adaptive scheme to change the barring factor *p* dynamically.

3.1 Load-estimation Scheme

In [11], it estimates the number of devices according to the probability of the idle preambles. In [10], it estimates the number of devices according of the use of preambles. Based on the two methods, we improve these methods by taking into preamble detection probability account. Finally, we propose the two methods to predict the number of the devices in the next slot and evaluate the accuracy of the prediction. In this paper, we call these two methods as load-estimation scheme based on the idle preambles and load-estimation scheme based on the Markov Chain respectively.

Load-estimation Scheme Based on the Idle Preambles

In this scheme, the eNB counts the number of idle preamble $L_{idle,i}$ among the total number of preambles in the i^{th} slot L_i. The probability of the preamble being idle $p_{idle,i}$ is calculated from $L_{idle,i}$,

$$p_{idle,i} = \frac{L_{idle,i}}{Li} \quad . \tag{6}$$

The probability of the preamble being idle can also be determined by

$$p_{idle,i} = \left(1 - \frac{1}{L_i} + \frac{1}{e*L_i}\right)^{M_i} \quad . \tag{7}$$

In (7), there are two reasons for the preamble being idle. Firstly, for a specific device, it doesn't select the preamble actually. Secondly, the device select the preamble actually but the preamble can't be detected by the eNB due to the effects of the radio channel. Because many device just transmit the preamble for once, we consider the preamble detection probability is $1 - \frac{1}{e}$ for every transmission. And we will adjust to the result in the below analysis. The number of active devices M_i in the i^{th} slot can be estimated as

$$M_i' = \frac{\log(p_{idle,i})}{\log(1-\frac{1}{L_i}+\frac{1}{e*L_i})} \quad . \tag{8}$$

From M_i*'* and the ACB factor $p_{,i}$ in the i^{th} slot, the eNB can predict the number of active devices in the $(i + 1)^{th}$ slot.

$$N_{i+1} \cong N_i = \frac{M_i'}{p_i} \ . \qquad \textbf{(9)}$$

However, the preamble detection probability is relevant to the transmission times of preambles and the probability is fixed as $1-\frac{1}{e}$ in the above equation. Thus we should adjust to the equation and the new result is as follows.

$$N_{i+1} \cong N_i = \frac{M_i'\alpha}{p_i} \ . \qquad \textbf{(10)}$$

To reduce the effect of prediction error, the eNB can predict the number of devices arrivals in the $(i+1)^{\text{th}}$ slot by averaging the estimation value of the current slot and the previous slot, as shown in the below equation.

$$N_{i+1} = \frac{\frac{M_i'\alpha}{p_i}+\frac{M_{i-1}'\alpha}{p_{i-1}}}{2} \ . \qquad \textbf{(11)}$$

Finally we find the appropriate value of α is 0.8 after a number of simulation tests.

Load-estimation Scheme Based on the Markov Chain

In this scheme, the eNB can predict N_{i+1} according to the status of preambles. The set of preambles that no device to select is denoted as I and $|I|(|I| = 0, 1, \ldots, K)$ is the cardinality of I. The set of preambles that exactly one device to select is denoted as S and $|S|(|S| = 0, 1, \ldots, K)$ is the cardinality of S. The set of preambles that two or more than two devices to select is denoted as C and $|C|(|C| = 0, 1, \ldots, K)$ is the cardinality of C. So $|I|$ means the number of idle preambles; $|S|$ means the number of successful preamble or the number of device which access successfully; $|C|$ means the number of collisional preamble. We denote the status of preambles as the combination of $(|I|, |S|, |C|)$, namely $|I| + |S| + |C| = K$. we can obtain the total number of combinations of preambles' status from the above equation

$$\textstyle\sum_{i=1}^{K}\binom{i}{1} = \frac{1}{2}(K+1)(K+2) \ . \qquad \textbf{(12)}$$

but M_i is unknown by the eNB and $|I|$, $|S|$, $|C|$ are known by the eNB.

The eNB can estimate the random variable M_i based on the values of $|I|$, $|S|$ and $|C|$, and the problem can be formulated by

$$M_i' = \arg \max_{0 \le m \le N} \{\Pr(M_i = m| \, |I| = i, |S| = s, |C| = c)\} \ . \qquad \textbf{(13)}$$

where $i, s, c = 0, 1, \ldots K$; $m = 0, 1, \ldots N$. Based on the Bayes' theorem, the estimation equation becomes

$$M_i' = \arg \max_{0 \le m \le N} \left\{\frac{\Pr(M_i = m)}{\Pr(|I| = i, |S| = s, |C| = c)} \times \Pr(\,|I| = i, |S| = s, |C| = c|M_i = m)\right\} \ . \qquad \textbf{(14)}$$

Here, we use the maximum likelihood estimation to simplify the estimation. Then the result becomes

$$M_i' = \arg \max_{0 \le m \le N} \{\Pr(\,|I| = i, |S| = s, |C| = c|M_i = m)\} \ . \qquad \textbf{(15)}$$

M_i devices which select the preambles are equivalent to M devices which apply for the preambles one by one. Therefore, we establish a Markov Chain to estimate M_i using this transformation, and we can obtain the Transition Probability Matrix **P** based on the state transition.

In this paper, we denote (K, 0, 0) as the first state, $(K-1, 1, 0)$ as the second state, and so on. We denote P_{ij} as the probability of i^{th} state transform to the j^{th} state. Here, we define n as the n^{th} state of the preambles and the total number of states is $\frac{(K+1)(K+2)}{2}$. We can get that if the state is $(|I|, |S|, |C|)$, then $n = \frac{(2K+3-|C|)\times|C|}{2} + |S| + 1$.

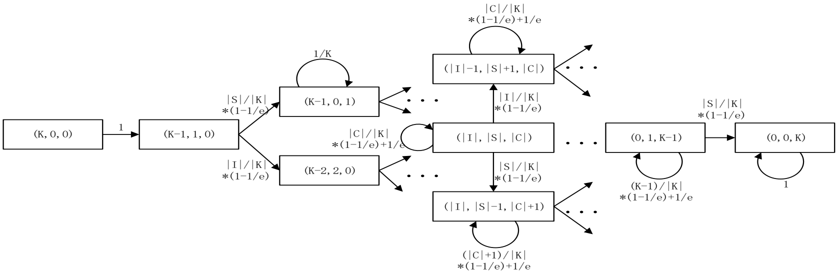


Fig. 4. The preambles' state transition diagram.

As shown in Fig. 4, if a device selects the preamble belonging to set *I*, the number of preambles in set *I* will minus one, the number of preambles in set *S* will add one and the number of preambles in set *C* is invariable. The transition of the states can be denoted as ($|I|, |S|, |C|$) to ($|I| - 1, |S| + 1, |C|$) and the transition probability P_{ij} is $\frac{|I|}{K} \times (1 - \frac{1}{e})$. If a device selects the preamble belonging to set S, the number of preambles in set C will add one, the number of preambles in set S will minus one and the number of preambles in set I is invariable. The transition probability P_{ij} is $\frac{|S|}{K} \times (1 - \frac{1}{e})$. If a device selects the preamble belonging to set C, the state will stay the same and the transition probability P_{ij} is$\frac{|C|}{K} \times \left(1 - \frac{1}{e}\right) + \frac{1}{e}$. Moreover, the transition probability of ($|I|, |S|, |C|$) to other states is 0. So we can obtain the Transition Probability Matrix **P**. One device selecting a preamble means a transformation of the state, so M_i is the steps of transformation. Thus the estimation of M_i is equivalent to find the steps of transformation which satisfies the formula (15). Because the devices which select the preambles are from the active devices which pass through the ACB procedure, and the probability is p_i. The prediction of N_{i+1} can be formulated as

$$N_{i+1} \cong N_i = \frac{M_i'}{p_i} \ . \tag{16}$$

However, the preamble detection probability is relevant to the transmission times of preambles and the probability is fixed as $1 - \frac{1}{e}$ in the above equation. Thus we should adjust to the equation and the new result is as follows.

$$N_{i+1} \cong N_i = \frac{M_i'\alpha}{p_i} \ . \tag{17}$$

Similar to the load-estimation scheme 1, to reduce the effect of prediction error, the eNB can predict the number of devices arrivals in the (i + 1)th slot by averaging the estimation value of the current slot and the previous slot, as shown in the below equation.

$$N_{i+1} = \frac{\frac{M_i'\alpha}{p_i} + \frac{M_{i-1}'\alpha}{p_{i-1}}}{2} \ . \tag{18}$$

Finally we find the appropriate value of α is 0.8 after a number of simulation tests.

3.2 The Adjustment of the ACB Factor p

It is critical to find the optimal value of p to accommodate the massive devices. In [9], the authors propose a method to obtain the optimal value of p. We will improve the method proposed by [9] and take the preamble detection probability into consideration.

the expected number of successful preamble transmissions in each time slot should be maximized. By taking the derivative of (5) with respect to p, we can obtain

$$\frac{d}{dp}E(D_i \mid N_i = n) = n(1-\frac{1}{e})(1-\frac{np}{Ke})(1-\frac{p}{Ke})^{n-2} \ . \qquad \textbf{(19)}$$

When K ≥ n, we have $\frac{d}{dp}E(D_i \mid N_i = n) \geq 0$. Thus the value is growing up gradually and the maximum value is achieved when p = 1. That is, when the preamble number is greater than the number of backlogged devices, the ACB factor should be set to 1. When K< n, we set $\frac{d}{dp}E(D_i \mid N_i = n) = 0$, and obtain

$$p^* = \frac{Ke}{n} \ . \qquad \textbf{(20)}$$

Because we regard the preamble detection probability as $1-\frac{1}{e}$ in the above equations for simplicity, however, the preamble detection probability is relevant to the transmission times of preambles. Therefore, we should add a parameter to adjust to the result. Thus we have

$$p^* = min\left(1, \frac{Ke \times \alpha}{n}\right) \ . \qquad \textbf{(21)}$$

Finally we find the appropriate value of α is 0.3 after a number of simulation tests. Fig.5 illustrates the load-aware ACB scheme based on the prediction result.

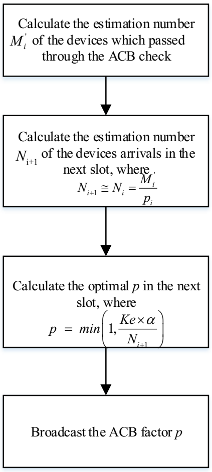

Calculate the estimation number M_i' of the devices which passed through the ACB check
Calculate the estimation number N_{i+1} of the devices arrivals in the next slot, where $N_{i+1} \cong N_i = \frac{M_i'}{p_i}$
Calculate the optimal p in the next slot, where $p = min\left(1, \frac{Ke \times \alpha}{N_{i+1}}\right)$
Broadcast the ACB factor p

Fig. 5. load-aware ACB scheme .

4 Simulation Results

4.1 Simulation Parameters

The simulations consider the access success performance of 10000 to 50000 M2M devices in a single cell activating with a *Beta* distribution over 10 seconds. The RACH is configured to occur every 5 ms, with up to 54 preambles. During analysis, the simulation parameters are listed in Table 1, which has been agreed by 3GPP.

Table1. Parameters for simulation

Parameter	Value
Number of M2M devices	10000~50000
Distributed period for Beta traffic distribution	10 seconds
Beta function(α,β)in Beta traffic distribution	(3,4)
PRACH configuration	2 PRACH per 10ms
Back-off indicator	20ms
Max number of preamble transmission	10
Preamble number	54
Preamble detection probability for the ith preamble transmission	$1-\frac{1}{e^{i}}$

4.2 Comparison of Load-estimation Schemes

To illustrate the prediction performance of the proposed methods, we calculate the average absolute error, the average relative error and the average standard deviation between the predicted value from load-estimation scheme based on the idle preambles and the actual arrivals value. Meanwhile, we calculate the above factors for load-estimation scheme based on the Markov Chain. Furthermore, we calculate the above factors for another two methods without taking the preamble detection probability into consideration. Then, we compare the prediction performance for the four methods and the results are described in the Fig.6, Fig.7, and Fig.8 respectively. In these figures, traditional scheme based on the idle preambles and traditional scheme based on the Markov Chain mean the scheme proposed in [11] and the scheme proposed in [10] respectively, which doesn't take into account the effects of radio channels. We can observe from these figures that load-estimation scheme based on the Markov Chain we proposed in this paper outperforms the other methods in the prediction for the M2M devices arrivals value in the next slot and the predicted error is acceptable.

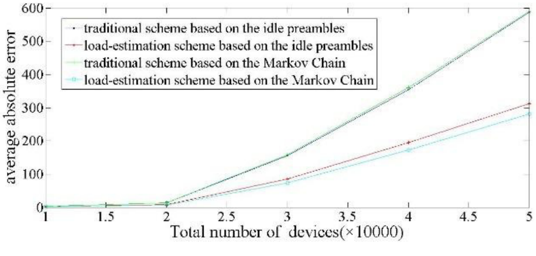


Fig. 6. The average absolute error of the four prediction methods

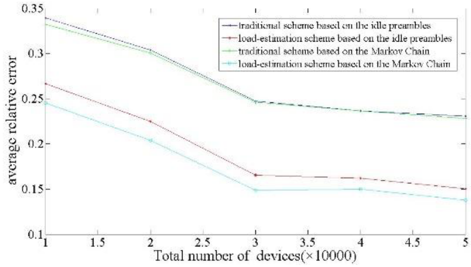


Fig. 7. The average relative error of the four prediction methods

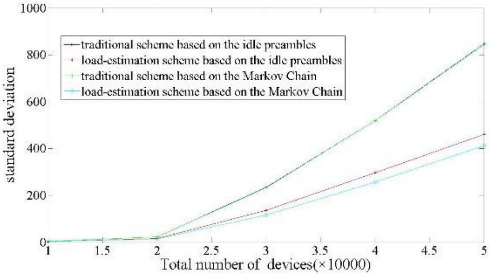


Fig. 8. The standard deviation of the four prediction methods

4.3 Comparison of Load-aware ACB Scheme and Traditional ACB Scheme

From the above analysis, we can figure out that load-estimation scheme based on the idle preambles and load-estimation scheme based on the Markov Chain are better than the other two methods without preamble detection probability in the prediction performance. In this section, we would prefer to select load-estimation scheme based on the Markov Chain to predict the arrivals value rather than load-estimation scheme based on the idle preambles from the above analysis. Then, we compare the performance of the proposed load-aware ACB scheme and the traditional ACB scheme which is proposed in [10] in the following metrics: Access success probability, Average retransmission times and Collision rate. Fig.9 depicts the access success probability for the two schemes and we can figure out the access success probability of the load-aware ACB scheme is higher than the traditional ACB scheme. Meanwhile, we can observe from the Fig.9 that the access success probability is almost equal to 1 in the load-aware ACB scheme.

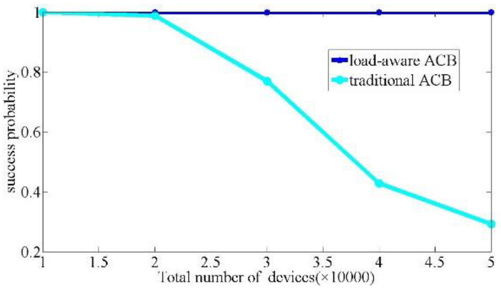


Fig. 9. The success probability with different number of devices

We can also figure out from the Fig.10 and Fig.11 that the average retransmission times and collision rate of the proposed load-aware ACB scheme is relatively lower than the traditional ACB scheme. Meanwhile, the average retransmission times and collision rate will grow as the total number of the M2M devices grows.

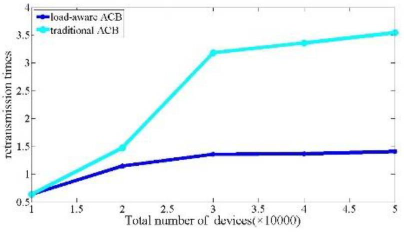


Fig. 10. The retransmission times with different number of devices

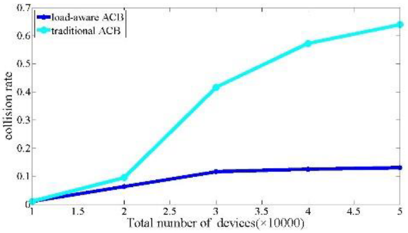


Fig. 11. The collision rate with different number of devices

5 Conclusions

In this paper, we propose a load-aware ACB scheme with considering the effects of radio channel to alleviate the congestion when a great number of M2M devices prepare to access the eNB within a short time. Meanwhile, compared to the traditional ACB scheme, the proposed ACB scheme improve the access success performance. Simulation results indicate the load-aware ACB scheme can improve the access success probability and reduce the average retransmission times and collision rate.

Acknowledgements. The research work reported in this paper is supported by the National Nature Science Foundation of China (61372113) and National Science and Technology Major Project (No.2015ZX03001025).

References

[1] G. Wu, S. Talwar, K. Johnsson, N. Himayat, and K. D. Johnson.: M2M: From mobile to embedded Internet. IEEE Comm. Magazine, vol. 49, no. 4, pp. 36–43, Apr. 2011.

[2] Machina Research Sector Report.: Machine-to-Machine (M2M) communication in consumer electronics 2012-22. Feb. 2013.

[3]T. P. C. de Andrade, C. A. Astudillo and N. L. S. da Fonseca.: The impact of massive machine type communication devices on the access probability of human-to-human users in LTE networks. 2014 IEEE Latin-America Conference on Communications (LATINCOM), Cartagena de Indias, 2014, pp. 1-6.

[4] S.-Y. Lien, K.-C. Chen, and Y. Lin.: Toward ubiquitous massive accesses in 3GPP machine-to-machine communications. IEEE Commun. Mag.,vol. 49, no. 4, pp. 66-74, Apr. 2011.

[5] G. Wang, X. Zhong, S. Mei, and J. Wang.: An adaptive medium access control mechanism for cellular based machine to machine (M2M) communication. in Proc. IEEE International Conference on Wireless Information Technology and Systems (ICWITS) 2010, pp. 1-4, Aug 2010.

[6] 3GPP.: Access class barring and overload protection.3GPP TR23.898 V7.0.0, Mar. 2005.

[7] 3GPP TR 37.868 V11.2.0 (2011-09): Study on RAN Improvements for Machine-type Communications. Sept. 2011.

[8] S. Gharbi and N. Zangar.: Adaptive multiple Access Class Barring factors for M2M communications in LTE-A Networks.2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC), Las Vegas, NV, 2015, pp. 605-606.

[9] S. Duan; V. Shah-Mansouri; Z. Wang; V. Wong.: D-ACB: Adaptive Congestion Control Algorithm for Bursty M2M Traffic in LTE Networks. in IEEE Transactions on Vehicular Technology, vol.PP, no.99, pp.1-1.Feb.2016

[10] H. He, Q. Du, H. Song, W. Li, Y. Wang and P. Ren.: Traffic-aware ACB scheme for massive access in machine-to-machine networks. 2015 IEEE International Conference on Communications (ICC), London, 2015, pp. 617-622.

[11] C. Y. Oh, D. Hwang.: Joint Access Control and Resource Allocation for Concurrent and Massive Access of M2M Devices. in IEEE Transactions on Wireless Communications, vol. 14, no. 8, pp. 4182-4192, Aug. 2015.

[12] C. M. Chou, C. Y. Huang and C. Y. Chiu.: Loading prediction and barring controls for machine type communication. 2013 IEEE International Conference on Communications (ICC), Budapest, 2013, pp. 5168-5172.

[13] 3GPP TSG RAN WG2 #71 R2-104663.: [70bis#11] LTE: MTC LTE simulations. ZTE, Madrid, Spain, 23rd Aug. 2010.

[14] A. K. Gupta and S. Nadarajah, Handbook of Beta Distribution and Its Applications. CRC Press, 2004.

[15] U. Phuyal, A. T. Koc, Mo-Han Fong, R. Vannithamby.: Controlling access overload and signaling congestion in M2M networks. in Proc. Conference Record of the Forty Sixth Asilomar Conference on Signals,Systems and Computers (ASILOMAR) 2012, pp. 591-595, Nov. 2012.

[16] A. Laya, L. Alonso.: Is the Random Access Channel of LTE and LTE-A Suitable for M2M Communications? A Survey of Alternatives. IEEE Communications Surveys & Tutorials, vol. 16, no. 1, pp. 4-16, First Quarter 2014.

[17] 3GPP TS 36.211.: E-UTRA Physical Channels and Modulation.

An Energy-Efficient Macro-assisted Sleep-Mode Scheme in Heterogeneous Networks

Xuhui Yang[1], Wen'an Zhou[1], Da Li[1]
[1] Beijing University of Posts and Telecomunications Beijing, China
jimmyyoung24@outlook.com, zhouwa@bupt.edu.cn, freedani@bupt.edu.cn

Abstract. Ultra-dense small cell deployment is seen as a necessary means to address the explosive mobile traffic growth in the near future. However, a large-scale small cell deployment can substantially increase the network energy consumption with strong ecological and economic implications. In this article, we introduce an energy-efficient macro assisted sleep mode scheme in heterogeneous networks to reduce cellular networks' power consumption. The designed scheme takes into account that (i) macro base station's power consumption is varying with the load, (ii) macro starts a UE-small cell connection procedure as soon as a connected UE is required to start receiving data is potentially suboptimal in terms of energy consumption. Furthermore, we present the procedures of the proposed macro assisted sleep mode scheme. By our calculation and analysis, our scheme can yield a further 5% energy savings with respect to the amount of energy savings obtained with existing schemes.

1 Introduction

Fueled by the popularization of mobile devices and the increased number of broadband services, wireless traffic experienced an exponential growth in recent years. Such growth is very likely to continue in the near future owing to new cloud-based services and data-hungry applications [1]. In order to be able to serve such a big amount of traffic, wireless networks shall increase their capacity.

Ultra-dense small cell deployment is seen as a necessary means to address the huge capacity demand [2], [3]. However, as base stations are responsible for the large amount of energy consumed in cellular networks, deploying such large number of small cells poses two main challenges: the cost of the network infrastructure and its environmental footprint. Both issues can be solved by designing an energy efficient mobile network: the decrease of the network energy consumption will result in lower operational cost for the infrastructure and lower greenhouse gas emissions.

While there are various distinctive approaches to reduce energy consumptions in a mobile cellular network, adopting renewable energy resources or improving design of certain hardware is often prohibitive due to the cost of replacing, and installing new equipment. By comparison, sleep mode techniques, does not require changes to current network architecture, and takes advantage of changing traffic patterns on daily

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_8

or weekly basis. Hence, sleep mode have the potential to save a significant amount of energy, as shown in various studies.

Decoupling coverage and capacity provisioning could make it easier to deploy and manage a large number of small cells. DOCOMO proposes the concept of Phantom Cell which gives the ideas of high-frequency small cells configured no Cell specific Reference Signal (CRS) [4]. In the Phantom Cell Concept, the control plane (C-plane) and the (U-plane) are separated [5], as shown in Figure 1. In such systems, effective macro-assisted energy savings schemes based on sleep mode techniques can be implemented with minimal additional signaling. Thus, in this paper we introduce an energy efficient macro assisted sleep mode scheme in heterogeneous networks.

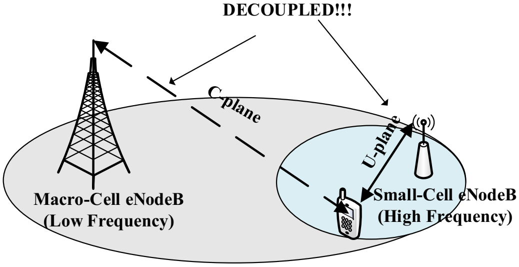


Fig. 1. Heterogeneous networks with the C-plane and the U-plane are separated.

We note that there are several previously published papers, for example [6], [7], which already studied the base stations sleep mode in heterogeneous networks. However, only a few papers considered macro assisted scenarios. Among these, we cite [9], where the authors show that in macro-controlled small cells, effective macro-assisted energy savings schemes can be employed to reduce the network energy consumption at practically no additional cost to the network and no hit to user quality of service (QoS). However, [9] does not consider that:

- The energy consumed by macro cell is not constant, but rather variable with the traffic load (approximately linearly);
- When a small cell is switched off, the macro cell assumes the load of that small cell, hence the macro cell energy consumption increases.
- Macro cells have data capacity, thus, in some case a connection to a small cell is not needed: the macro cell base station could handle the transmission of the arriving file with its own resources, especially when the file is relatively small.

The scheme introduced in this paper considers the aforementioned aspects, and thus minimizes the energy consumption in heterogeneous networks and provides a more energy efficient as well as a more realistic sleep mode scheme.

The rest of this paper is organized as follows. In Section 2, we introduce the macro-controlled heterogeneous networks architecture and the energy model considered within the scope of this paper. In Section 3 we propose our macro assisted sleep mode scheme. Section 4 provides numeric analysis of the proposed scheme and illustrate energy savings. Section 5 concludes the paper.

2 System Model

2.1 System Architecture

The system in this paper is macro-controlled small cells heterogeneous networks, such as the PCC architecture [5], and the Macro-assisted Data-Only Carrier System in [11]. Such macro-controlled system advocates the separation of the C-plane and the U-plane. Such system, illustrated in Figure 2, is comprised of two overlaid heterogeneous networks:

- A macro cell network, whose primary objective is to offer a C-plane coverage to users. These macro cells operate in lower frequency bands (e.g., 2 GHz);
- A small cell network with small cell BSs, responsible for delivering a high throughput to connected users. These small cells operate in higher frequency bands (e.g., 3.5 GHz), and are connected to a macro cell through a backhaul link.

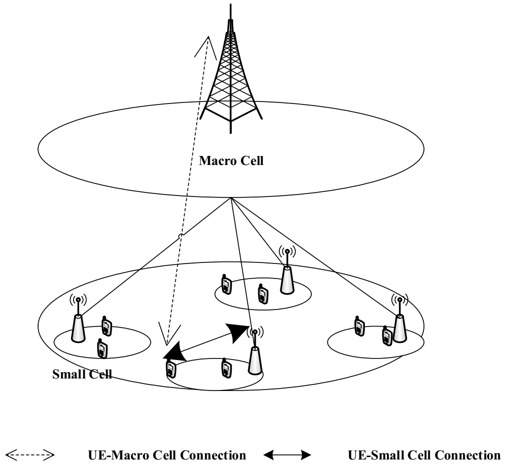

Macro Cell
Small Cell
<----> UE-Macro Cell Connection
◄──► UE-Small Cell Connection

Fig. 2. System architecture of macro-controlled small cells, a heterogeneous network.

Additionally, we consider two small cell states, namely on and sleep. In the on state, all hardware components in the small cell base station are fully switched on, small cells consume the highest amount of energy. In the sleep state, some of the hardware components are either completely switched off or operated in low-power modes. The exact components to be switched off are a function of the specific hardware architecture and the particular energy saving algorithm.

2.2 Energy Model

In order to evaluate the potential of the system energy efficiency, an appropriate base station power consumption model is needed. The energy model we consider in this paper is introduced in [8], based on the following assumptions:

- The power consumption of a macro base station is equal to $a\ell + W_M$, where $\ell \in [0,1]$ is the traffic load of the macro base station normalized to its maximum capacity, and W_M is the power consumption when $\ell = 0$, a is a constant value;

$$P_M = a\ell + W_M \tag{1}$$

- The power consumption of a small base station is constant (independent from the traffic load variations) and equal to W_S.

$$P_S = W_S \tag{2}$$

3 Energy Efficient Macro Assisted Sleep Mode Scheme

3.1 Problem Description

The goal of sleep mode technology is to minimize the energy consumption while guaranteeing that the QoS requirements of each user are fulfilled. At a given time granularity, this system optimization problem can be formulated as follows:

$$\text{Minimize} \quad \sum_{n=1}^{N}[C_n E_S^{on} + (1 - C_n)E_S^{sleep}] + E_M \tag{3}$$

$$\text{Subject to:} \quad R_k \geq \rho_k, \forall k \in \{1,...,K\}, \tag{4}$$

$$C_n \in \{0,1\}, \tag{5}$$

where N is the total number of small cells deployed in the coverage of macro cell, K is the number of users in the system, E_S^{on} is the energy consumed by a small cell in the on state, E_S^{sleep} is the energy consumed by a small cell in the sleep state, E_M is the energy consumed by the macro cell, R_k is the throughput (or data rate) obtained by UE k, ρ_k is the data rate requirement of user k,

$$\text{and} \quad C_n = \begin{cases} 1 \text{ if small cell n is in the on state} \\ 0 \text{ if small cell n is in the sleep state} \end{cases}$$

Expression (4), imposing that the data rate requirement of each UE in the system be fulfilled, means that each BS in the system need to have enough bandwidth to fulfil the data rate requirement of all the UEs it is serving. Thus, it is possible to rewrite expression (4) as:

$$\sum_{k=1}^{K} \frac{\rho_k}{\omega_{n,k}} \alpha_{n,k} \le B_{total,n}, \quad \forall n \in \{1,...,N\}, \tag{6}$$

$$\sum_{k=1}^{K} \frac{\rho_k}{\omega_{macro,k}} \left(1 - \sum_{n=1}^{N} \alpha_{n,k}\right) \le B_{total,macro}, \tag{7}$$

$$\sum_{n=1}^{N} \alpha_{n,k} \le 1, \quad \forall k \in \{1,...,K\}, \tag{8}$$

$$\alpha_{n,k} \in \{0,1\}, \quad \forall n \in \{1,...,N\}, k \in \{1,...,K\}, \tag{9}$$

where $\omega_{n,k}$ ($\omega_{macro,k}$) is the spectral efficiency of the link between small cell n (macro cell) and user k, in bit/s/Hz, $B_{total,n}$ ($B_{total,macro}$) is the total bandwidth available on small cell n (macro cell), in Hz,

and $\alpha_{n,k} = \begin{cases} 1 \text{ if small cell n is serving user k} \\ 0 \text{ if small cell n is not serving user k} \end{cases}$

To minimize the total power consumption of the system, a small cell should be in sleep state whenever it is not serving any user, but in the on state when serving at least one user. This means C_n can be expressed as follows:

$$C_n = \begin{cases} 1 \text{ if} \sum_{k=1}^{K} \alpha_{n,k} \ge 1 \\ 0 \text{ if} \sum_{k=1}^{K} \alpha_{n,k} = 0 \end{cases} \quad \forall n \in \{1,...,N\} \tag{10}$$

Additionally, for a given time granularity, expression (3) can be rewritten as follows, according to the energy model introduced in Section 2 (expression (1) and expression (2)).

$$\sum_{n=1}^{N} [C_n W_S + (1 - C_n) W_S^{sleep}] + a\ell + W_M \tag{11}$$

It has been shown in [9] that problems of this form are not guaranteed to exist a unique optimal solution. In this paper, we propose a heuristic macro-assisted sleep mode scheme based on the following heuristics:

- To minimize the power consumption of small cells, we have to minimize the total time that small cells spend in the on state.
- As we considered heterogeneous network that comprised of macro cell and small cells, we should take into account the macro cell energy consumption increase when a small cell is switched off.

The first heuristic can be achieved by finding and switching on the best small cell for user to connect to, so that the transmission of files from each small cell to users can be finished quickly to be able to put small cells to sleep state as soon as possible. A good way to find the best small cell for a user to connect to is to find small cell which providing the highest SINR to that user. To utilize the second heuristic, when a file arrives for a specific user, if the macro identifies there are UE-associated small cells in the sleep state, the macro cell can make the decision to make the UE connect to the small cell network or not based on the traffic load variation of macro base station. At a given timeslot, the power consumption that make the UE not connect to the small cell can be expressed as:

$$\sum_{m=1}^{M} W_S + a\left(\ell + \Delta\ell\right) + W_M \tag{12}$$

where M is the total number of small cells in the on state, $\Delta\ell$ is the traffic load variation of macro because of the macro cell assumes the load of the arrived file. The power consumption that make the UE connect to small cell is:

$$\sum_{m=1}^{M+1} W_S + a\ell + W_M \tag{13}$$

Therefore, we have:

$$\Delta E = W_S - a\Delta\ell \tag{14}$$

ΔE is the power consumption difference between UE connect to the small cell network and not. Thus the macro can make the decision to make the UE connect to the small cell network if $W_S < a\Delta\ell$, vice versa.

3.2 An Energy Efficient Macro Assisted Sleep Mode Scheme

Best Small Cell Selection and Small Cell Sleep to On Procedure

The small cell base stations can be configured to reside in the sleep state by default and move to the on state as explained next. The flow chart of our macro-assisted sleep mode scheme is illustrated in Figure 3 and the decision-making process is:

The transition of small cell from sleep to on state is controlled by the macro via the backhaul. In the macro-controlled small cells, UE first connect to the candidate macro cell. After connecting to the candidate macro cell, the UE will get time and frequency synchronization with the macro. In energy savings schemes presented in the literature, when there is any UE-associated small cell in sleep state, it is generally assumed that the macro changes an appropriate small cell to the on state and starts a UE-small cell connection procedure as soon as a connected UE is required to start receiving data [6], [7], [8], [9]. In this paper, we propose that this behavior is potentially suboptimal in terms of energy consumption, since in some cases a connection to a small cell is not needed. In our scheme, macro decide whether the UE connect to the small cell network or not based on expression (14): if $W_S < a\Delta\ell$, which means the power consumption that make the UE connect to the small cell network is less than make the UE stays connected to macro cell, and vice versa.

If macro decide to make UE connect to small cell network, the best small cell select procedure is performed. The decision is performed using the SINR values of UE small cell links, which can be obtained by information reporting from the small cells via their backhaul links. Then, macro builds a list of small cell candidates for the UE, and sorting them by UE-small cell link SINR values. Finally, chose the first small cell of the sleep list as the small cell for UE to connect to.

It should be noted that, $\Delta\ell$ which stands for the traffic variation of macro cell can be obtained by based on the size of the file to be received. Nevertheless, the size of a file is an application-layer piece of information, only available to the two ends of the communication, thus we maybe need cross-layer information exchange or obtaining the file size information via UE.

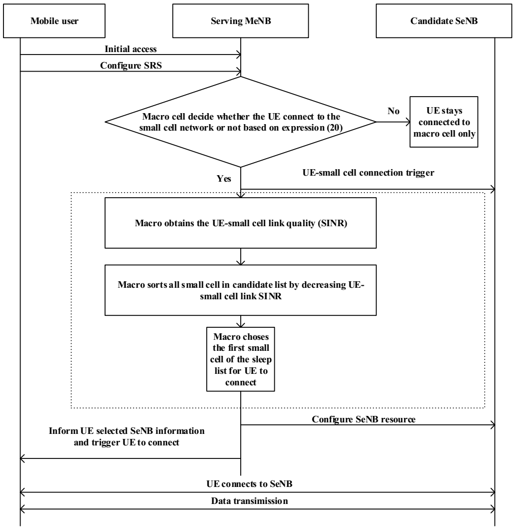

Mobile user
Serving MeNB
Candidate SeNB
Initial access
Configure SRS
Macro cell decide whether the UE connect to the small cell network or not based on expression (20)
No
UE stays connected to macro cell only
Yes
UE-small cell connection trigger
Macro obtains the UE-small cell link quality (SINR)
Macro sorts all small cell in candidate list by decreasing UE-small cell link SINR
Macro choses the first small cell of the sleep list for UE to connect
Configure SeNB resource
Inform UE selected SeNB information and trigger UE to connect
UE connects to SeNB
Data transimission

Fig. 3. Procedures of the proposed macro assisted sleep mode scheme used to determine the best small cell for a UE to connect to.

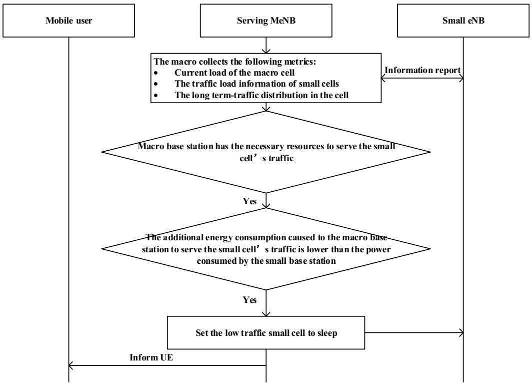

Mobile user
Serving MeNB
Small eNB
The macro collects the following metrics:
• Current load of the macro cell
• The traffic load information of small cells
• The long term-traffic distribution in the cell
Information report
Macro base station has the necessary resources to serve the small cell' s traffic
Yes
The additional energy consumption caused to the macro base station to serve the small cell' s traffic is lower than the power consumed by the small base station
Yes
Set the low traffic small cell to sleep
Inform UE

Fig. 4. Small cell On to Sleep Procedure

Small cell On to Sleep Procedure

A small cell may be putted to sleep state as soon as: (i) the macro base station has the necessary resources to serve the small cell's traffic; (ii) the additional energy consumption caused to the macro base station to serve the small cell's traffic when it is sleeping is lower than the power consumed by the small base station. The main characteristic of our macro assisted sleep mode scheme is that the state of the small cell can be determined by the macro cell which is different from the current LTE system. Macro cell can take into account the long-term traffic distribution in the cell, the state and capacity of small cells, and the moving speed of terminals. Thus, the optimized centralized decision can be achieved. The flow chart of our small cell on to sleep procedure we proposed is illustrated in Figure 4.

4 Evaluation Results

This Section focuses on the performance of the proposed scheme by estimation based on the research result. We consider specific values of the parameters introduced in Sections 2 and 3 according to [8]. According to [12], the number of pico base stations for a given capacity is shown in Table 1. According to [8], the parameters is shown in Table 2.

Table 1. Density of small cell base stations.

System Capacity (GB/H/km^2)	Small cell Density (BSs/km^2)
50	35
100	190
150	300
200	500
250	1000

Table 2. System Parameters values.

Parameter	Value
W_M	800W
a	500W
W_S	13W
W_0	4.3W

Using the traffic model provide in [8], we can calculate the data volume percentage of different traffics. Big files that satisfies $W_S < a\Delta\ell$ represented by FTP traffic currently take up to 22% of the total data volume. We choose random arrival traffic for study. Three schemes are compared: a) traditional no sleep scheme; b) macro starts a UE-small cell connection procedure as soon as a connected UE is required to start receiving data, as presented in the state-of-art sleep mode schemes [9], [10]; c) macro decides whether UE connect to small cells network or not based on the traffic load variation of macro cell, which is proposed in Section 3 in this paper

Figure 5 shows the energy per km^2 consumed in one day for the three considered schemes varying the capacity requirements. We notice that the use of sleep mode reduces the system energy consumption. Moreover, the proposed macro assisted sleep mode scheme in this paper gives a further reduction in the system energy consumption. Figure 6 shows the percentage of power savings by schemes b) and c) with respect to the fully on small cell scheme a). We notice that our scheme can yield energy savings of up to 10% compared to not using sleep mode, and can yield a further 5% energy savings with respect to the amount of energy savings obtained with existing sleep mode schemes.

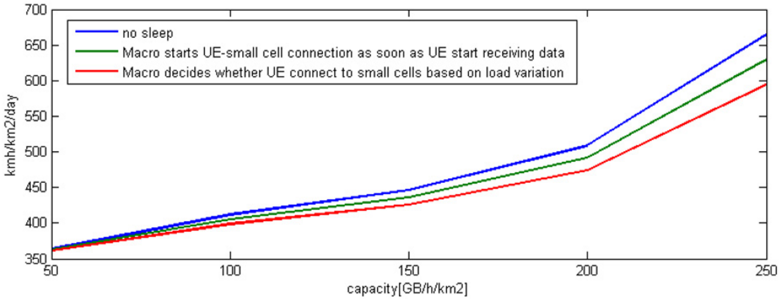


Fig. 5. Energy consumed for the three considered schemes varying the capacity requirements.

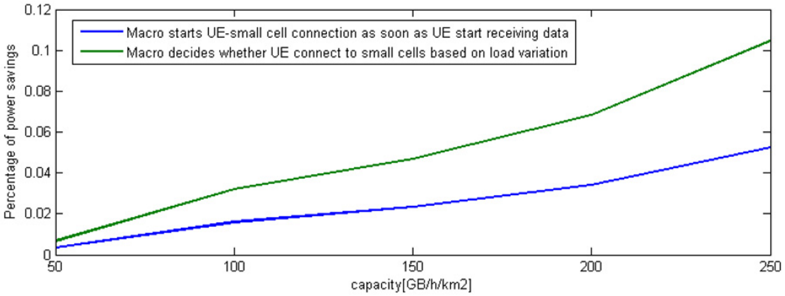


Fig. 6. Percentage of power savings by schemes b) and c) with respect to the fully on small cell scheme a).

Finally, we have to notice the fact that the power consumption of macro base stations has a significant dependency on the traffic load [8], assuming a constant macro base station power consumption leads to an overestimation of the energy savings with the use of sleep mode technology.

5 Conclusions

In this paper we have presented an energy efficient macro assisted sleep mode scheme in the macro-controlled heterogeneous networks. From the results of our analysis, we showed that the power consumption of macro base station has a significant dependency on the traffic load. Moreover, it is important to model the dependency of the macro base station power consumption on the traffic load to correctly determine whether the UE connect to the small cell network or not. By our calculation and analysis, our scheme can yield energy savings of up to 10% compared to not using sleep mode, and can yield a further 5% energy savings with respect to the amount of energy savings obtained with existing sleep mode schemes.

However, the proposed scheme in this paper is based on the assumption that the size of each arriving file is known by the macro cell serving the UE to which the file needs to be sent. This is difficult to implement in practice in the current LTE standard. In addition, energy-aware load balancing between macro and small cells still needs further study

Acknowledgement

This work is supported by the National Science and Technology Major Project (No. 2015ZX03001025).

References

1. H. Peng et al., “Ultra Dense Network: Challenges, Enabling Technologies and New Trends”, China Communications, Feb. 2016
2. N Bhushan, J Li, D Malladi et al., “Network densification: the dominant theme for wireless evolution into 5G”, Commun. Mag. IEEE. 52(2), 82–89 (2014)
3. PK Agyapong, M Iwamura et al., “Design considerations for a 5G network architecture”, IEEE Commun. Mag. 52(11), 65–75 (2014)
4. Kishiyama, Y.; Benjebbour, A.; Nakamura, T.; Ishii, H., “Future steps of LTE-A: evolution toward integration of local area and wide area systems,” Wireless Communications, IEEE, vol.20, no.1, pp.12, 18, February 2013
5. H Ishii, Y Kishiyama, H Takahashi, in Globecom Workshops (GC Wkshps), 2012 IEEE. A novel architecture for LTE-B: C-plane/U-plane split and Phantom Cell concept (Anaheim, CA, 2012), pp. 624–630
6. I. Ashraf, F. Boccardi, and L. Ho, “Sleep mode techniques for small cell deployments,” IEEE Commun. Mag., vol. 49, no. 8, pp. 72-79, Aug. 2011
7. S. Cai, L. Xiao, H. Yang, J. Wang, and S. Zhou, “A cross-layer optimization of the joint macro and picocell deployment with sleep mode for green communications,” in Proc. 22nd WOCC, May 2013, pp. 225-230
8. P. Dini, M. Miozzo, N. Bui, and N. Baldo, “A model to analyze the energy savings of base station sleep mode in LTE HetNets,” in Proc. IEEE GreenCom, IEEE iThings/CPSCom, Aug. 2013, pp. 1375-1380

9. E. Ternon, P. Agyapong, and A. Dekorsy, “Performance Evaluation of Macro-assisted Small Cell Energy Savings Schemes,” 2015, accepted for Publication to EURASIP Journal on Wireless Communications and Networking
10. Chang Liu et al., “Performance analysis of macro-assisted data-only carrier system in 5G energy-efficient heterogeneous networks,” IEEE GreenCom, Nov. 2014, pp. 1-6
11. Xing Zhang; Jiaxin Zhang; Wenbo Wang; Yan Zhang; Chih-Lin I, “Macro-assisted Data-only Carrier for Small Cell Enhancement in 5G Cellular Systems,” IEEE Commun. Mag., vol. 53, no. 5, pp. 223-231, May 2015
12. K.Hiltunen, “Comparison of different network densification alternatives from the lte downlink performance point of view,” in IEEE Vehicular Technology Conference (VTC Fall) 2011, San Francisco, US, Sept. 2011

A QoE Estimation Model for Video Streaming over 5G Millimeter Wave Network

Yanjun Hou, Wen’an Zhou, Lijun Song, Mengyu Gao
School of computer science, Beijing University of Posts andTelecommunications
Beijing, China
605575422@qq.com, zhouwa@bupt.edu.cn , songlj1992@gmail.com , itsmengyu@163.com

Abstract. With the rapid development of mobile communication, human's demand for communication capacity increasing, a fifth-generation communications network (5G) have gained popularity because of enormous amount of spectrum in the millimeter wave (mmWave) bands. The 5G network is not only to provide people with more high data transfer rates and lower latency, but also to provide users with more meaningful and personalized service based on the users and their understanding of the service required. Nowadays, video streaming service plays an increasing important role in human's daily life. However, the existing video streaming service application scenarios are mostly in Long Term Evolution network (LTE) and wireless network, there are few articles about studying video streaming service quality evaluation method in 5G network, so we conduct a study on the quality of experience (QoE) of video streaming service under 5G mmWave network scenario with NS-3. Under this scenario, video streaming system is created. By obtaining the quality of service (QoS) parameters of the network scenario, the non-linear regression function is used to predict the QoE of our video streaming service. The fit coefficient of the result shows that our model is powerful.

1 Introduction

With the popularity of the smart phone and smart tablet, the propagation of these smart devices has leads to an explosive increase of mobile traffic. According [1], the

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_9

traffic load on conventional cellular networks is predicted to be increased by 1000 times in the next 10 years [2]. To meet the need of the user's demand of the amount of mobile traffic to attain a high quality of experience , many researchers have proposed some new technologies to enhance the capacity of current mobile communication systems based on frequency bands below 3GHz such as carrier aggregation, multiple input and multiple output(MIMO)[3].But these new technologies cannot solve the traffic increasement problem fundamentally, and the capacity enhancement based on narrow bandwidth is easy to reach a limit. So it is emergency to study a wider bandwidth than that of existing mobile communication system. To solve the problem with the increasement of mobile traffic, some researchers and industrial experts have begun the study of millimeter wave band in the 30-300GHz that is a new radio band in emerging 5G mobile communication systems [4][5][6][7][8]. Because most of the millimeter waveband is underutilized, the 5G mobile communication system can make the best of the wide and continuous band to gain the higher bandwidth and bit rate.

With the popularity of mobile video streaming service, the Quality of Experience (QoE) of the service is also becoming more and more important. The competition among service providers is intensifying and becomes fiercer than ever before. In order to win the market in the fierce competition, the service provider must ensure that the service that they provided get a much higher acceptance. Thus they need a degree of user acceptance as a standard service rating method. Now, the QoE is a widely adopted service metric. According to the International Telecommunication Union-Telecommunication Standardization Sector ITU-T P.10/G.100 Recommendation, the QoE can be defined as “the overall acceptability of an application or service, as perceived subjectively by the end-user.” Based on this definition, more and more researchers have already begun the research on it. With the study on this field, the operators can master the key factor that affects the user evaluation and through the adjustment and optimization of the key factors, they can improve the quality of user experience and provide users with higher estimation of the service.

But now, the most of the existing research on QoE of the mobile streaming media focused on the existing mobile communication network and wireless network, few of them study the QoE method in 5G network. But the 5G millimeter wave network should be used in the near future and it can make the best of the wide and continuous band to obtain the higher bandwidth and bit rate gains. So in this paper, the main contributions can be summarized as follows: First, the millimeter wave network through ns-3 platform is built. Second, in the application layer we build video transmission system. Finally, Evalvid video evaluation tools and non-linear regression algorithm are used to evaluate the video QoE by using objective methods that reflect the QoS to QoE. The remainder of this paper is structured as follows. Section II reviews the current research works on 5G network simulation and QoE evaluation for video streaming service. In Section III, the architecture of video streaming system over 5G network is presented. Then in Section IV the result and analysis of the model is showed. Finally, the conclusion is stated in Section V.

2 Related work

2.1 5G millimeter wave simulation work with video streaming system

With the rapid growth of user demand for mobile traffic, spectrum shortages and the need of capacity requirement become a very obvious contradiction. Overcoming network bandwidth bottlenecks is a key issue in a fifth generation communications networks. On the other hand, since the bandwidth of the millimeter wave range from 30 to 300GHz, millimeter-wave communication become a hot topic in 5G network because of its enormous bandwidth [9].

O.Ayach et.al showed original theoretical results on the capacity and converge of cellular networks using millimeter wave [5]. K.Wang et al. have been pushing millimeter wave bands for 5G cellular networks with evidences of millimeter wave propagation measurements [6]. D.Love with his industrial team proposed to use mm-wave beam-forming both for access and backhaul in small cell networks [7]. W. Roh and his industrial team revealed the effectiveness of mm-wave beam-forming to improve system capacity of future cellular networks [8]. All researches are important for the study of millimeter wave network. But, the analyses in these papers are limited to link level or system level with homogeneous networks.

Menglei Zhang et.al in [10] [11] have established an end-to-end millimeter wave simulation system, the simulation system presented by TCP congestion avoidance algorithm control to show the performance of a millimeter wave link. The framework has implement the model spectrum millimeter wave beam-forming models, and can be quantified analysis in the millimeter-wave link, to achieve performance transport layer and application layer protocols. In our system, the simulation environment is to use millimeter wave link model to build video streaming client and server side, which achieves the goal of spread of video streaming [10] [11].

2.2 Quality of Experience of video streaming service

Future 5G networks will provide users with high bandwidth content, high speed in excess of 10Gbs, various mobility levels, and more solutions to save energy and cost, but the most important thing is that it must meet the satisfaction of users, the so called quality of experience [12]. Traditionally, in mobile communications and wireless networks the QoS is considered in order to evaluate the performance of the network given a service with a guaranteed service level [13]. QoS is a complex concept that includes network parameters such as packet loss, jitter, and delay. The operator obtained the networks parameters to predict the QoE of users.

Nowadays because the parameters and the mathematic way that researchers have adopted are different, there are many models to predict the QoE of users. Karan Mitra et.al in [14] proposed a model based on user behavior and QoE model context-aware and they use the P2P video streaming to complete the experiment. The system used user behaviors and context -related variables to model a network node in Bayesian network by building links between nodes from user behavior data and context data to draw the final QoE. Finally, they used obtained QoE value to optimize the network and got a high predict accurate. Y. Kanget.al in [15] present a no-reference, content-based Quality of Experience (QoE) estimation model for video streaming service over wireless networks. He and his team have used radial basis function networks (RBFN) which is a feed-forward artificial neural network with excellent approximating ability. The result shows that performs well in terms of high estimation accuracy, high Pearson correlation coefficient, low root mean square error, and small computational time. There are many researches in [16][17][18][19]also have proposed their way to show their accuracy. But nowadays researches are mostly focus on 3G, 4G or wireless network. They have not try to predict the QoE in 5G millimeter wave network. So we make our experiment to simulate millimeter wave network using ns-3 simulation tools to transmit our video under this network.

In summary, most of researches are focused on 3G, 4G or wireless network, they have not taken the 5G millimeter wave networks characterizes into consideration. Because of the enormous bandwidth of the 5G millimeter wave network, it will be utilized in the near future. So we begin the study on 5G millimeter wave network and by using the ns-3 simulation network to simulate 5G millimeter wave network and in this scenario I transmit the video streaming from the server to client successfully. Finally, from the parameters such as packet loss, jitter, delay, the non-linear regression method is used to build the parameters to our mean opinion score (MOS) that predict the QoE of our models.

3 The architecture of video streaming system over 5G network

Fig. 1 showed the architecture of our system. From the figure we can see it includes two parts, one is to build millimeter wave network by using ns-3 simulation tools. The other is that the estimation model for video streaming system that created in millimeter wave network with Evalvid tools. Then the two parts is introduced as follows.

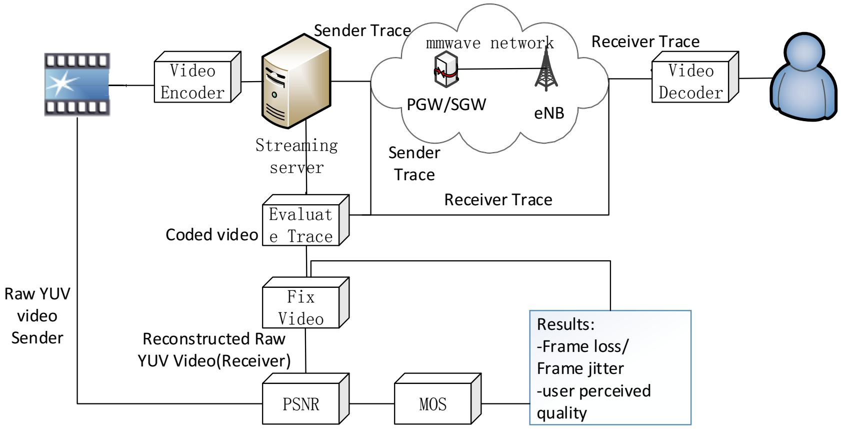


Fig. 1. Architecture of video streaming system over 5G network.

3.1 5G millimeter wave simulation network for our video streaming system

NS-3 is a discrete event simulator; it is beginning from an open source project in 2006[20]. The millimeter wave model of ns-3 is from a communication network center at New York University [10] [11].The entire model includes a basic implementation of millimeter wave devices, which comprises the propagation and channel model, the physical (PHY) layer, and the MAC layer. The module completely is developed with C++. It has a very robust architecture. The network model is described in Fig. 2.

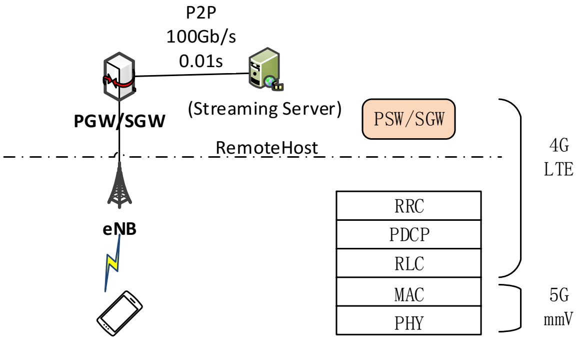


Fig. 2. 5G millimeter wave end-to-end network with video streaming system.

MAC: Millimeter wave MAC layer is designed to meet the ultra-low latency and high data rate demands, as presented in [21], thus following a flexible frame structure. A hybrid automatic repeat request (HARQ) is also implemented in order to better react to channel quality fluctuations.

PHY: Physical layer functions are: (1): a fully customizable time division duplex (TDD) frame structure and transport policy implementation, as shown in Table I. (2): to achieve channel model frame structure parameters, including millimeter wave transmission path loss model, MIMO technology to achieve beam-forming technology, the channel configuration parameters. (3) Realize the receiving end decoding error model. (4) Adaptive interference model based on CQI feedback loop to achieve [11]. By using large point-to-point MIMO technology, spatial multiplexing of a large number of data streams in wireless communications using millimeter wave network can be achieved. Through multiple antenas to achieve multiple input and multiple output, it can make full use of space resources and can exponentially increase the channel capacity without increasing the spectrum resources and the antenna transmit power. In the ns-3 module, it model the mmWave channel as a combination of cluster, each composed of several sub paths. And in order to support phased-array antennas, a new AntennaArrayModel class is developed, which contains a complex beam forming vector.

Table 1. Parameters for configuring the millimeter wave frame structure [11].

Parameter Name	Default Value	Description
SymbolPerSlot	30	Number of OFDM symbols per slot
SymbolLength	4.16μs	Length of one OFDM symbol in μs
SlotsPerSubframe	8	Number of slots in one subframe
SubframePerFrame	10	Number of subframes in one frame
NumReferenceSymbols	6	The number of reference OFDM symbols per slot
TDDControlDataPattern	"ccdddddd"	The control (c) and data(d) pattern
SubcarriersPerSubband	48	Number of subcarriers in each sub-band
SubbandsPerRB	18	Number of sub-bands in one resource block

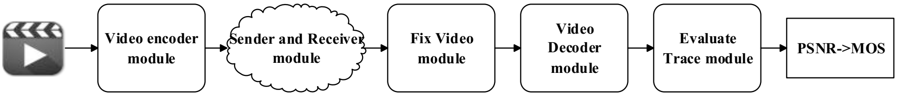


Fig. 3. Estimation model for video streaming system.

3.2 Estimation model for video streaming system with Evalvid tools

Our video streaming system includes five parts as we can see from the Fig. 3. Firstly, videos are encoded with videos encoder, then it is sent to client side from server side. After millimeter wave network transmission, users can receive the videos in the client side. Then we fix the videos from the trace file which is generated from the network and the sender file that are generated by Evalvid. Next, we will decode the videos to YUV file. Finally, Evalvid tool is used to calculate the PSNR of the videos and map the PSNR value to MOS. Thereinafter, we will describe the five parts in detail.

Video encoder module: The first step of our system is that we should encode our raw YUV format video to *.m4v with MPEF-4 encoder and then create *.mp4 file containing the video frames and a hint track which describes how to packetize the frames for transport with RTP.

Sender and Receiver module: The second step of the system is that we used mp4trace tool from Evalvid that produce the sender trace file named *.st that are transmitted to the client side over millimeter wave network.

Fix Video module: Then, we can reconstruct mp4 video from the sender trace and receiver trace file named *.mp4.

Video Decoder module: we will decode the mp4 file to YUV file with FFmpeg tool.

Evaluate Trace module: Finally, we will calculate packet / frame loss, delay, and jitter from three trace file that are sd_*, rd_* and *.st .By comparing the raw video and the reconstructed video, we can calculate the PSNR of the video and map it to the MOS value of the user as showed in Table II.

Table 2. PSNR-MOS.

PSNR	MOS	Estimated Quality
>37	5	Excellent
31-37	4	Good
25-31	3	Reasonable
20-25	2	Poor
<20	1	Bad

4 Results and Analysis

In our test, we select 12 videos that every video has two different samples because the network scenario has two different error model. And the video is 352X288 resolution, and decoded with mpeg-4 codec. The Table III shows a part of our test result.

Table 3. the Result of Video Test.

Video	Packet loss	Delay(s)	Jitter(s)	MOS
Australia	0.047	10.388	0.024	4.0967
City	0.105	10.009	0.017	2.0189
Coastguard	0.108	9.66	0.018	2.8936
Container	0.224	9.88	0.02	2.1188
Crew	0.055	11.1	0.033	3.8019
Football	0.123	11.1	0.03	2.6629
Foreman	0.065	1.05	0.021	2.8192

As we can see some videos which is transmitted from the millimeter wave network can get 4 grade. But most of video only can get 2 or 3 grade. Because the function that maps the parameters from the network level to MOS mostly are exponential function and liner function, which has a good performance on grade fitting. So we use the IQX hypothesis from the [15] to fit the data with the software matlab. In detail, we use multiple non-linear regression and exponential function to fit our data. The formula (1) is as follow:

$$\mathrm{QoE} = \alpha + \beta_1 * e^{\mathrm{loss}} + \beta_2 * e^{\mathrm{delay}} + \beta_3 * e^{\mathrm{jitter}}. \quad \textbf{(1)}$$

α、 β_1、 β_2、 β_3 are four unknown parameters . But at the beginning, we need to normalize the QoS parameters with formula (2).

$$y_i = \frac{x_i - \mathrm{MIN}_i}{\mathrm{MAX}_i - \mathrm{MIN}_i} \quad i \in \{\mathrm{loss}, \mathrm{delay}, \mathrm{jitter}\}. \quad \textbf{(2)}$$

Meanwhile, x_i 、 y_i respectively represent parameters before and after normalization, MAX_i、 MIN_i respectively represent the max value and min value among them.

Then we use regression function with matlab to get the parameters.Finally, we get the 0.8366 fit coefficient. And α、 β_1、 β_2、 β_3 respectively are 6.0148、-1.4010、-0.0898、-0.0585. From the parameters we know the packet loss and delay have greater effects on the QoE than delay and jitter.

Then we compare our QoE that we fit with the MOS that we map the PSNR to. Fig.5 described the relationship between them. From the test case we can clearly see that our fit QoE is closer to the MOS.

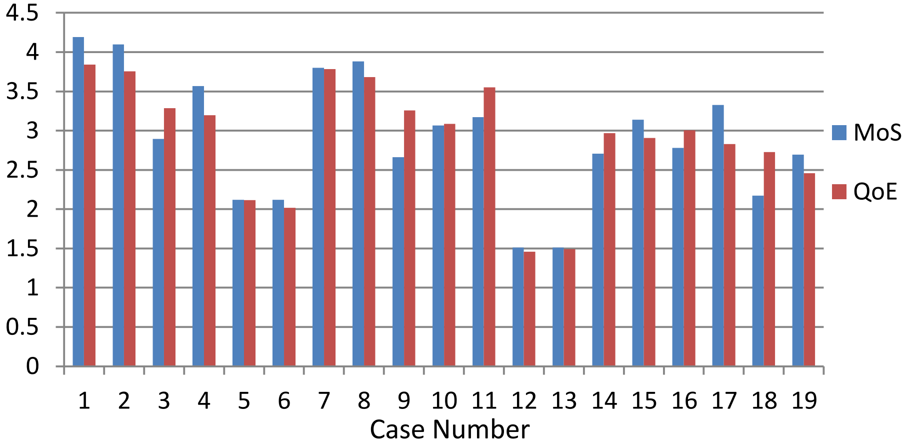


Fig. 5. Contract between the MOS and QoE.

5 Conclusions

In this paper, we propose a novel QoE model for evaluating video streaming in 5G millimeter wave scenario. We build an end-to-end streaming system on millimeter wave network with NS-3 simulation tools. Then we used Evalvid tool to obtain the video trace transmit from the millimeter wave network. By reconstructing the video and comparing the raw video and reconstructed video we get the PSNR value then map the PSNR value to MOS value. Finally, through non-linear regression and exponential function we get the formula to map the QoS parameters to MOS value. We get our QoE model for video estimation for 5G millimeter wave scenario. Because the millimeter wave link that we used is not mature enough, the packet loss and delay are unavoidable, so we get a little bit packet loss and a lot delay so that the fit coefficient is not very good. For future study, the link level to improve the system performance to transmit our video will be optimized and a more accurate estimation model for millimeter wave network will be built.

Acknowledgments: The research work reported in this paper is supported by the National Nature Science Foundation of China (61372113) and National Science and Technology Major Project (No.2015ZX03001025).

References

1. Cisco VNI Forecast, Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013—2018, Cisco Public Information, Feb. 2014.
2. Sakaguchi, Kei, et al. Millimeter-wave Evolution for 5G Cellular Networks. Ieice Transactions on Communications E98.B.3(2014):388-402.
3. Bae, Jung Sook, et al. Architecture and performance evaluation of Millimeter wave based 5G mobile communication system. International Conference on Information and Communication Technology Convergence 2014:847-851.
4. Pi, Zhouyue, and F. Khan. An introduction to millimeter-wave mobile broadband systems. IEEE Communications Magazine 49.6(2011):101-107.
5. S. Akoum, O. El Ayach,et al, Coverage and capacity in millimeter wave cellular systems, 2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), Pacific Grove, CA, 2012, pp. 688-692.
6. T. S. Rappaport, S. Sun,et al. Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!. Access IEEE 1.1(2013):335-349.
7. S. Hur, T. Kim, D. J. Love, et al. Millimeter Wave Beamforming for Wireless Backhaul and Access in Small Cell Networks. IEEE Transactions on Communications 61.10(2013):4391-4403.
8. T. Kim, J. Park, J. Seol, et al. Tens of Gbps support with millimeter wave beamforming systems for next generation communications. GLOBECOM 2013 - 2013 IEEE Global Communications Conference 2013:3685-3690.
9. Niu, Yong, et al. A survey of millimeter wave communications (millimeter wave) for 5G: opportunities and challenges. Wireless Networks 21.8(2015):2657-2676.
10. Zhang, M., Mezzavilla, et al. Transport Layer Performance in 5G millimeter wave Cellular. (2016).
11. Mezzavilla, M., Dutta, et al. 5G Millimeter wave Module for the ns-3 Network Simulator. ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems ACM, 2015:283-290.

12. Pierucci, Laura. The quality of experience perspective toward 5G technology. IEEE Wireless Communications 22.4(2015):10-16.
13. 3GPP Technical Specification TS 23.107 V11.0.0, Qualityof Services (QoS) Concept and Architecture, June 2012
14. Mitra, K. Zaslavsky, et al. Context-Aware QoE Modelling, Measurement, and Prediction in Mobile Computing Systems. Mobile Computing, IEEE Transactions on, vol.14, no.5, pp.920,936, May 2015 doi: 10.1109/TMC.2013.155
15. Y. Kang, H. Chen, and L. Xie, An artificial-neural-network-based QoE estimation model for video streaming over wireless networks, in Communications in China (ICCC), 2013 IEEE/CIC International Conference on. IEEE, 2013, pp. 264–269.
16. M. A. Santos, J. Villalón et al. A novel QoE aware multicast mechanism for video communications over IEEE 802.11WLANs, IEEE J. Sel. Areas Commun., vol. 30, no. 7, pp. 1205–1214,Aug. 2012
17. A. U. Mian, Z. Hu, et al. A decision theoretic approach for in-service QoE estimation and prediction of P2P live video streaming systems based on user behavior modeling and context awareness. JICS,vol. 10, no. 11, pp. 3429–3436, 2013.
18. T. De Pessemier,et al.Quantifying the influence of rebuffering interruptions on the user's quality of experience during mobile video watching. IEEE Trans.Broadcast., vol. 59, no. 1, pp. 47–61, Mar. 2013.
19. S.-O. Lee and D.-G. Sim. Hybrid bitstream-based video quality assessment method for scalable video coding. Opt. Eng., vol. 51, no. 6, pp. 067403-1–067403-9, Jun. 2012.
20. Ns-3 Network Simulator; https://www.nsnam.org/
21. S. Dutta, M. Mezzavilla, et al. Frame structure design and analysis for millimeter wave cellular systems. in arXiv:1512.05691 [cs.NI], Dec. 2015.

An Energy-Efficient Process Replication Algorithm in Virtual Machine Environments

Tomoya Enokido and Makoto Takizawa

Abstract Server cluster systems are widely used to realize fault-tolerant, scalable, and high performance application services with virtual machine technologies. In order to provide reliable application services, multiple replicas of each application process can be redundantly performed on multiple virtual machines. On the other hand, a server cluster system consumes a large amount of electric energy since multiple replicas of each application process are performed on multiple virtual machines. It is critical to discuss how to realize not only reliable but also energy-efficient server cluster systems. In this paper, we propose the redundant energy consumption laxity based (RECLB) algorithm to select multiple virtual machines for redundantly performing each application process in presence of server faults so that the total energy consumption of a server cluster and the average computation time of each process can be reduced. We evaluate the RECLB algorithm in terms of the total energy consumption of a server cluster and the average computation time of each process compared with the basic round-robin (RR) algorithm.

1 Introduction

Various types of business and industrial information services like data centers [10] require scalable, high performance, and fault-tolerant information systems like cloud computing systems [10]. These computing systems are realized virtual machines [5] with server cluster systems [2, 3, 4]. A server cluster system is composed

Tomoya Enokido
Faculty of Business Administration, Rissho University, Tokyo, Japan
e-mail: eno@ris.ac.jp

Makoto Takizawa
Department of Advanced Sciences, Faculty of Science and Engineering,
Hosei University, Tokyo, Japan
e-mail: makoto.takizawa@computer.org

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_10

of a large number of servers and multiple virtual machines are installed in each server in order to increase the resource utilization of servers. In fault-tolerant information systems, application processes which provide application services have to be reliably performed in presence of server faults [8]. One way to provide a fault-tolerant application service is that multiple replicas of each application process are performed on multiple virtual machines in a server cluster. However, a large amount of electric energy is consumed in a server cluster system since replicas of each application process are performed on multiple virtual machines which are performed on multiple servers. It is necessary to realize not only fault-tolerant but also energy efficient server cluster systems with virtual machines as discussed in Green computing [10].

In order to design energy-efficient server cluster systems [2, 3, 4], it is necessary to define a computation model and power consumption model of servers to perform application processes on virtual machines. In our previous studies [5], we measured power consumption of servers to perform *computation type application processes* (*computation processes*) which mainly consumes CPU resources of servers and derived the computation model of a virtual machine and power consumption model of a server to perform computation processes on multiple virtual machines from the experimentations. In this paper, we consider computation processes.

In this paper, we propose the *redundant energy consumption laxity based* (*RECLB*) algorithm to select multiple virtual machines for redundantly performing each application process in presence of server faults so that the total energy consumption of a server cluster and the average computation time of each process can be reduced. In this paper, we assume some servers in a server cluster might stop by fault. If a server stops by fault, every virtual machine performed on the server stops. Hence, replicas of each computation process have to be performed on multiple virtual machines which are performed on different servers in a serve cluster. Here, if at least one virtual machine is operational, a computation process is successfully performed even if some servers stop by fault. In the RECLB algorithm, a set of multiple virtual machines where the total energy consumption laxity of a server cluster is the minimum is selected for redundantly performing multiple replicas of each computation process. We evaluate the RECLB algorithm in terms of the total energy consumption of a homogeneous server cluster and computation time of each request process compared with the basic round-robin (RR) algorithm [9]. The evaluation results show the total energy consumption of a homogeneous server cluster and computation time of each request process can be more reduced in the RECLB algorithm than the RR algorithm.

In section 2, we define the computation model of a virtual machine and power consumption model of a server. In section 3, we discuss the RECLB algorithm. In section 4, we evaluate the RECLB algorithm compared with the RR algorithm.

2 System Model

2.1 Computation Model of a Virtual Machine

Let S be a cluster of servers s_1, ..., s_n ($n \geq 1$). Let C_t be a set of cores c_{1t}, ..., c_{lt} ($l \geq 1$) and nc_t be the total number of cores in a server s_t. We assume Hyper-Threading Technology [7] is enabled on a CPU. Let TH_t be a set of threads th_{1t}, ..., th_{qt} ($q \geq 1$) in a server s_t. Let ct_t be the number of threads on each core c_{ht} in a server s_t. Threads $th_{(h-1)\cdot ct_t+1}$, ..., $th_{h\cdot ct_t}$ ($1 \leq h \leq l$) are bounded to a core c_{ht}. Let nt_t be the total total number of threads in a server s_t, i.e. $nt_t = nc_t \cdot ct_t$. Let V_t be a set of virtual machines VM_{1t}, ..., VM_{qt} ($q \geq 1$) in a server s_t. Each virtual machine VM_{kt} holds one virtual CPU and is bounded to a thread th_{kt} in a server s_t. In this paper, we assume any virtual machine does not migrate to another server in a server cluster S. A virtual machine VM_{kt} is referred to as *active* iff (if and only if) the virtual machine VM_{kt} is initiated on a thread th_{kt} and at least one process is performed on the virtual machine VM_{kt}. A virtual machine VM_{kt} is *idle* iff the virtual machine VM_{kt} is initiated on a thread th_{kt} but no process is performed on the virtual machine VM_{kt}. A virtual machine is *stopped* iff the virtual machine is not initiated on any thread. A core c_{ht} is referred to as *active* iff at least one virtual machine VM_{kt} is active on a thread th_{kt} in the core c_{ht}. A core c_{ht} is *idle* if the core c_{ht} is not active.

We consider *computation processes*, where CPU resources are mainly consumed. A term *process* stands for a computation process in this paper. Let rd^i be the *redundancy* of a process p^i. A notation p^i_{kt} stands for a *replica* of a process p^i performed on a virtual machine VM_{kt}. On receipt of a process p^i, the load balancer K selects a set VMS^i ($|VMS^i| = rd^i$) of virtual machines in the server cluster S and forwards the process p^i to every virtual machines VM_{kt} in the set VMS^i as shown in Figure 1. We assume servers in a server cluster S might stop by fault. Let NF be the maximum number of servers which concurrently stop by fault in the cluster S. If a server s_t stops by fault, every virtual machine VM_{kt} performed on the server s_t stops. Hence, replicas of each process p^i have to be performed on rd^i virtual machines performed on different servers in a server cluster S. We assume $NF + 1 \leq rd^i \leq n$ and replicas of each process p^i are performed on rd^i virtual machines performed on different servers. This means, each client cl^i can receive at least one reply r^i_{kt} from a virtual machine VM_{kt} even if NF servers stop by fault in a server cluster S. On receipt of a process p^i, a replica p^i_{kt} is created and performed on a virtual machine VM_{kt}. Then, the virtual machine VM_{kt} sends a reply r^i_{kt} to the load balancer K. The load balancer K takes only the first reply r^i_{kt} and ignores every other reply.

Replicas which are being performed and already terminate at time τ are *current* and *previous*, respectively. Let $CP_{kt}(\tau)$ be a set of current replicas on a virtual machine VM_{kt} at time τ and $NC_{kt}(\tau)$ be $|CP_{kt}(\tau)|$. Let T^i_{kt} be the total computation time of a replica p^i_{kt} [msec]. $minT^i_{kt}$ shows the minimum computation time of a replica p^i_{kt} where the replica p^i_{kt} is exclusively performed on a virtual machine VM_{kt} and the other virtual machines are not active in a server s_t. We assume $minT^i_{1t} = minT^i_{2t} = \cdots = minT^i_{qt}$ in a server s_t, i.e. the maximum computation rate of every virtual machine

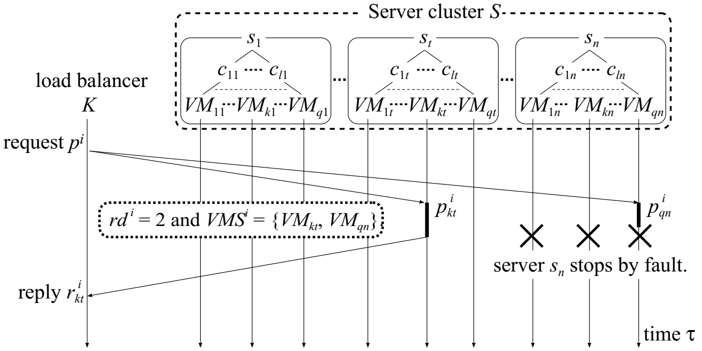


Fig. 1 System model.

VM_{kt} in the server s_t is the same. $minT^i = min(minT^i_{k1}, ..., minT^i_{kn})$. $minT^i = minT^i_{kt}$ on the fastest server s_t. We assume one virtual computation step is performed for one time unit on a virtual machine VM_{kt} in the fastest server s_t. That is, the maximum computation rate $Maxf_{kt}$ of the fastest virtual machine VM_{kt} is 1 [vs/msec]. We assume $Maxf_{1t} = Maxf_{2t} = \cdots = Maxf_{qt}$ in a server s_t. $Maxf = max(Maxf_{k1}$, ..., $Maxf_{kn})$. A replica p^i_{kt} is considered to be composed of VS^i_{kt} virtual computation steps. $VS^i_{kt} = minT^i_{kt} \cdot Maxf = minT^i_{kt}$ [vs].

The computation rate $f^i_{kt}(\tau)$ of a replica p^i_{kt} performed on a virtual machine VM_{kt} at time τ is defined as follows [5]:

$$f^i_{kt}(\tau) = \alpha_{kt}(\tau) \cdot VS^i/(minT^i_{kt} \cdot NC_{kt}(\tau)) \cdot \beta_{kt}(nv_{kt}(\tau)). \tag{1}$$

Here, $\alpha_{kt}(\tau)$ is the *computation degradation ratio* of a virtual machine VM_{kt} at time τ $(0 \leq \alpha_{kt}(\tau) \leq 1)$. $\alpha_{kt}(\tau_1) \leq \alpha_{kt}(\tau_2) \leq 1$ if $NC_{kt}(\tau_1) \geq NC_{kt}(\tau_2)$. $\alpha_{kt}(\tau) = 1$ if $NC_{kt}(\tau) \leq 1$. Here, $\alpha_{kt}(\tau)$ is assumed to be $\varepsilon_{kt}^{NC_{kt}(\tau)-1}$ where $0 \leq \varepsilon_{kt} \leq 1$. The maximum computation rate $maxf^i_{kt}$ of a replica p^i_{kt} is VS^i_{kt} / $minT^i_{kt}$ $(0 \leq f^i_{kt}(\tau) \leq maxf^i_{kt}$ $\leq 1)$ where the replica p^i_{kt} is exclusively performed on a virtual machine VM_{kt} and only VM_{kt} is active on a core. Let $nv_{kt}(\tau)$ be the number of active virtual machines on a core which performs a virtual machine VM_{kt} at time τ. Let $\beta_{kt}(nv_{kt}(\tau))$ be the *performance degradation ratio* of a virtual machine VM_{kt} at time τ $(0 \leq \beta_{kt}(nv_{kt}(\tau))$ $\leq 1)$ where multiple virtual machines are active on the same core. $\beta_{kt}(nv_{kt}(\tau)) = 1$ if $nv_{kt}(\tau) = 1$. The formula (1) means the computation rate $f^i_{kt}(\tau)$ of each replica p^i_{kt} performed on a virtual machine VM_{kt} at time τ decreases as the number of current replicas increases on the virtual machine VM_{kt}. The computation rate of a virtual machine decreases as the number of active virtual machines increases on the same core. The computation rate of a virtual machine performed on a core c_{ht} is independent of active virtual machines performed on another core c_{ft} $(h \neq f)$.

Suppose that a replica p^i_{kt} starts and terminates on a virtual machine VM_{kt} at time st^i_{kt} and et^i_{kt}, respectively. Here, $T^i_{kt} = et^i_{kt}$ - st^i_{kt} and $\sum_{\tau=st^i_{kt}}^{et^i_{kt}} f^i_{kt}(\tau) = VS^i$. At time st^i_{kt} a replica p^i_{kt} starts, the computation laxity $lc^i_{kt}(\tau) = VS^i$ [vs]. The computation laxity $lc^i_{kt}(\tau)$ [vs] of a replica p^i_{kt} at time τ is given as follows:

$$lc^i_{kt}(\tau) = VS^i - \sum_{x=st^i_{kt}}^{\tau} f^i_{kt}(x). \tag{2}$$

2.2 Power Consumption Model of a Server

$E_t(\tau)$ shows the electric power [W] of a server s_t at time τ. Let $maxE_t$ and $minE_t$ be the maximum and minimum electric power [W] of a server s_t, respectively. Let $ac_t(\tau)$ be the number of active cores in a server s_t at time τ. $minC_t$ shows the electric power [W] where at least one core c_{ht} is active on a server s_t. Let $mv_t(\tau)$ be the number of virtual machines which concurrently migrate from a server s_t to the other servers at time τ. Let cE_t be the electric power [W] where one core gets active on a server s_t. Let mE_t be the electric power [W] where one virtual machine migrates from a server s_t to the other server s_u.

The electric power $E_t(\tau)$ [W] of a server s_t to perform processes on virtual machines at time τ is given as follows [5]:

$$E_t(\tau) = minE_t + \sigma_t(\tau) \cdot (minC_t + ac_t(\tau) \cdot cE_t) + \lambda_t(\tau) \cdot (mv_t(\tau) \cdot mE_t). \tag{3}$$

Here, $\sigma_t(\tau) = 1$ if at least one core c_{ht} is active on a server s_t at time τ. Otherwise, $\sigma_t(\tau) = 0$. $\lambda_t(\tau) = 1$ if at least one virtual machine migrates from a server s_t to another server at time τ. Otherwise, $\lambda_t(\tau) = 0$.

The total energy consumption $TE_t(\tau_1, \tau_2)$ [Ws] of a server s_t from time τ_1 to τ_2 is $\int_{\tau 1}^{\tau 2} E_t(\tau)\, d\tau$. The processing power $PE_t(\tau)$ [W] of a server s_t at time τ is $E_t(\tau)$ - $minE_t$. The total processing energy consumption $TPE_t(\tau_1, \tau_2)$ of a server s_t from time τ_1 to τ_2 is $\int_{\tau 1}^{\tau 2} PE_t(\tau)\, d\tau$.

3 Selection Algorithm

The total processing energy consumption laxity $tpel_t(\tau)$ [Ws] shows how much electric energy a server s_t has to consume to perform every current replica on every active virtual machine in the server s_t at time τ. Suppose a load balancer K receives a new request process p^{new} and allocates a replica p^{new}_{kt} to a virtual machine VM_{kt} performed on a server s_t at time τ. Here, the replica p^{new}_{kt} is added to the current replica set $CP_{kt}(\tau)$ of a virtual machine VM_{kt}, i.e. $CP_{kt}(\tau) = CP_{kt}(\tau) \cup \{p^{new}_{kt}\}$. Let $\mathbf{CP}_t(\tau)$ be a family $\{CP_{1t}(\tau), ..., CP_{qt}(\tau)\}$ of current replica sets of every virtual machine VM_{kt} in a server s_t at time τ. The total energy consumption laxity $tpel_t(\tau)$ of a server s_t at time τ is obtained by the **ELaxity**(s_t, τ) procedure [6]:

```
ELaxity($s_t$, $\tau$) { /* a term VM stands for a virtual machine. */
    if every $CP_{kt}(\tau) = \phi$ in $\mathbf{CP}_t(\tau)$ and $mv_t(\tau) = 0$, return(0);
    for each core $c_{ht}$ in a server $s_t$, {
        $nv$ = the number of active VMs on $c_{ht}$ at time $\tau$;
        $mv$ = the number of VMs migrating from $c_{ht}$ at time $\tau$;
        if $nv \geq 1$, $ac_t(\tau) = ac_t(\tau) + 1$; /* count of active cores on $s_t$.*/
        $mv_t(\tau) = mv_t(\tau) + mv$; /* count of VMs migrating from $s_t$.*/
        for each $VM_{kt}$ on a core $c_{ht}$, {
            for each $p^i_{kt} \in CP_{kt}(\tau)$, {
                $lc^i_{kt}(\tau + 1) = lc^i_{kt}(\tau) - f^i_{kt}(\tau)$;
                if $lc^i_{kt}(\tau + 1) \leq 0$, $CP_{kt}(\tau) = CP_{kt}(\tau) - \{p^i_{kt}\}$;
            }
        }
    }
    $tpel_t(\tau) = E_t(\tau) - minE_t$; /* processing power consumption */
    return($tpel_t(\tau) + ELaxity(s_t, \tau + 1)$);)
}
```

We discuss the *redundant energy consumption laxity based* (*RECLB*) algorithm to select multiple virtual machines to redundantly perform each process in presence of server fault so that the total processing energy consumption of a server cluster and average computation time of each replica can be reduced. Let $TPE^S_{kt}(\tau)$ be the total processing energy consumption laxity of a server cluster S where a replica p^i_{kt} of a new request process p^i is allocated to a virtual machine VM_{kt} at time τ. Suppose a load balancer K receives a new request process p^i at time τ. Then, the load balancer K selects a set VMS^i of rd^i virtual machines in a server cluster S by the following procedure **RECLB**(p^i, τ):

```
RECLB($p^i$, $\tau$) { /* a term VM stands for a virtual machine. */
    $VMS^i = \phi$;
    while ($rd^i > 0$) {
        for each $VM_{kt}$ in a server cluster $S$, {
            $CP_{kt}(\tau) = CP_{kt}(\tau) \cup \{p^i\}$;
            $TPE^S_{kt}(\tau) = \sum_{t=1}^{n}$ ELaxity($s_t$, $\tau$);
        }
        $vm$ = a virtual machine $VM_{kt}$ where $TPE^S_{kt}(\tau)$ is the minimum;
        $VMS^i = VMS^i \cup \{vm\}$;
        $S = S - \{s_t\}$; /* the server $s_t$ which performs the virtual machine $vm$ is removed. */
        $rd^i = rd^i - 1$;
    }
    return($VMS^i$);
}
```

Suppose there are three servers s_1, s_2, and s_3 in a server cluster S, i.e. $S = \{s_1, s_2, s_3\}$. Each server s_t is equipped with a single-core CPU and two threads th_{1t} and th_{2t} are bounded to the single-core. Hence, two virtual machines VM_{1t} and VM_{2t} are performed on each server s_t. Suppose a load balancer K receives a new request process p^i at time τ and the redundancy rd^i of the process p^i is two (rd^i = 2). Then, the load balancer K calculates the total processing energy consumption laxity $TPE^S_{kt}(\tau)$ where a replica p^i_{kt} of the process p^i is allocated to each virtual machine VM_{kt} ($k = \{1, 2\}$ and $t = \{1, 2, 3\}$) according to the **ELaxity**(s_t, τ) procedure. Suppose $TPE^S_{11}(\tau)$ is the minimum, i.e. $TPE^S_{11}(\tau) \leq TPE^S_{kt}(\tau)$ ($k = \{1, 2\}$ and $t = \{1, 2, 3\}$). Then, the load balancer K includes a virtual machine VM_{11} into the set VMS^i (= $\{VM_{11}\}$) and removes the server s_1 which performs the virtual machine VM_{11} from the server set S (= $\{s_2, s_3\}$). Next, the load balancer K calculates the total processing energy consumption laxity $TPE^S_{kt}(\tau)$ where a replica p^i_{kt} is allocated to each virtual machine VM_{kt} ($k = \{1, 2\}$ and $t = \{2, 3\}$) in the server set S (= $\{s_2, s_3\}$). Suppose $TPE^S_{23}(\tau)$ is the minimum. Then, the load balancer K includes a virtual machine VM_{23} into the set VMS^i (= $\{VM_{11}, VM_{23}\}$) and removes the server s_3 which performs the virtual machine VM_{23} from the server set S (= $\{s_2\}$). Here, $rd^i = |VMS^i| = 2$. The load balancer K forwards the process p^i to a pair of virtual machines VM_{11} and VM_{23} in the set VMS^i.

4 Evaluation

We evaluate the RECLB algorithm in terms of the total processing energy consumption of a homogeneous server cluster S and average computation time of each process p^i compared with the basic round-robin (RR) [9] algorithm.

A homogeneous cluster S is composed of five servers $s_1, ..., s_5$ (n = 5) as shown in Table 1. Every server s_t ($1 \leq t \leq 5$) follows the same computation model and the same power consumption model. Every server s_t is equipped with a dual-core CPU (nc_t = 2). Hyper-Threading Technology [7] is enabled on a CPU in every server s_t. Two threads are bounded for each core in a server s_t, i.e. ct_t = 2. The number of threads nt_t in each server s_t is four, i.e. $nt_t = nc_t \cdot ct_t = 2 \cdot 2 = 4$. $minE_t$ = 14.8 [W], $minC_t$ = 6.3 [W], cE_t = 3.9 [W], mE_t = 1.25 [W], and $maxE_t$ = 33.8 [W]. The parameters of each server s_t are obtained from the experiment [5]. Each virtual machine VM_{kt} holds one virtual CPU and is bounded to a thread th_{kt} in a server s_t (k = 1, ..., 4 and t = 1, ..., 5). Hence, there are twenty virtual machines in the server cluster S. Every virtual machine VM_{kt} follows the same computation model as shown in Table 1. The maximum computation rate $Maxf_{kt}$ is 1 [vs/msec]. The parameter ε_{kt} in the computation degradation ratio $\alpha_{kt}(\tau)$ is 1. The performance degradation ratios $\beta_{kt}(1)$ = 1 and $\beta_{kt}(2)$ = 0.5. The parameters of each server s_t and each virtual machine VM_{kt} are obtained from the experiment [5]. We assume the fault probability fr_t for every server s_t is the same fr = 0.1. We assume any virtual machine does not migrate to the other servers.

Table 1 Parameters of each server s_t and each virtual machine VM_{kt}.

Server	nc_t	ct_t	nt_t	$minE_t$	$minC_t$	cE_t	mE_t	$maxE_t$
s_t	2	2	4	14.8 [W]	6.3 [W]	3.9 [W]	1.25 [W]	33.8 [W]

Virtual machine	$Maxf_{kt}$	ε_{kt}	$\beta_{kt}(1)$	$\beta_{kt}(2)$
VM_{kt}	1 [vs/msec]	1	1	0.5

(k = 1, ..., 4, t = 1, ..., 5, and $minT^i$ = 1 [msec]).

The number m of processes p^1, ..., p^m ($0 \leq m \leq 10{,}000$) are issued in the simulation. The starting time of each process p^i is randomly selected in a unit of one millisecond between 1 and 60 [sec]. The minimum computation time $minT^i$ of every process p^i is assumed to be 1 [msec]. This means it takes one milli-second to exclusively perform a replica p^i_{kt} on a virtual machine VM_{kt} if the other virtual machines are not active on the same core where the virtual machine VM_{kt} is performed. We assume the redundancy rd^i for each process p^i is the same rd (= $\{1, 2, 3, 4, 5\}$) and N_{fault} = rd - 1 holds.

Let $TPEC^m_{tm}(rd)$ be the total processing energy consumption [Ws] to perform the total number m of processes ($0 \leq m \leq 10{,}000$) with redundancy rd (= $\{1, 2, 3, 4, 5\}$) obtained in the tmth simulation. The total processing energy consumption $TPEC^m_{tm}(rd)$ is measured 5 times for each redundancy rd (= $\{1, 2, 3, 4, 5\}$) and each number m of processes. The average total processing energy consumption $ATPEC^m(rd)$ [Ws] to perform the total number m of processes with redundancy rd is calculated as $\sum^5_{tm=1} TPEC^m_{tm}(rd)$ / 5. Here, $ATPEC^m_{algo}(rd)$ stands for the average total processing energy consumption with an algorithm type $algo \in$ {RECLB, RR} to perform the total number m of processes with redundancy rd. Figure 2 shows the average total processing energy consumption $ATPEC^m_{algo}(rd)$ of the RECLB and RR algorithms where the fault probability fr for every server s_t is 0.1. In the RECLB and RR algorithms, the average total processing energy consumption increases as the number m of processes increases. In the RECLB algorithm, a virtual machine VM_{kt} where the total processing energy consumption laxity of a server cluster S is the minimum is selected for each replica. Hence, the average total processing energy consumption to perform the number m of processes can be more reduced in the RECLB algorithm than the RR algorithm for each redundancy rd.

The computation time T^i for each process p^i is the computation time T^i_{kt} of a replica p^i_{kt} earliest committed in the replicas of the process p^i. The computation time T^i for each process p^i is measured 5 times for each redundancy rd and each number m of processes. Let $T^{i,m}_{tm}(rd)$ be the computation time T^i [msec] of a process p^i obtained in the tm-th simulation for redundancy rd and total number m of processes. Here, $AT^m_{algo}(rd)$ stands for the average computation time [msec] of each process with an algorithm type $algo \in$ {RECLB, RR} to perform the total number m of processes with redundancy rd. The average computation time $AT^m_{algo}(rd)$ is calculated as $\sum^5_{tm=1}\sum^m_{i=1} T^{i,m}_{tm}(rd)$ / ($m \cdot 5$). Figure 3 shows the average computation time

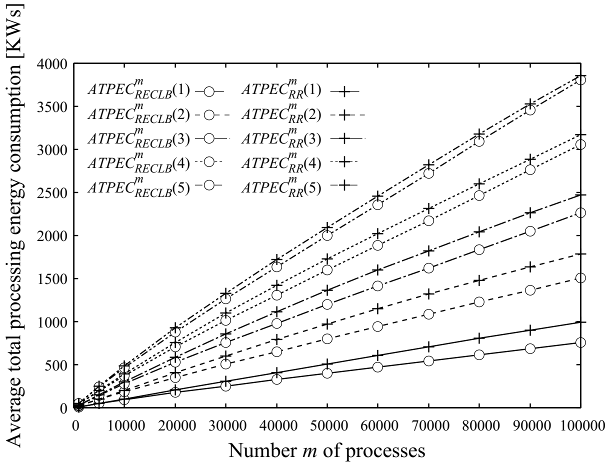


Fig. 2 Average total processing energy consumption $ATPEC^m_{algo}(rd)$ in the homogeneous server cluster S (number m of processes are issued in 60 [sec] and $fr = 0.1$).

$AT^m_{algo}(rd)$ in the RECLB and RR algorithms where the fault probability fr for every server s_t is 0.1. The average computation time $AT^m_{algo}(rd)$ increases as the number m of processes and redundancy rd increase in the RECLB and RR algorithms. For $1 \leq rd \leq 4$ and $0 \leq m \leq 10{,}000$, the average computation time in the RECLB algorithm is the same as the RR algorithm. For $rd = 5$ and $0 \leq m \leq 10{,}000$, the average computation time in the RECLB algorithm can be more reduced than the RR algorithm since the computation resources in the homogeneous server cluster S can be more efficiently utilized in the RECLB algorithm than the RR algorithm.

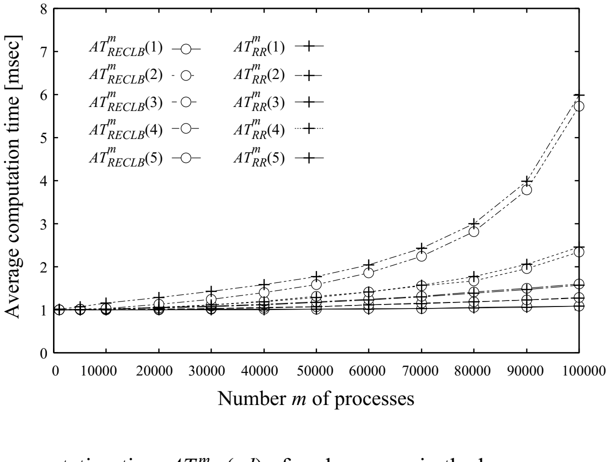


Fig. 3 Average computation time $AT^m_{algo}(rd)$ of each process in the homogeneous server cluster S (number m of processes are issued in 60 [sec] and $fr = 0.1$).

Following the evaluation, we conclude the RECLB algorithm is more useful in a homogeneous server cluster than the RR algorithm.

5 Concluding Remarks

In this paper, we proposed the RECLB algorithm to select multiple servers for redundantly performing each computation process issued by a client in presence of server fault so that the total energy consumption of a server cluster and the average computation time of each process can be reduced. In the RECLB algorithm, a set of multiple virtual machines where the total processing energy consumption laxity of a server cluster is the minimum is selected to perform multiple replicas of each process. We evaluated the RECLB algorithm in terms of the total energy consumption of a homogeneous server cluster and computation time of each process compared with the RR algorithm. The average total processing energy consumption of the homogeneous server cluster and computation time of each process are shown to be more reduced in the RECLB algorithm than the RR algorithm.

References

1. Enokido, T., Aikebaier, A., and Takizawa, M.: A model for reducing power consumption in peer-to-peer systems. IEEE Systems Journal, **4**(2), pp. 221–229, (2010).
2. Enokido, T., Aikebaier, A., and Takizawa, M.: Process allocation algorithms for saving power consumption in peer-to-peer systems. IEEE Trans. on Industrial Electronics, **58**(6), pp. 2097–2105, (2011).
3. Enokido, T. and Takizawa, M.: Integrated power consumption model for distributed systems IEEE Trans. on Industrial Electronics, **60**(2), pp. 824–836, (2013).
4. Enokido, T., Aikebaier, A., and Takizawa, M.: An extended simple power consumption model for selecting a server to perform computation type processes in digital ecosystems. IEEE Trans. on Industrial Informatics, **10**(2), pp. 1627–1636, (2014).
5. Enokido, T. and Takizawa, M.: Power consumption and computation models of virtual machines to perform computation type application processes. Proc. of the 9th International Conference on Complex, Intelligent and Software Intensive Systems (CISIS-2015), pp. 126–133, (2015).
6. Enokido, T. and Takizawa, M.: An energy-efficient load balancing algorithm to perform computation type application processes for virtual machine environments. Proc. of the 18th International Conference on Network-Based Information Systems (NBiS-2015), pp. 32–39, (2015).
7. Intel: Intel xeon processor 5600 series : the next generation of intelligent server processors. http://www.intel.com/content/www/us/en/processors/xeon/xeon-5600-brief.html, (2010).
8. Lamport, R., Shostak, R., and Pease, M.: The byzantine generals problems. ACM Trans. on Programing Language and Systems, **4**(3), pp.382–401, (1982).
9. LVS project: Job scheduling algorithms in linux virtual server. http://www.linuxvirtual server.org/docs/scheduling.html, (2010).
10. Natural Resources Defense Council (NRDS): Data center efficiency assessment - scaling up energy efficiency across the data center lndustry: Evaluating key drivers and barriers -. http://www.nrdc.org/energy/files/data-center-efficiency-assessment-IP.pdf, (2014).

Comparison Analysis by WMN-GA Simulation System for Different WMN Architectures, Distributions and Routing Protocols Considering TCP

Tetsuya Oda, Admir Barolli, Ryoichiro Obukata, Leonard Barolli, Fatos Xhafa and Makoto Takizawa

Abstract In this paper, we evaluate the performance of two WMN architectures considering Packet Delivery Ratio (PDR), throughput, delay and energy metrics. For simulations, we used ns-3 and Transmission Control Protocol (TCP). We compare the performance for Hybrid Wireless Mesh Protocol (HWMP) and Optimized Link State Routing (OLSR) for normal and uniform distributions of mesh clients by sending multiple Constant Bit Rate (CBR) flows in the network. The simulation results show that the PDR for both distributions and architectures is almost the same, but the PDR of HWMP is a little bit better than OLSR. the throughput is better for normal distribution and I/B WMN architecture in case of HWMP. However, for OLSR and normal distribution, the throughput of Hybrid WMN is a little bit higher than I/B WMN. For both distributions the delay of both architectures is better for HWMP compared with OLSR. For both WMN architectures and routing protocols for normal and uniform distributions, respectively. For normal distribution, the energy decreases sharply, because of the high density of nodes, thus the nodes spend

Tetsuya Oda and Leonard Barolli
Department of Information and Communication Engineering, Fukuoka Institute of Technology (FIT), 3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811–0295, Japan e-mail: oda.tetsuya.fit@gmail.com, barolli@fit.ac.jp

Admir Barolli
Department of Information Technology, Aleksander Moisiu University of Durres, L.1, Rruga e Currilave, Durres, Albania e-mail: admir.barolli@gmail.com

Ryoichiro Obukata
Graduate School of Engineering, Fukuoka Institute of Technology (FIT), 3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811–0295, Japan e-mail: obukenkyuu@gmail.com

Fatos Xhafa
Technical University of Catalonia Department of Languages and Informatics Systems C/Jordi Girona 1-3, 08034 Barcelona, Spain e-mail: fatos@lsi.upc.edu

Makoto Takizawa
Hosei University, 3-7-2, Kajino-Machi, Koganei-Shi, Tokyo 184–8584, Japan e-mail: makoto.takizawa@computer.org

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_11

more energy. So, the uniform distribution has better performance compared with normal distribution considering the energy parameter.

1 Introduction

Wireless Mesh Networks (WMNs) can be seen as a special type of wireless ad-hoc networks [1]. WMNs are based on mesh topology, in which every node (representing a server) is connected through wireless links to one or more nodes, enabling thus the information transmission in more than one path. The path redundancy is a robust feature of mesh topology. Compared to other topologies, mesh topology does not need a central node, allowing networks based on it to be self-healing. These characteristics of networks with mesh topology make them very reliable and robust networks to potential server node failures.

There are a number of application scenarios for which the use of WMNs is a very good alternative to offer connectivity at a low cost. It should also mentioned that there are applications of WMNs which are not supported directly by other types of wireless networks such as cellular networks, ad hoc networks, wireless sensor networks and standard IEEE 802.11 networks. There are many applications of WMNs in Neighboring Community Networks, Corporative Networks, Metropolitan Area Networks, Transportation Systems, Automatic Control Buildings, Medical and Health Systems, Surveillance and so on.

In WMNs, the mesh routers provide network connectivity services to mesh client nodes. The good performance and operability of WMNs largely depends on placement of mesh routers nodes in the geographical deployment area to achieve network connectivity, stability and client coverage.

In our previous work [2, 3], we considered the version of the mesh router nodes placement problem in which we are given a grid area where to deploy a number of mesh router nodes and a number of mesh client nodes of fixed positions (of an arbitrary distribution) in the grid area. We used WMN-GA system to optimize the location of mesh routers the network connectivity.

In this work, we use the topology generated by WMN-GA system and evaluate by simulations the performance of two different routing protocols and distributions of mesh clients considering two architectures of WMNs by sending multiple Constant Bit Rate (CBR) flow in the network. For simulations, we use ns-3 and Transmission Control Protocol (TCP). As evaluation metrics we considered PDR, throughput, delay and energy.

The structure of the paper is as follows. In Section 2, we discuss the related work. In Section 3, we make an explanation of architectures of WMNs. In Section 4, we make an overview of Hybrid Wireless Mesh Protocol (HWMP) and Optimized Link State Routing (OLSR) routing protocols. In Section 5, we show the description and design of the simulation system. In Section 6, we show the simulation results. Finally, conclusions and future work are given in Section 7.

2 Related Work

Until now, many researchers performed valuable research in the area of multi-hop wireless networks by computer simulations and experiments [4]. Most of them are focused on throughput improvement and they do not consider mobility [5].

WMNs are attracting a lot of attention from wireless research. Node placement problems have been investigated for a long time in the optimization field due to numerous applications in location science (facility location, logistics, services, etc.).

The main issue of WMNs is to achieve network connectivity and stability as well as QoS in terms of user coverage. Several heuristic approaches are found in the literature for node placement problems in WMNs [6, 7, 8, 9]. As node placement problems are known to be computationally hard to solve for most of the formulations [10, 11], GAs have been recently investigated as effective resolution methods. However, GAs require the user to provide values for a number of parameters and a set of genetic operators to achieve the best GA performance for the problem [12, 13, 14, 15, 16, 17, 18].

3 Architectures of WMNs

In this section, we describe the architectures of WMN. The architecture of the nodes in WMNs [19] can be classified according to the functionalities they offer as follows:

Infrastructure/Backbone WMNs: This type of architecture (also known as infrastructure meshing) is the most used and consists of a grid of mesh routers which are connected to different clients. Moreover, routers have gateway functionality thus allowing Internet access for clients. This architecture enables integration with other existing wireless networks and is widely used in neighboring communities.

Client WMNs: Client meshing architecture provides a communications network based on peer-to-peer over client devices (there is no the role of mesh router). In this case we have a network of mesh nodes which provide routing functionality and configuration as well as end-user applications, so that when a packet is sent from one node to another, the packet will jump from node to node in the mesh of nodes to reach the destination.

Hybrid WMNs: This architecture combines the two previous ones, so that mesh clients are able to access the network through mesh routers as well as through direct connection with other mesh clients. Benefiting from the advantages of the two architectures, Hybrid WMNs can connect to other networks (Internet, Wi-Fi, and sensor networks) and enhance the connectivity and coverage due to the fact that mesh clients can act as mesh routers.

4 Overview of Routing Protocol

4.1 HWMP

Hybrid Wireless Mesh Protocol (HWMP) defined in IEEE 802.11s, is a basic routing protocol for a wireless mesh network. It is based on AODV [20] and tree-based routing. It relies on peer link management protocol by which each mesh point discovers and tracks neighboring nodes. If any of these are connected to a wired backhaul, there is no need for HWMP, which selects paths from those assembled by compiling all mesh point peers into one composite map.

HWMP protocol is hybrid, because it supports two kinds of path selection protocols. Although these protocols are very similar to routing protocols, but bear in mind, that in case of IEEE 802.11s these use MAC addresses for "routing", instead of IP addresses. Therefore, we use the term "path" instead of "route" and thus "path selection" instead of "routing".

HWMP is intended to displace proprietary protocols used by vendors like Meraki for the same purpose, permitting peer participation by open source router firmware.

4.2 OLSR

The OLSR protocol [21] is a pro-active routing protocol, which builds up a route for data transmission by maintaining a routing table inside every node of the network. The routing table is computed upon the knowledge of topology information, which is exchanged by means of Topology Control (TC) packets.

OLSR makes use of `HELLO` messages to find its one hop neighbours and its two hop neighbours through their responses. The sender can then select its Multi Point Relays (MPR) based on the one hop node which offer the best routes to the two hop nodes. By this way, the amount of control traffic can be reduced. Each node has also an MPR selector set which enumerates nodes that have selected it as an MPR node. OLSR uses TC messages along with MPR forwarding to disseminate neighbour information throughout the network. Host Network Address (HNA) messages are used by OLSR to disseminate network route advertisements in the same way TC messages advertise host routes.

5 Simulation Description and Design

5.1 Positioning of mesh routers by WMN-GA system

We use WMN-GA system [22] for node placement problem in WMNs. A bi-objective optimization is used to solve this problem by first maximizing the number

Table 1 Input parameters of WMN-GA system.

Parameters	Values
Number of clients	48
Number of routers	16, 20, 24, 28, 32
Grid width	32 units
Grid height	32 units
Independent runs	10
Number of generations	200
Population size	64
Selection method	Linear Ranking
Crossover rate	80 %
Mutate method	Single
Mutate rate	20 %
Distribution of clients	Normal, Uniform

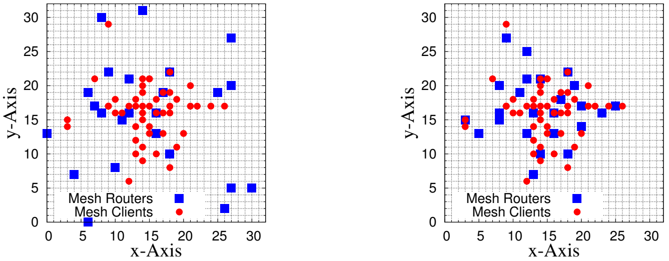
of connected routers in the network and then the client coverage. The input parameters of WMN-GA system are shown in Table 1. In Fig. 1 and Fig. 2, we show the location of mesh routers and clients for first generations and the optimized topologies generated by WMN-GA system for normal and uniform distributions, respectively.

In Fig. 3 and Fig. 4 are shown the simulation results of Size of Giant Component (SGC) vs. number of generations. After few generations, all routers are connected with each other.

Then, we optimize the position of routers in order to cover as many mesh clients as possible. We consider normal and uniform distributions of mesh clients, which are similar with nodes concentrated in event-site environment. The simulation results of SGC and Number of Covered Mesh clients (NCM) are shown in Table. 2.

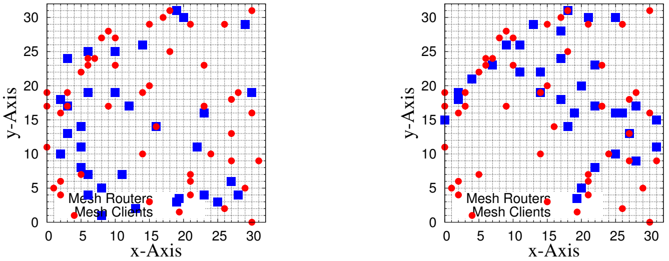
5.2 Simulation Description

We conduct simulations using ns-3 simulator. The simulations in ns-3 are done for number of generations 1 and 200. The area size is considered 640m×640m (or 32 units×32 units) and the number of mesh routers is from 16 to 32. We used HWMP and OLSR routing protocols and sent multiple CBR flows over TCP. The pairs source-destination are the same for all simulation scenarios. Log-distance path loss model and constant speed delay model are used for the simulation and other parameters are shown in Table 3.



(a) Number of generations: 1 (8, 12) (b) Number of generations: 200 (24, 47)

Fig. 1 Location of mesh routers by WMN-GA system, (m, n): m is number of connected mesh routers, n is number of covered mesh clients of normal distribution.



(a) Number of generations: 1 (15, 18) (b) Number of generations: 200 (32, 35)

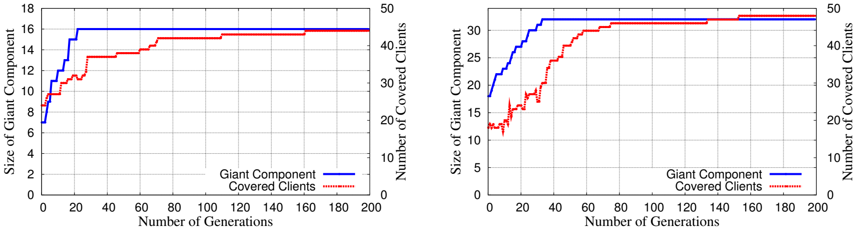
Fig. 2 Location of mesh routers by WMN-GA system, (m, n): m is number of connected mesh routers n is number of covered mesh clients of uniform distribution.

Table 2 Evaluation of WMN-GA system.

Number of mesh routers	Normal Distribution SGM	 NCN	Uniform Distribution SGC	 NCM
16	16	44	16	21
20	20	46	20	22
24	24	47	24	27
28	28	48	28	33
32	32	48	32	35

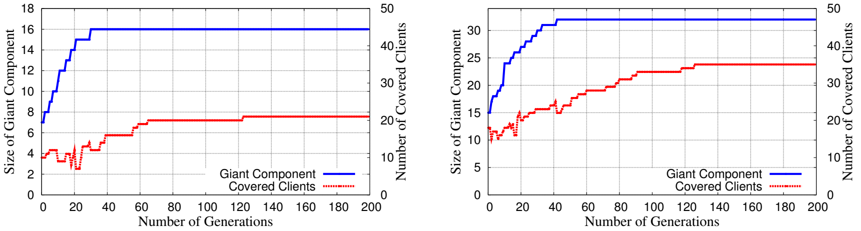
5.3 NS-3

The ns-3 simulator is developed and distributed completely in the C++ programming language, because it better facilitated the inclusion of C-based implementation code [23]. The ns-3 architecture is similar to Linux computers, with internal interface and application interfaces such as network interfaces, device drivers and sockets. The goals of ns-3 are set very high: to create a new network simulator aligned with modern research needs and develop it in an open source community. Users



(a) Number of mesh routers: 16 (b) Number of mesh routers: 32

Fig. 3 SGC and NCM vs. number of generations for Normal Distribution.



(a) Number of mesh routers: 16 (b) Number of mesh routers: 32

Fig. 4 SGC and NCM vs. number of generations for Uniform Distribution.

Table 3 Simulation parameters for ns-3.

Parameters	Values
Area Size	640m×640m
Number of mesh routers	24, 32
Distributions of mesh clients	Normal, Uniform
Number of mesh clients	48
MAC	IEEE 802.11s
Propagation loss model	Log-distance Path Loss Model
Propagation delay model	Constant Speed Model
Routing protocol	HWMP, OLSR
Transport protocol	TCP
Application type	CBR
Packet size	1024 bytes
Number of source nodes	10
Number of destination nodes	1
Transmission energy	17.4 mA
Receiving energy	19.7 mA
Simulation time	60 sec

of ns-3 are free to write their simulation scripts as either *C++ main()* programs or *Python* programs. The ns-3's low-level API is oriented towards the power-user but more accessible "helper" APIs are overlaid on top of the low-level API.

In order to achieve scalability of a very large number of simulated network elements, the ns-3 simulation tools also support distributed simulation. The ns-3 support standardized output formats for trace data, such as the pcap format used by network packet analyzing tools such as tcpdump, and a standardized input format such as importing mobility trace files from ns-2 [24].

The ns-3 simulator is equipped with *Pyviz* visualizer, which has been integrated into mainline ns-3, starting with version 3.10. It can be most useful for debugging purposes, i.e. to figure out if mobility models are what you expect, where packets are being dropped. It is mostly written in Python and it works both with Python and pure C++ simulations. The function of ns-3 visualizer is more powerful than network animator (*nam*) of ns-2 simulator.

The ns-3 simulator has models for all network elements that comprise a computer network. For example, network devices represent the physical device that connects a node to the communication channel. This might be a simple Ethernet network interface card or a more complex wireless IEEE 802.11 device.

The ns-3 is intended as an eventual replacement for popular ns-2 simulator. The ns-3's wifi models a wireless network interface controller based on the IEEE 802.11 standard [25]. The ns-3 provides models for these aspects of 802.11:

1. Basic 802.11 DCF with infrastructure and ad hoc modes.
2. 802.11a, 802.11b, 802.11g and 802.11s physical layers.
3. QoS-based EDCA and queueing extensions of 802.11e.
4. Various propagation loss models including Nakagami, Rayleigh, Friis, LogDistance, FixedRss, and so on.
5. Two propagation delay models, a distance-based and random model.
6. Various rate control algorithms including Aarf, Arf, Cara, Onoe, Rraa, ConstantRate, and Minstrel.

6 Simulation Results

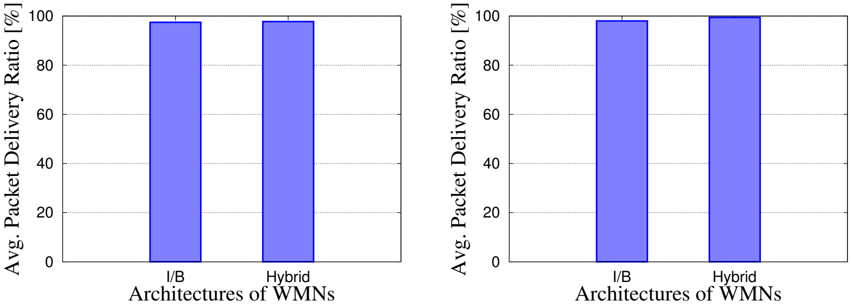
In this section, we present the simulation results. For evaluation, we used the PDR, throughput, delay and energy metrics. We analyze and compare the simulation results considering normal and uniform distributions, HWMP and OLSR routing protocols, TCP, and I/B WMN and Hybrid WMN architectures.

In Fig. 5 and Fig. 6, we show the simulation results for PDR metric. The simulation results show that the PDR for both distributions and architectures is almost the same, but the PDR of HWMP is a little bit better than OLSR.

In Fig. 7 and Fig. 8, we show the simulation results of throughput metric. The throughput is better for normal distribution and I/B WMN architecture in case of HWMP. However, for OLSR and normal distribution, the throughput of Hybrid WMN is a little bit higher than I/B WMN.

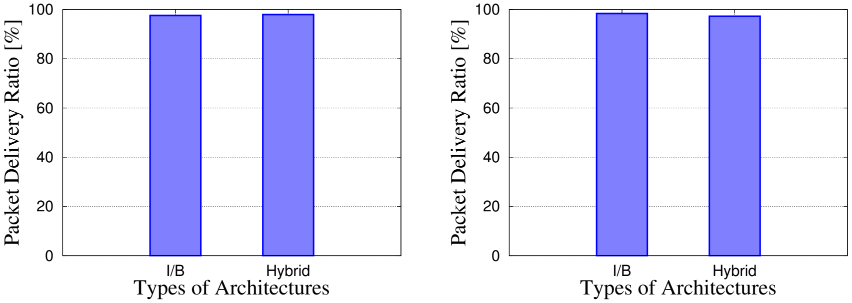
In Fig. 9 and Fig. 10, for both distributions the delay of both architectures is better for HWMP compared with OLSR.

In Fig. 11(a), Fig. 11(b), Fig. 12(a) and Fig. 12(b), we show the remaining energy for both WMN architectures and routing protocols for normal and uniform distributions, respectively. For normal distribution, the energy decreases sharply, because of the high density of nodes, thus the nodes spend more energy. So, the uniform distribution has better performance compared with normal distribution considering the energy parameter.



(a) Normal distribution (b) Uniform distribution

Fig. 5 Results of average PDR for different distributions and architectures considering HWMP.

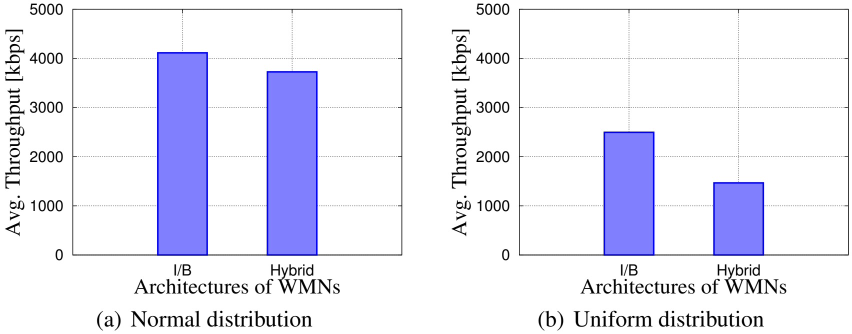


(a) Normal distribution (b) Uniform distribution

Fig. 6 Results of average PDR for different distributions and architectures considering OLSR.

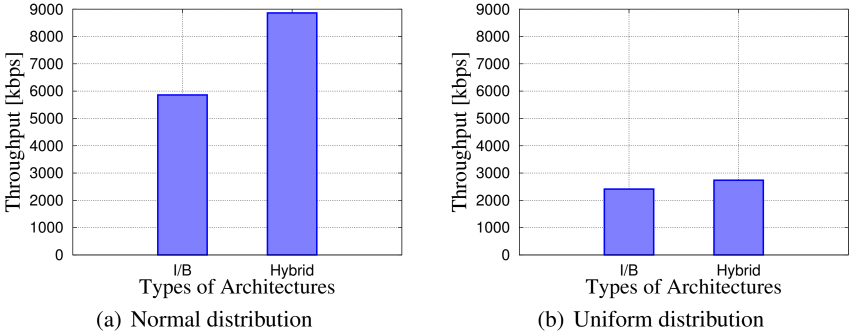
7 Conclusions

In this paper, we evaluated by simulations the performance of WMNs considering PDR, throughput, delay and energy metrics. We considered two architectures of WMNs. The topologies of WMNs are generated using WMN-GA system with area



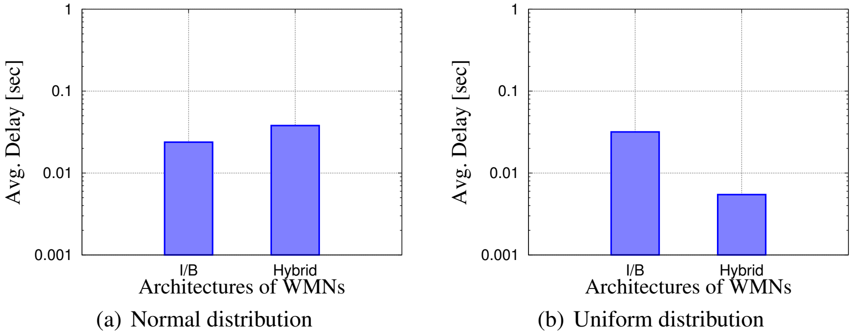
(a) Normal distribution (b) Uniform distribution

Fig. 7 Results of average throughput for different distributions and architectures considering HWMP.



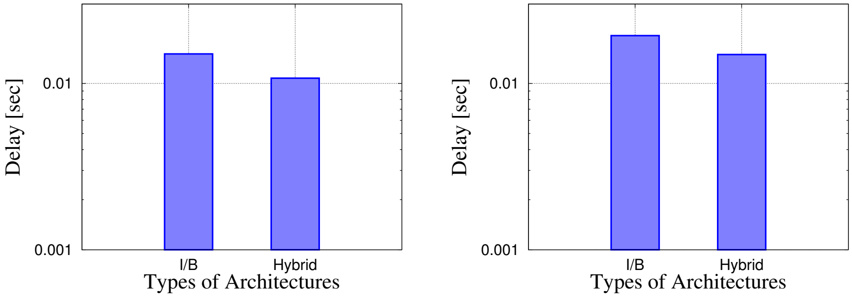
(a) Normal distribution (b) Uniform distribution

Fig. 8 Results of average throughput for different distributions and architectures considering OLSR.



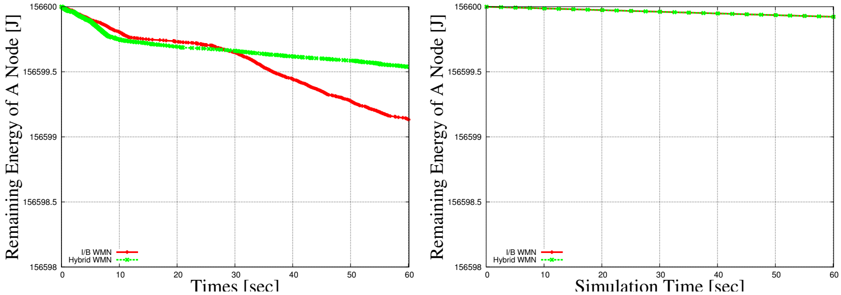
(a) Normal distribution (b) Uniform distribution

Fig. 9 Results of average delay for different distributions and architectures considering HWMP.



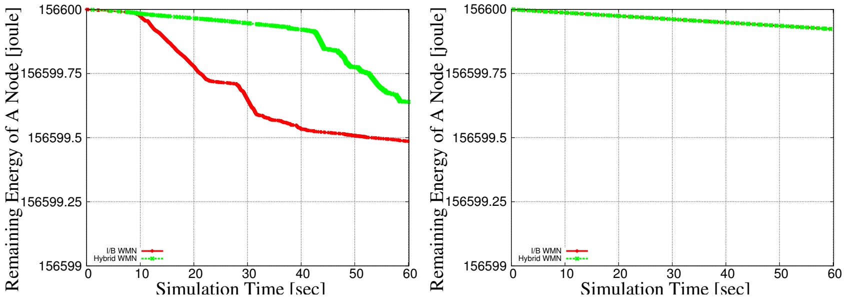
(a) Normal distribution (b) Uniform distribution

Fig. 10 Results of average delay for different distributions and architectures considering OLSR.



(a) Normal distribution (b) Uniform distribution

Fig. 11 Results of remaining energies for different distributions and architectures considering HWMP.



(a) Normal distribution (b) Uniform distribution

Fig. 12 Results of remaining energies for different distributions and architectures considering OLSR.

size 640m×640m. The clients are distributed in the grid using normal and uniform distributions.

We carried out the simulations using ns-3 simulator, and HWMP and OLSR routing protocols. We transmitted multiple CBR flows over TCP. For simulations, we considered log-distance path loss model and constant speed delay model. From simulations, we found the following results.

1. The PDR for both distributions and architectures is almost the same, but the PDR of HWMP is a little bit better than OLSR.
2. The throughput is better for normal distribution and I/B WMN architecture in case of HWMP. However, for OLSR and normal distribution, the throughput of Hybrid WMN is a little bit higher than I/B WMN.
3. For both distributions the delay of both architectures is better for HWMP compared with OLSR.
4. For both WMN architectures and routing protocols for normal and uniform distributions, respectively. For normal distribution, the energy decreases sharply, because of the high density of nodes, thus the nodes spend more energy. So, the uniform distribution has better performance compared with normal distribution considering the energy parameter.

In this work, we consider TCP, two distributions, routing protocols and architectures of WMNs. In the future work, we would like to make extensive simulations for different density of mesh clients and grid sizes.

References

1. I.F. Akyildiz, X. Wang, W. Wang, “Wireless Mesh Networks: A Survey”, In Computer Networks, Vol. 47, No. 4, pp. 445-487, 2005.
2. T. Oda, A. Barolli, F. Xhafa, L. Barolli, M. Ikeda, M. Takizawa, “WMN-GA: A Simulation System for WMNs and Its Evaluation Considering Selection Operators”, Journal of Ambient Intelligence and Humanized Computing (JAIHC), Springer, Vol. 4, No. 3, pp. 323-330, June 2013.
3. M. Ikeda, T. Oda, E. Kulla, M. Hiyama, L. Barolli and M. Younas, “Performance Evaluation of WMN Considering Number of Connections Using NS-3 Simulator”, The Third International Workshop on Methods, Analysis and Protocols for Wireless Communication (MAPWC 2012), pp. 498-502, Victoria, Canada, November 12-14, 2012.
4. E. Nordstrom, “APE - a large scale ad hoc network testbed for reproducible performance tests”, Master’s thesis, Uppsala University, 2002.
5. R. Draves, J. Padhye, and B. Zill, “Comparison of routing metrics for static multi-hop wireless networks,h in SIGCOMM-04, pp. 133-144, 2004.
6. S. N. Muthaiah and C. Rosenberg, “Single Gateway Placement in Wireless Mesh Networks”, In Proc. of 8th International IEEE Symposium on Computer Networks, Turkey, pp. 4754-4759, 2008.
7. M. Tang, “Gateways Placement in Backbone Wireless Mesh Networks”, International Journal of Communications, Network and System Sciences, Vol. 2, No.1, pp. 45-50, 2009.
8. A. Franklin, C. Murthy “Node Placement Algorithm for Deployment of Two-Tier Wireless Mesh Networks”, In: IEEE GLOBECOM-2007, pp. 4823-4827, 2007.

9. T. Vanhatupa, M. Hännikäinen and T.D. Hämäläinen, “Genetic Algorithm to Optimize Node Placement and Configuration for WLAN Planning”, In Proc. of 4th International Symposium on Wireless Communication Systems, pp. 612-616, 2007.
10. A. Lim, B. Rodrigues, F. Wang and Zh. Xua, “$k-$Center Problems with Minimum Coverage”, Theoretical Computer Science, Vol. 332, No. 1-3, pp. 1-17, 2005.
11. J. Wang, B. Xie, K. Cai and D. P. Agrawal, “Efficient Mesh Router Placement in Wireless Mesh Networks”, MASS, Pisa, Italy, pp. 9-11, 2007.
12. X. Yao, “An empirical study of genetic operators in genetic algorithms”, in EUROMICRO 93 Nineteenth EUROMICRO Symposium on Microprocessing and Microprogramming on Open System Design: Hardware, Software and Applications, pp. 707-714, 1993.
13. J. Denzinger and J. Kidney, “Evaluating different genetic operators in the testing for unwanted emergent behavior using evolutionary learning of behavior”, in IEEE/WIC/ACM International Conference on Intelligent Agent Technology, pp. 23-29, 2006.
14. M. Odetayo, gEmpirical study of the interdependencies of genetic algorithm parameters”, in 23rd EUROMICRO Conference, New Frontiers of Information Technology, pp. 639-643, 1997.
15. F. Xhafa, B. Duran, A. Abrahamy, and K. Daha, “Tuning struggle strategy in genetic algorithms for scheduling in computational grids”, Neural Network World, Vol. 18, No. 3, pp. 209-225, 2008.
16. F. Xhafa, L. Barolli, and A. Durresi, “An experimental study on genetic algorithms for resource allocation on grid systems”, Journal of Interconnection Networks, Vol. 8, No. 4, pp. 427-443, 2007.
17. F. Xhafa, C. Sanche, and L. Barolli, “Ad hoc and neighborhood search methods for placement of mesh routers in wireless mesh networks”, in ICDCS Workshops of the IEEE 29th International Conference on Distributed Computing Systems (ICDCS-09), pp. 400-405, 2009.
18. T. Oda, D. Elmazi, A. Barolli, S. Sakamoto, L. Barolli, F. Xhafa, “A Genetic Algorithm Based System for Wireless Mesh Networks: Analysis of System Data Considering Different Routing Protocols and Architectures”, Journal of Soft Computing (SOCO), Springer, Published online: 31 March 2015, DOI: 10.1007/s00500-015-1663-z, pp. 1-14, 2015.
19. F. Xhafa, C. Sanchez, and L. Barolli, “Locals Search Algorithms for Efficient Router Nodes Placement in Wireless Mesh Networks”, in International Conference on Network-Based Information Systems (NBiS), pp. 572-579, 2009.
20. C. Perkins, E. Belding-Royer, and S. Das, “Ad hoc On-Demand Distance Vector(AODV) Routing, RFC3561”, Technical report, Nokia Research Center, University of California, University of Cincinnati, 2003.
21. T. Clausen and P. Jacquet, “Optimized Link State Routing Protocol (OLSR)”, RFC 3626 (Experimental), 2003.
22. T. Oda, A. Barolli, E. Spaho, F. Xhafa, L. Barolli, M. Takizawa, ”Evaluation of WMN-GA for Different Mutation Operators”, International Journal of Space-Based and Situated Computing (IJSSC), Inderscience, Vol. 2. No. 3, pp. 149-157, 2012.
23. “ns-3”, https://www.nsnam.org/.
24. “The Network Simulator-ns-2”, http://www.isi.edu/nsnam/ns/.
25. IEEE 802.11, “Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications”, IEEE Computer Society Std., June 2007. [Online]. Available: http://standards.ieee.org/getieee802/download/802.11-2007.pdf

Comparison Analysis by WMN-GA Simulation System for Different WMN Architectures, Normal and Uniform Distributions, DCF and EDCA Functions

Admir Barolli, Tetsuya Oda, Makoto Ikeda, Leonard Barolli, Fatos Xhafa and Makoto Takizawa

Abstract Wireless Mesh Networks (WMNs) are attracting a lot of attention from wireless network researchers. Node placement problems have been investigated for a long time in the optimization field due to numerous applications in location science. In this paper, we evaluate the performance of two WMN architectures considering throughput, delay, jitter and fairness index metrics. For simulations, we used ns-3 and Optimized Link State Routing (OLSR). We compare the performance of Distributed Coordination Function (DCF) and Enhanced Distributed Channel Access (EDCA) for normal and uniform distributions of mesh clients by sending multiple Constant Bit Rate (CBR) flows in the network. The simulation results show that for normal distribution, the throughput of I/B WMN is higher than Hybrid WMN architecture. For uniform distribution, in case of I/B WMN, the throughput of EDCA is a little bit higher than Hybrid WMN. However, for Hybrid WMN, the throughput of DCF is higher than EDCA. For normal distribution, the delay and jitter of Hybrid WMN is lower compared with I/B WMN. For uniform distribution, the delay and

Admir Barolli
Department of Information Technology, Aleksander Moisiu University of Durres, L.1, Rruga e Currilave, Durres, Albania e-mail: admir.barolli@gmail.com

Fatos Xhafa
Technical University of Catalonia Department of Languages and Informatics Systems C/Jordi Girona 1-3, 08034 Barcelona, Spain e-mail: fatos@lsi.upc.edu

Tetsuya Oda, Makoto Ikeda and Leonard Barolli
Department of Information and Communication Engineering, Fukuoka Institute of Technology (FIT), 3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811–0295, Japan e-mail: oda.tetsuya.fit@gmail.com, m-ikeda@fit.ac.jp, barolli@fit.ac.jp

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_12

jitter of both architectures are almost the same. However, in the case of DCF for 20 flows, the delay and jitter of I/B WMN is a lower compared with Hybrid WMN. In normal distribution case, the fairness index of 10 and 20 flows is higher than 30 flows for both WMN architectures. For I/B architecture the fairness index of DCF is higher than EDCA. However, for Hybrid WMN, the fairness index of EDCA is higher than DCF. For uniform distribution, the fairness index of 10 flows is higher than other flows for both WMN architectures.

1 Introduction

Wireless Mesh Networks (WMNs) [1] are important networking infrastructures. These networks are made up of wireless nodes, organized in a mesh topology, where mesh routers are interconnected by wireless links and provide Internet connectivity to mesh clients.

WMNs distinguish for their low cost nature that makes them attractive for providing wireless Internet connectivity. Moreover, such infrastructure can be used to deploy community networks, metropolitan area networks, municipal and, corporative networks, and to support applications for urban areas, medical, transport and surveillance systems.

The main issue of WMNs is to achieve network connectivity and stability as well as QoS in terms of user coverage. This problem is very closely related to the family of node placement problems in WMNs [3, 4, 2, 5], among them, the mesh router mesh nodes placement. We consider the version of the mesh router nodes placement problem in which we are given a grid area where to deploy a number of mesh router nodes and a number of mesh client nodes of fixed positions (of an arbitrary distribution) in the grid area. The objective is to find a location assignment for the mesh routers to the cells of the grid area that maximizes the network connectivity and client coverage.

As node placement problems are known to be computationally hard to solve for most of the formulations [6], [7], Genetic Algorithms (GAs) has been recently investigated as effective resolution method.

In our previous work [8, 9, 10], we used mesh router nodes placement system that is based on Genetic Algorithms (GAs) to find an optimal location assignment for mesh routers in the grid area in order to maximize the network connectivity and client coverage.

In this work, we use the topology generated by WMN-GA system and evaluate by simulations the performance of uniform distribution of mesh clients considering two architectures and two MAC protocols by sending multiple Constant Bit Rate (CBR) flows in the network. For simulations, we use ns-3 and Optimized Link State Routing (OLSR). As evaluation metrics we considered throughput, one-way delay, jitter and fairness.

The rest of the paper is organized as follows. Architectures of WMNs are presented in Section 2. In Section 3, we show the description and design of the simulation system. In Section 4, we discuss the simulation results. Finally, conclusions and future work are given in Section 5.

2 Architectures of WMNs

In this section, we describe the architectures of WMN. The architecture of the nodes in WMNs [11, 12, 13, 14] can be classified according to the functionalities they offer as follows:

Infrastructure/Backbone WMNs: This type of architecture (also known as infrastructure meshing) is the most used and consists of a grid of mesh routers which are connected to different clients. Moreover, routers have gateway functionality thus allowing Internet access for clients. This architecture enables integration with other existing wireless networks and is widely used in neighboring communities.

Client WMNs: Client meshing architecture provides a communications network based on peer-to-peer over client devices (there is no the role of mesh router). In this case we have a network of mesh nodes which provide routing functionality and configuration as well as end-user applications, so that when a packet is sent from one node to another, the packet will jump from node to node in the mesh of nodes to reach the destination.

Hybrid WMNs: This architecture combines the two previous ones, so that mesh clients are able to access the network through mesh routers as well as through direct connection with other mesh clients. Benefiting from the advantages of the two architectures, Hybrid WMNs can connect to other networks (Internet, Wi-Fi, and sensor networks) and enhance the connectivity and coverage due to the fact that mesh clients can act as mesh routers.

3 Simulation Description and Design

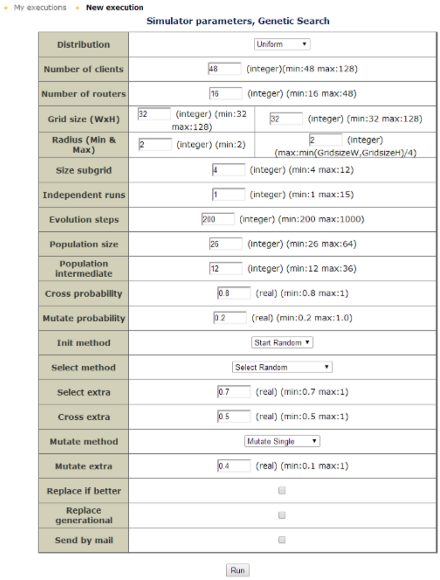
3.1 GUI of WMN-GA System

The WMN-GA system can generate instances of the problem using different distributions of client and mesh routers.

The GUI interface of WMN-GA is shown in Fig. 1. The left site of the interface shows the GA parameters configuration and on the right side are shown the network configuration parameters.

For the network configuration, we use: distribution, number of clients, number of mesh routers, grid size, radius of transmission distance and the size of subgrid.

For the GA parameter configuration, we use: number of independent runs, GA evolution steps, population size, population intermediate size, crossover probability, mutation probability, initial methods, select method.



Simulator parameters, Genetic Search

Distribution	Uniform ▾	
Number of clients	48 (integer)(min:48 max:128)	
Number of routers	16 (integer) (min:16 max:48)	
Grid size (WxH)	32 (integer) (min:32 max:128)	32 (integer) (min:32 max:128)
Radius (Min & Max)	2 (integer) (min:2)	2 (integer) (max:min(GridsizeW,GridsizeH)/4)
Size subgrid	4 (integer) (min:4 max:12)	
Independent runs	1 (integer) (min:1 max:15)	
Evolution steps	200 (integer) (min:200 max:1000)	
Population size	26 (integer) (min:26 max:64)	
Population intermediate	12 (integer) (min:12 max:36)	
Cross probability	0.8 (real) (min:0.8 max:1)	
Mutate probability	0.2 (real) (min:0.2 max:1.0)	
Init method	Start Random ▾	
Select method	Select Random ▾	
Select extra	0.7 (real) (min:0.7 max:1)	
Cross extra	0.5 (real) (min:0.5 max:1)	
Mutate method	Mutate Single ▾	
Mutate extra	0.4 (real) (min:0.1 max:1)	
Replace if better	□	
Replace generational	□	
Send by mail	□	

Fig. 1 GUI tool for WMN-GA system.

3.2 Positioning of mesh routers by WMN-GA system

We use WMN-GA system for node placement problem in WMNs. A bi-objective optimization is used to solve this problem by first maximizing the number of connected routers in the network and then the client coverage. The input parameters of WMN-GA system are shown in Table 1. In Fig. 2, we show the location of mesh routers and clients for first generations and the optimized topologies generated by WMN-GA system for normal and uniform distribution.

In Fig. 4 are shown the simulation results of Size of Giant Component (SGC) and Number of Covered Mesh Clients (NCMC) vs. number of generations. After few generations, all routers are connected with each other.

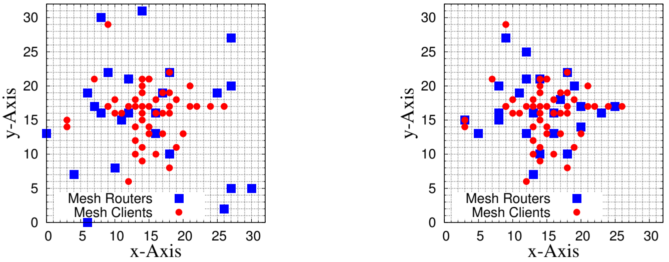
Then, we optimize the position of routers in order to cover as many mesh clients as possible. The simulation results of SGC and NCMC are shown in Table 2.

3.3 Simulation Description

We conduct simulations using ns-3 simulator. The simulations in ns-3 are done for number of generations 1 and 200. The area size is considered 640m×640m (or 32 units×32 units) and the number of mesh routers is from 16 to 32. We used DCF, EDCA and OLSR routing protocol and sent multiple CBR flows over UDP. The pairs source-destination are the same for all simulation scenarios. Log-distance path

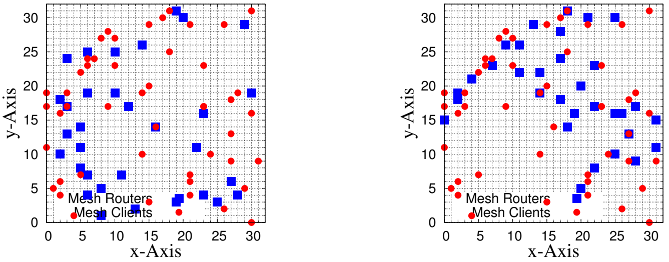
Table 1 Input parameters of WMN-GA system.

Parameters	Values
Number of clients	48
Number of routers	16, 24, 32
Grid width	32 [units]
Grid height	32 [units]
Independent runs	10
Number of generations (NG)	200
Population size	64
Selection method	Linear Ranking
Crossover rate	80 [%]
Mutate method	Single
Mutate rate	20 [%]
Distribution of clients	Normal, Uniform



(a) Number of generations: 1 (8, 12) (b) Number of generations: 200 (32, 35)

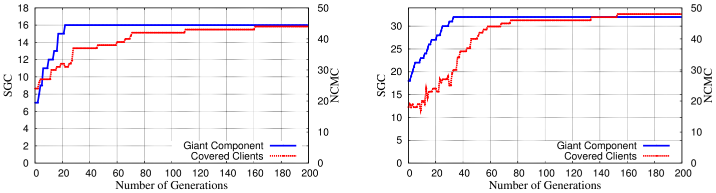
Fig. 2 Location of mesh routers by WMN-GA system for normal distribution; (m, n): m is number of connected mesh routers, n is number of covered mesh clients.



(a) Number of generations: 1 (8, 12) (b) Number of generations: 200 (32, 35)

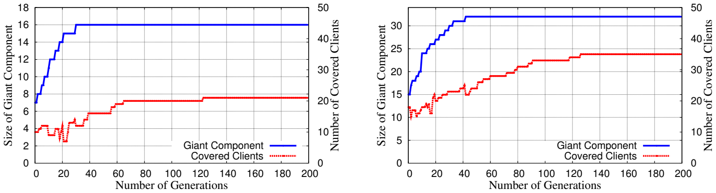
Fig. 3 Location of mesh routers by WMN-GA system for uniform distribution; (m, n): m is number of connected mesh routers, n is number of covered mesh clients.

loss model and constant speed delay model are used for the simulation and other parameters are shown in Table 3.



(a) Number of mesh routers: 16 (b) Number of mesh routers: 32

Fig. 4 SGC and NCMC vs. number of generations for normal distribution.



(a) Number of mesh routers: 16 (b) Number of mesh routers: 32

Fig. 5 SGC and NCMC vs. number of generations for uniform distribution.

Table 2 Evaluation of WMN-GA system.

Number of	Normal Distribution		Uniform Distribution	
mesh routers	SGC	NCMC	SGC	NCMC
16	16	44	16	21
20	20	46	20	22
24	24	47	24	27
28	28	48	28	33
32	32	48	32	35

3.4 NS-3

The ns-3 simulator [15] is developed and distributed completely in the C++ programming language, because it better facilitated the inclusion of C-based implementation code. The ns-3 architecture is similar to Linux computers, with internal interface and application interfaces such as network interfaces, device drivers and sockets. The goals of ns-3 are set very high: to create a new network simulator aligned with modern research needs and develop it in an open source community. Users of ns-3 are free to write their simulation scripts as either *C++ main()* programs or *Python* programs. The ns-3's low-level API is oriented towards the power-user but more accessible "helper" APIs are overlaid on top of the low-level API.

In order to achieve scalability of a very large number of simulated network elements, the ns-3 simulation tools also support distributed simulation. The ns-3 support standardized output formats for trace data, such as the pcap format used by

Table 3 Simulation parameters for ns-3.

Parameters	Values
Area Size	640[m]×640[m]
Distributions of mesh clients	Normal, Uniform
Number of mesh routers	16
Number of mesh clients	48
PHY protocol	IEEE 802.11b
Propagation loss model	Log-distance Path Loss Model
Propagation delay model	Constant Speed Model
MAC protocols	DCF, EDCA
Routing protocol	OLSR
Transport protocol	UDP
Application type	CBR
Packet size	1024 [Bytes]
Number of source nodes	10, 20, 30
Number of destination node	1
Transmission current	17.4 [mA]
Receiving current	19.7 [mA]
Simulation time	600 [sec]

network packet analyzing tools such as tcpdump, and a standardized input format such as importing mobility trace files from ns-2 [16].

The ns-3 simulator is equipped with *Pyviz* visualizer, which has been integrated into mainline ns-3, starting with version 3.10. It can be most useful for debugging purposes, i.e. to figure out if mobility models are what you expect, where packets are being dropped. It is mostly written in Python and it works both with Python and pure C++ simulations. The function of ns-3 visualizer is more powerful than network animator (*nam*) of ns-2 simulator.

The ns-3 simulator has models for all network elements that comprise a computer network. For example, network devices represent the physical device that connects a node to the communication channel. This might be a simple Ethernet network interface card or a more complex wireless IEEE 802.11 device.

The ns-3 is intended as an eventual replacement for popular ns-2 simulator. The ns-3's wifi models a wireless network interface controller based on the IEEE 802.11 standard [17]. The ns-3 provides models for these aspects of 802.11:

1. Basic 802.11 DCF with infrastructure and ad hoc modes.
2. 802.11a, 802.11b, 802.11g and 802.11s physical layers.
3. QoS-based EDCA and queueing extensions of 802.11e.
4. Various propagation loss models including Nakagami, Rayleigh, Friis, LogDistance, FixedRss, and so on.
5. Two propagation delay models, a distance-based and random model.
6. Various rate control algorithms including Aarf, Arf, Cara, Onoe, Rraa, ConstantRate, and Minstrel.

3.5 Overview of DCF and EDCA Protocols

In our study we concentrate on two distributed access methods: DCF from legacy 802.11 [18] and EDCA from 802.11e [19]. The centralised access methods, Point Coordination Function (PCF) [18] and Hybrid Controlled Channel Access (HCCA) [19] are not considered as they are rarely implemented in hardware devices [20].

3.5.1 DCF

DCF is a random access scheme based on the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) scheme. A legacy DCF station with a packet to send will first sense the medium for activity. If the channel is idle for a Distributed Inter-Frame Space (DIFS), the station will attempt to transmit after a random back-off period. This period is referred as the Contention Window (CW). The value for the CW is chosen randomly from a range $[0, 2^n - 1]$, i.e.

$$CW_{min} \leq CW \leq CW_{max} \tag{1}$$

where n is PHY dependent. Initially, CW is set to the minimum number of slot times CW_{min}, which is defined per PHY in microseconds [18]. The randomly chosen CW value, referred as the back-off counter, is decreased each slot time if the medium remains idle. If during any period the medium becomes busy, the back-off counter is paused and resumed only when the medium becomes idle. On reaching zero, the station transmits the packet in the physical channel and awaits an acknowledgment (ACK). The transmitting station then performs a post back-off, where the back-off procedure is repeated once more. This is to allow other stations to gain access to the medium during heavy contention.

If the ACK is not received within a Short Inter-Frame Space (SIFS), it assumes that the frame was lost due to collision or being damaged. The CW value is then increased exponentially and the back-off begins once again for retransmission. This is referred as the Automatic Repeat Request (ARQ) process. If the following retransmission attempt fails, the CW is again increased exponentially, up until the limit CW_{max}. The retransmission process will repeat for up to 4 or 7 times, depending on whether the short retry limit or long retry limit is used. Upon reaching the retry limit the packet is considered lost and discarded. The retry limit is manufacturer dependent and can vary considerably.

3.5.2 Enhanced Distributed Channel Access (EDCA)

The enhanced access method EDCA builds on the legacy DCF process and introduces four different Access Categories (ACs) or traffic classes for service differentiation at the MAC layer. This is achieved by varying the size of CW in the backoff mechanism on a per category basis. Service differentiation is provided by the following methods:

Arbitration Inter-Frame Space (AIFS)

This is similar to the DIFS used in DCF, except the AIFS can vary according the access category;

Variable Contention Window

By giving higher priority traffic smaller contention windows, less time is spent in the back-off state, resulting in more frequent access to the medium.

Transmission Opportunity (TxOP)

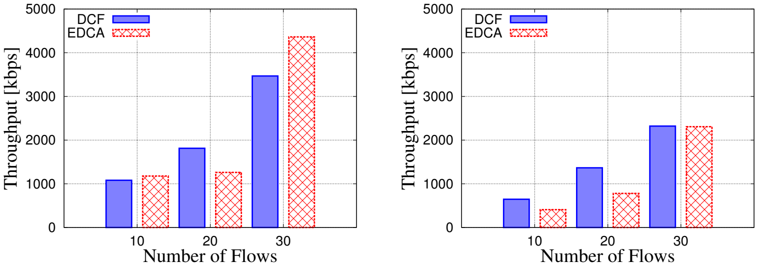
This allows a station that has access to the medium to transmit a number of data units without having to contend for access to the medium. In fact this is a form of frame bursting. The TxOP limit is defined per traffic class.

Multiple AC queues can exist on a single station, contending with each other for the physical medium. This is regarded as virtual contention.

3.6 Overview of OLSR Routing Protocol

The OLSR protocol [21] is a pro-active routing protocol, which builds up a route for data transmission by maintaining a routing table inside every node of the network. The routing table is computed upon the knowledge of topology information, which is exchanged by means of Topology Control (TC) packets.

OLSR makes use of HELLO messages to find its one hop neighbours and its two hop neighbours through their responses. The sender can then select its Multi Point Relays (MPR) based on the one hop node which offer the best routes to the two hop nodes. By this way, the amount of control traffic can be reduced. Each node has also an MPR selector set which enumerates nodes that have selected it as an MPR node. OLSR uses TC messages along with MPR forwarding to disseminate neighbour information throughout the network. Host Network Address (HNA) messages are used by OLSR to disseminate network route advertisements in the same way TC messages advertise host routes.



(a) I/B WMN (b) Hybrid WMN

Fig. 6 Results of average throughput considering normal distribution.

4 Simulation Results

We used the throughput, delay, jitter and fairness index metrics to evaluate the performance of WMNs for two architectures considering DCF and EDCA functions, and normal and uniform distributions.

In Fig. 6 and Fig. 7, we show the simulation results of throughput. For normal distribution, the throughput of I/B WMN is higher than Hybrid WMN architecture. For uniform distribution, in case of I/B WMN, the throughput of EDCA is a little bit higher than Hybrid WMN. However, for Hybrid WMN, the throughput of DCF is higher than EDCA.

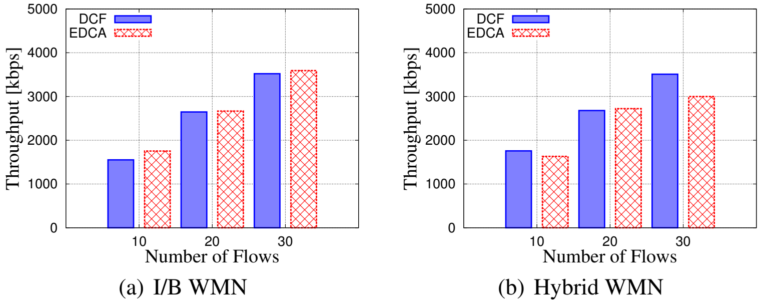
In Fig. 8, Fig. 9, Fig. 10 and Fig. 11, for normal distribution, the delay and jitter of Hybrid WMN is lower compared with I/B WMN. In uniform distribution case, the delay and jitter of both architectures are almost the same. However, in the case of DCF for 20 flows, the delay and jitter of I/B WMN is lower compared with Hybrid WMN.

In Fig. 12 and Fig. 13, we show the fairness index. For normal distribution, the fairness index of 10 and 20 flows is higher than 30 flows for both WMN architectures. For I/B architecture the fairness index of DCF is higher than EDCA. However, for Hybrid WMN, the fairness index of EDCA is higher than DCF. In uniform distribution case, the fairness index of 10 flows is higher than other flows for both WMN architectures.

5 Conclusions

In this work, we presented WMN-GA system and applied it for node placement problem in WMNs. We evaluated the performance of WMN-GA system for normal and uniform distributions of mesh clients considering DCF, EDCA and OLSR protocols.

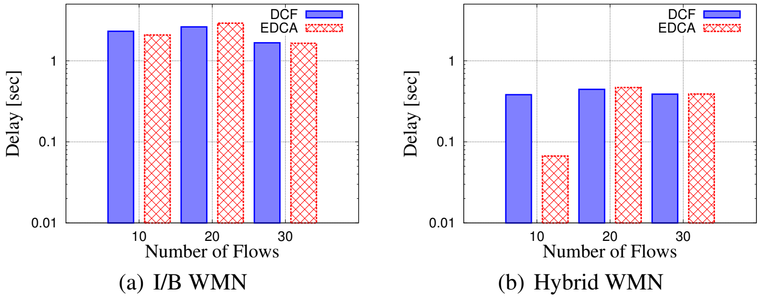
From the simulations we conclude as follows.



(a) I/B WMN

(b) Hybrid WMN

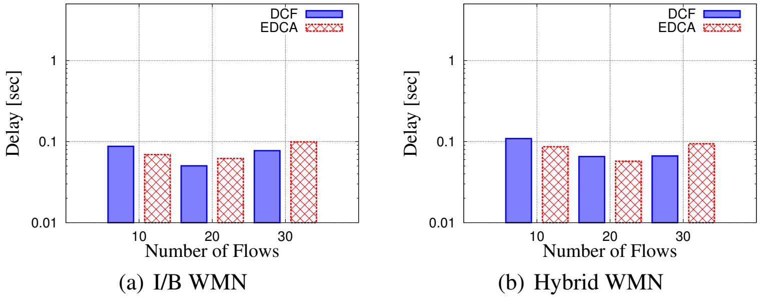
Fig. 7 Results of average throughput considering uniform distribution.



(a) I/B WMN

(b) Hybrid WMN

Fig. 8 Results of average delay considering normal distribution.

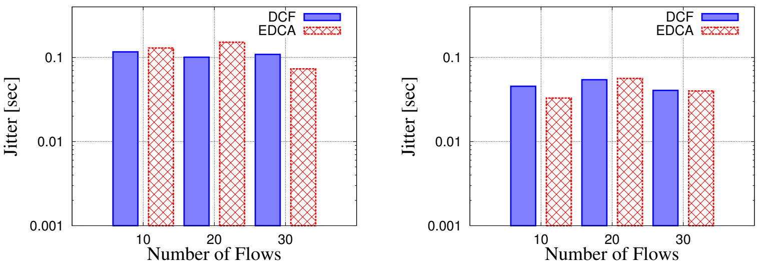


(a) I/B WMN

(b) Hybrid WMN

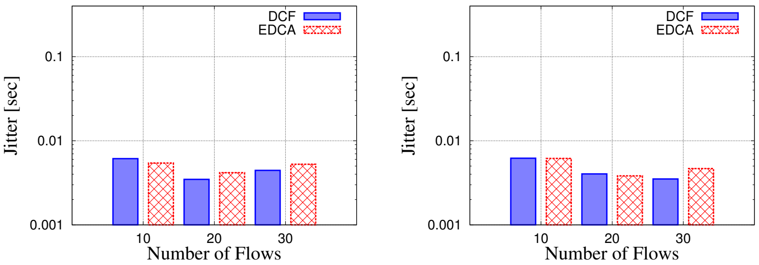
Fig. 9 Results of average delay considering uniform distribution.

- For normal distribution, the throughput of I/B WMN is higher than Hybrid WMN architecture. For uniform distribution, in case of I/B WMN, the throughput of EDCA is a little bit higher than Hybrid WMN. However, for Hybrid WMN, the throughput of DCF is higher than EDCA.
- For normal distribution, the delay and jitter of Hybrid WMN is lower compared with I/B WMN. For uniform distribution, the delay and jitter of both architectures



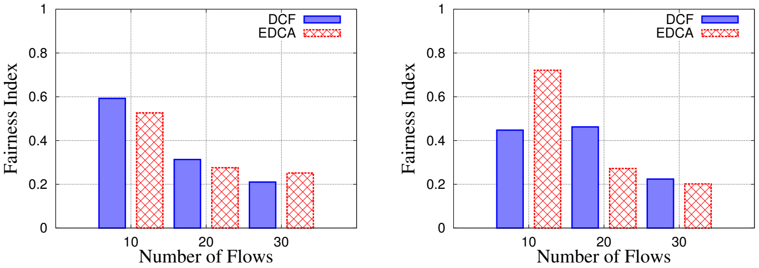
(a) I/B WMN (b) Hybrid WMN

Fig. 10 Results of average jitter considering normal distribution.



(a) I/B WMN (b) Hybrid WMN

Fig. 11 Results of average jitter considering uniform distribution.



(a) I/B WMN (b) Hybrid WMN

Fig. 12 Results of fairness index considering normal distribution.

are almost the same. However, in the case of DCF for 20 flows, the delay and jitter of I/B WMN is a lower compared with Hybrid WMN.

- In normal distribution case, the fairness index of 10 and 20 flows is higher than 30 flows for both WMN architectures. For I/B architecture the fairness index of DCF is higher than EDCA. However, for Hybrid WMN, the fairness index of EDCA is higher than DCF. For uniform distribution, the fairness index of 10 flows is higher than other flows for both WMN architectures.

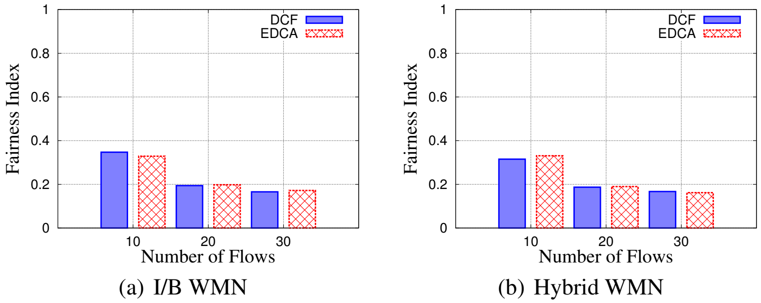


Fig. 13 Results of fairness index considering uniform distribution.

References

1. I.F. Akyildiz, X. Wang, W. Wang, "Wireless Mesh Networks: A Survey", In Computer Networks, Vol. 47, No. 4, pp. 445-487, 2005.
2. A. Franklin, C. Murthy "Node Placement Algorithm for Deployment of Two-Tier Wireless Mesh Networks", In: IEEE GLOBECOM-2007, pp. 4823-4827, 2007.
3. S. N. Muthaiah and C. Rosenberg, "Single Gateway Placement in Wireless Mesh Networks", In Proc. of 8th International IEEE Symposium on Computer Networks, Turkey, pp. 4754-4759, 2008.
4. M. Tang, "Gateways Placement in Backbone Wireless Mesh Networks", International Journal of Communications, Network and System Sciences, Vol. 2, No.1, pp. 45-50, 2009.
5. T. Vanhatupa, M. Hännikäinen and T.D. Hämäläinen, "Genetic Algorithm to Optimize Node Placement and Configuration for WLAN Planning", In Proc. of 4th International Symposium on Wireless Communication Systems, pp. 612-616, 2007.
6. A. Lim, B. Rodrigues, F. Wang and Zh. Xua, "$k-$Center Problems with Minimum Coverage", Theoretical Computer Science, Vol. 332, No. 1-3, pp. 1-17, 2005.
7. J. Wang, B. Xie, K. Cai and D. P. Agrawal, "Efficient Mesh Router Placement in Wireless Mesh Networks", MASS, Pisa, Italy, pp. 9-11, 2007.
8. T. Oda, A. Barolli, F. Xhafa, L. Barolli, M. Ikeda, M. Takizawa, "WMN-GA: A Simulation System for WMNs and Its Evaluation Considering Selection Operators", Journal of Ambient Intelligence and Humanized Computing (JAIHC), Springer, Vol. 4, No. 3, pp. 323-330, June 2013
9. M. Ikeda, T. Oda, E. Kulla, M. Hiyama, L. Barolli and M. Younas, "Performance Evaluation of WMN Considering Number of Connections Using NS-3 Simulator", The Third International Workshop on Methods, Analysis and Protocols for Wireless Communication (MAPWC 2012), pp. 498-502, Victoria, Canada, November 12-14, 2012.
10. T. Oda, D. Elmazi, A. Barolli, S. Sakamoto, L. Barolli, F. Xhafa, "A Genetic Algorithm Based System for Wireless Mesh Networks: Analysis of System Data Considering Different Routing Protocols and Architectures", Journal of Soft Computing (SOCO), Springer, Published online: 31 March 2015, DOI: 10.1007/s00500-015-1663-z, pp. 1-14, 2015.
11. F. Xhafa, C. Sanchez, and L. Barolli, "Locals Search Algorithms for Efficient Router Nodes Placement in Wireless Mesh Networks", in International Conference on Network-Based Information Systems (NBiS), pp. 572-579, 2009.
12. T. Oda, A. Barolli, E. Spaho, L. Barolli, F. Xhafa, "Analysis of Mesh Router Placement in Wireless Mesh Networks Using Friedman Test", Proc. of The 28th IEEE International Conference on Advanced Information Networking and Applications (IEEE AINA), pp. 289-296, Victoria, Canada, May 2014,

13. T. Oda, S. Sakamoto, A. Barolli, M. Ikeda, L. Barolli, F. Xhafa, “A GA-Based Simulation System for WMNs: Performance Analysis for Different WMN Architectures Considering TCP”, 2014 Eighth International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA), pp. 120-126, Guangzhou, China, November 2014.
14. T. Oda, A. Barolli, E. Spaho, F. Xhafa, L. Barolli, M. Takizawa, ”Evaluation of WMN-GA for Different Mutation Operators”, International Journal of Space-Based and Situated Computing (IJSSC), Inderscience, Vol. 2. No. 3, pp. 149-157, 2012.
15. “ns-3”, https://www.nsnam.org/.
16. “The Network Simulator-ns-2”, http://www.isi.edu/nsnam/ns/.
17. IEEE 802.11, “Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications”, IEEE Computer Society Std., June 2007. [Online]. Available: http://standards.ieee.org/getieee802/download/802.11-2007.pdf
18. IEEE-SA, “IEEE 802.11 Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications”, 1999.
19. IEEE-SA, “IEEE 802.11e Amendment: Medium Access Control (MAC) Quality of Service (QoS) Enhancements”, 2005.
20. S. Mukherjee, P. Xiao-Hong, Q. Gao, “QoS Performances of IEEE 802.11 EDCA and DCF: A Testbed Approach”, 5th International Conference Wireless Communications, Networking and Mobile Computing (WiCom ’09), pp. 1-5, 2009.
21. T. Clausen and P. Jacquet, “Optimized Link State Routing Protocol (OLSR)”, RFC 3626 (Experimental), 2003.

A GA-Based Simulation System for WMNs: Performance Analysis for Different WMN Architectures Considering Uniform Distribution, Transmission Rate and OLSR Protocol

Keita Matsuo, Tetsuya Oda, Admir Barolli, Makoto Ikeda, Leonard Barolli and Fatos Xhafa

Abstract In this paper, we evaluate the performance of two WMN architectures considering throughput, delay, jitter and fairness index metrics. For simulations, we used ns-3. We compare the performance for two architectures considering transmission rate, uniform distribution and OLSR protocol. The simulation results show that for transmission rate 600 and 1200 [kbps], the throughput of Hybrid WMN is higher than transmission rate 100 [kbps]. For transmission rate 100 [kbps], the delay and jitter of Hybrid WMN is lower than other transmission rates. For transmission rate 100 [kbps], the fairness index of I/B WMN is higher than other transmission rates.

1 Introduction

Wireless Mesh Networks (WMNs) [1] are important networking infrastructures. These networks are made up of wireless nodes, organized in a mesh topology, where mesh routers are interconnected by wireless links and provide Internet connectivity to mesh clients.

The main issue of WMNs is to achieve network connectivity and stability as well as QoS in terms of user coverage. This problem is very closely related to the family of node placement problems in WMNs [3, 4, 2, 5], among them, the mesh router

Keita Matsuo, Tetsuya Oda, Makoto Ikeda and Leonard Barolli
Department of Information and Communication Engineering, Fukuoka Institute of Technology (FIT), 3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811–0295, Japan, e-mail: kt-matsuo@fit.ac.jp, oda.tetsuya.fit@gmail.com, m-ikeda@fit.ac.jp, barolli@fit.ac.jp

Admir Barolli
Department of Information Technology, Aleksander Moisiu University of Durres, L.1, Rruga e Currilave, Durres, Albania, e-mail: admir.barolli@gmail.com

Fatos Xhafa
Technical University of Catalonia Department of Languages and Informatics Systems C/Jordi Girona 1-3, 08034 Barcelona, Spain, e-mail: fatos@lsi.upc.edu

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_13

mesh nodes placement. We consider the version of the mesh router nodes placement problem in which we are given a grid area where to deploy a number of mesh router nodes and a number of mesh client nodes of fixed positions (of an arbitrary distribution) in the grid area. The objective is to find a location assignment for the mesh routers to the cells of the grid area that maximizes the network connectivity and client coverage.

As node placement problems are known to be computationally hard to solve for most of the formulations [6], [7], Genetic Algorithms (GAs) has been recently investigated as effective resolution method.

In our previous work [8, 9, 10, 11], we used mesh router nodes placement system that is based on Genetic Algorithms (GAs) to find an optimal location assignment for mesh routers in the grid area in order to maximize the network connectivity and client coverage.

In this work, we use the topology generated by WMN-GA system and evaluate by simulations the performance of uniform distribution of mesh clients considering two architectures and transmission rate by sending multiple Constant Bit Rate (CBR) flows in the network. For simulations, we use ns-3 and Optimized Link State Routing (OLSR). As evaluation metrics we considered throughput, delay, jitter and fairness.

The rest of the paper is organized as follows. Architectures of WMNs are presented in Section 2. In Section 3, we show the description and design of the simulation system. In Section 4, we discuss the simulation results. Finally, conclusions and future work are given in Section 5.

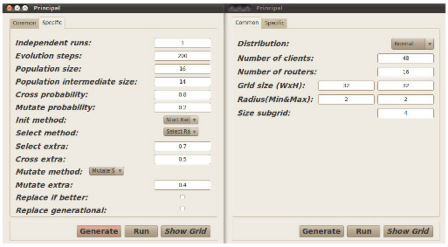
2 Architectures of WMNs

In this section, we describe the architectures of WMN. The architecture of the nodes in WMNs [12, 13, 14, 15] can be classified according to the functionalities they offer as follows:

Infrastructure/Backbone WMNs: This type of architecture (also known as infrastructure meshing) is the most used and consists of a grid of mesh routers which are connected to different clients. Moreover, routers have gateway functionality thus allowing Internet access for clients. This architecture enables integration with other existing wireless networks and is widely used in neighboring communities.

Client WMNs: Client meshing architecture provides a communications network based on peer-to-peer over client devices (there is no the role of mesh router). In this case we have a network of mesh nodes which provide routing functionality and configuration as well as end-user applications, so that when a packet is sent from one node to another, the packet will jump from node to node in the mesh of nodes to reach the destination.

Hybrid WMNs: This architecture combines the two previous ones, so that mesh clients are able to access the network through mesh routers as well as through direct connection with other mesh clients. Benefiting from the advantages of the two



Independent runs: 1
Evolution steps: 200
Population size: 16
Population intermediate size: 14
Cross probability: 0.8
Mutate probability: 0.2
Init method: Start Ran
Select method: Select Ra
Select extra: 0.7
Cross extra: 0.5
Mutate method: Mutate S
Mutate extra: 0.4
Replace if better: □
Replace generational: □

Distribution: Normal
Number of clients: 48
Number of routers: 16
Grid size (WxH): 32 32
Radius(Min&Max): 2 2
Size subgrid: 4

Fig. 1 GUI tool for WMN-GA system.

architectures, Hybrid WMNs can connect to other networks (Internet, Wi-Fi, and sensor networks) and enhance the connectivity and coverage due to the fact that mesh clients can act as mesh routers.

3 Simulation Description and Design

3.1 GUI of WMN-GA System

The WMN-GA system can generate instances of the problem using different distributions of client and mesh routers.

The GUI interface of WMN-GA is shown in Fig. 1. The left site of the interface shows the GA parameters configuration and on the right side are shown the network configuration parameters.

For the network configuration, we use: distribution, number of clients, number of mesh routers, grid size, radius of transmission distance and the size of subgrid.

For the GA parameter configuration, we use: number of independent runs, GA evolution steps, population size, population intermediate size, crossover probability, mutation probability, initial methods, select method.

3.2 Positioning of mesh routers by WMN-GA system

We use WMN-GA system for node placement problem in WMNs. A bi-objective optimization is used to solve this problem by first maximizing the number of connected routers in the network and then the client coverage. The input parameters of WMN-GA system are shown in Table 1. In Fig. 2, we show the location of mesh routers and clients for first generations and the optimized topologies generated by WMN-GA system for normal distribution.

In Fig. 3 are shown the simulation results of Size of Giant Component (SGC) and Number of Covered Mesh Clients (NCMC) vs. number of generations. After few generations, all routers are connected with each other.

Table 1 Input parameters of WMN-GA system.

Parameters	Values
Number of clients	48
Number of routers	16, 20, 24, 28, 32
Grid width	32 [units]
Grid height	32 [units]
Independent runs	10
Number of generations (NG)	200
Population size	64
Selection method	Linear Ranking
Crossover rate	80 [%]
Mutate method	Single
Mutate rate	20 [%]
Distribution of clients	Uniform

Table 2 Evaluation of WMN-GA system.

Number of	Uniform Distribution	
mesh routers	SGC	NCMC
16	16	21
20	20	22
24	24	27
28	28	33
32	32	35

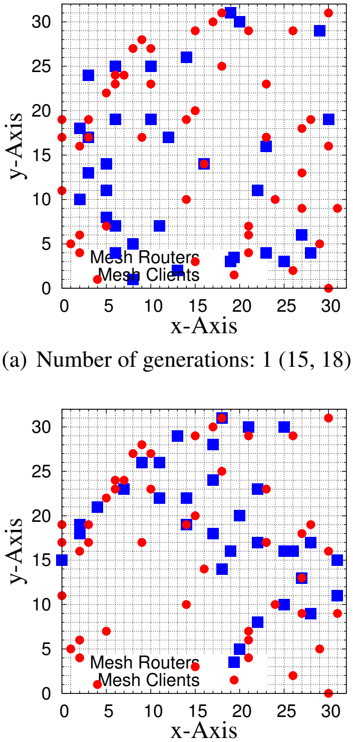
Then, we optimize the position of routers in order to cover as many mesh clients as possible. We consider normal distribution of mesh clients. The simulation results of SGC and NCMC are shown in Table 2.

3.3 Simulation Description

We conduct simulations using ns-3 simulator. The simulations in ns-3 are done for number of generations 1 and 200. The area size is considered 640 [m]×640 [m] (or 32 [units]×32 [units]) and the number of mesh routers is from 16 to 32. We used transmission rate, DCF and OLSR routing protocol and sent multiple CBR flows over UDP. The pairs source-destination are the same for all simulation scenarios. Log-distance path loss model and constant speed delay model are used for the simulation and other parameters are shown in Table 3.

3.4 NS-3

The ns-3 simulator [16] is developed and distributed completely in the C++ programming language, because it better facilitated the inclusion of C-based imple-



(a) Number of generations: 1 (15, 18)

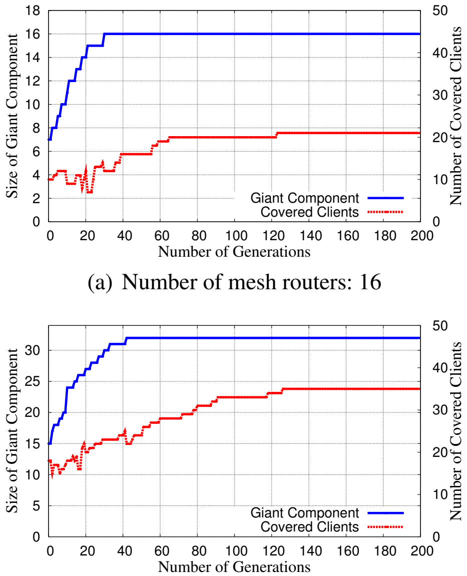
(b) Number of generations: 200 (32, 35)

Fig. 2 Node placement for uniform distribution, (m, n): m is SGC, n is NCMC.

mentation code. The ns-3 architecture is similar to Linux computers, with internal interface and application interfaces such as network interfaces, device drivers and sockets. The goals of ns-3 are set very high: to create a new network simulator aligned with modern research needs and develop it in an open source community. Users of ns-3 are free to write their simulation scripts as either *C++ main()* programs or *Python* programs. The ns-3's low-level API is oriented towards the power-user but more accessible "helper" APIs are overlaid on top of the low-level API.

In order to achieve scalability of a very large number of simulated network elements, the ns-3 simulation tools also support distributed simulation. The ns-3 support standardized output formats for trace data, such as the pcap format used by network packet analyzing tools such as tcpdump, and a standardized input format such as importing mobility trace files from ns-2 [17].

The ns-3 simulator is equipped with *Pyviz* visualizer, which has been integrated into mainline ns-3, starting with version 3.10. It can be most useful for debugging purposes, i.e. to figure out if mobility models are what you expect, where packets are being dropped. It is mostly written in Python and it works both with Python



(a) Number of mesh routers: 16

(b) Number of mesh routers: 32

Fig. 3 SGC and NCMC vs. number of generations for uniform distribution.

Table 3 Simulation parameters for ns-3.

Parameters	Values
Area Size	640[m]×640[m]
Distributions of mesh clients	Uniform distribution
Number of mesh routers	16
Number of mesh clients	48
PHY protocol	IEEE 802.11b
Propagation loss model	Log-distance Path Loss Model
Propagation delay model	Constant Speed Model
MAC protocols	DCF
Routing protocol	OLSR
Transport protocol	UDP
Application type	CBR
Packet size	1024 [Bytes]
Transmission rates	100, 600, 1200 [kbps]
Number of source nodes	10
Number of destination node	1
Transmission current	17.4 [mA]
Receiving current	19.7 [mA]
Simulation time	600 [sec]

and pure C++ simulations. The function of ns-3 visualizer is more powerful than network animator (*nam*) of ns-2 simulator.

The ns-3 simulator has models for all network elements that comprise a computer network. For example, network devices represent the physical device that connects a node to the communication channel. This might be a simple Ethernet network interface card or a more complex wireless IEEE 802.11 device.

The ns-3 is intended as an eventual replacement for popular ns-2 simulator. The ns-3's wifi models a wireless network interface controller based on the IEEE 802.11 standard [18]. The ns-3 provides models for these aspects of 802.11:

1. Basic 802.11 DCF with infrastructure and ad hoc modes.
2. 802.11a, 802.11b, 802.11g and 802.11s physical layers.
3. QoS-based EDCA and queueing extensions of 802.11e.
4. Various propagation loss models including Nakagami, Rayleigh, Friis, LogDistance, FixedRss, and so on.
5. Two propagation delay models, a distance-based and random model.
6. Various rate control algorithms including Aarf, Arf, Cara, Onoe, Rraa, ConstantRate, and Minstrel.

3.5 Overview of OLSR Routing Protocol

The OLSR protocol [19] is a pro-active routing protocol, which builds up a route for data transmission by maintaining a routing table inside every node of the network. The routing table is computed upon the knowledge of topology information, which is exchanged by means of Topology Control (TC) packets.

OLSR makes use of HELLO messages to find its one hop neighbours and its two hop neighbours through their responses. The sender can then select its Multi Point Relays (MPR) based on the one hop node which offer the best routes to the two hop nodes. By this way, the amount of control traffic can be reduced. Each node has also an MPR selector set which enumerates nodes that have selected it as an MPR node. OLSR uses TC messages along with MPR forwarding to disseminate neighbour information throughout the network. Host Network Address (HNA) messages are used by OLSR to disseminate network route advertisements in the same way TC messages advertise host routes.

4 Simulation Results

We used the throughput, delay, jitter and fairness index metrics to evaluate the performance of WMNs for two architectures considering transmission rate and uniform distribution.

In Fig. 4, we show the simulation results of throughput. For transmission rates 600 and 1200 [kbps], the throughput of both architectures are higher than transmission rate 100 [kbps].

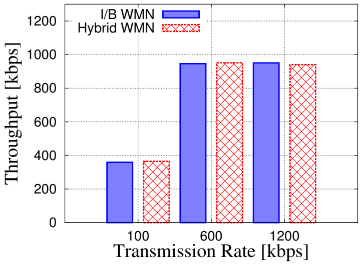


Fig. 4 Results of average throughput.

In Fig. 5 and Fig. 6, for transmission rate 100 [kbps], the delay and jitter of Hybrid WMN is lower than I/B WMN. But, for 1200 [kbps] the delay of I/B WMN is lower than Hybrid WMN.

In Fig. 7, we show the fairness index. For transmission rate 1200 [kbps], the fairness index of I/B WMN is higher than Hybrid WMN.

5 Conclusions

In this work, we presented WMN-GA system and applied it for node placement problem in WMNs. We evaluated the performance of WMN-GA system for uniform distribution of mesh clients considering transmission rate and OLSR protocols.

From the simulations we found that:

- For transmission rates 600 and 1200 [kbps], the throughput of both architectures are higher than transmission rate 100 [kbps].
- For transmission rate 100 [kbps], the delay and jitter of Hybrid WMN is lower than I/B WMN. But, for 1200 [kbps] the delay of I/B WMN is lower than Hybrid WMN.
- For transmission rate 1200 [kbps], the fairness index of I/B WMN is higher than Hybrid WMN.

In the future work, we would like to implement other systems and compare the performance with proposed system.

References

1. I.F. Akyildiz, X. Wang, W. Wang, “Wireless Mesh Networks: A Survey”, In Computer Networks, Vol. 47, No. 4, pp. 445-487, 2005.
2. A. Franklin, C. Murthy “Node Placement Algorithm for Deployment of Two-Tier Wireless Mesh Networks”, In: IEEE GLOBECOM-2007, pp. 4823-4827, 2007.

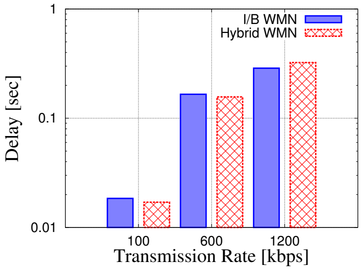


Fig. 5 Results of average delay.

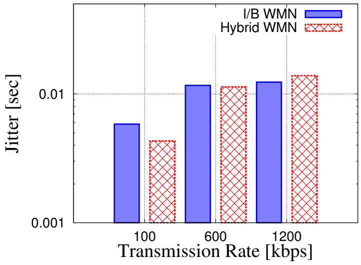


Fig. 6 Results of average jitter.

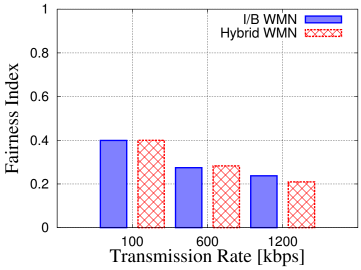


Fig. 7 Results of fairness index.

3. S. N. Muthaiah and C. Rosenberg, "Single Gateway Placement in Wireless Mesh Networks", In Proc. of 8th International IEEE Symposium on Computer Networks, Turkey, pp. 4754-4759, 2008.
4. M. Tang, "Gateways Placement in Backbone Wireless Mesh Networks", International Journal of Communications, Network and System Sciences, Vol. 2, No.1, pp. 45-50, 2009.
5. T. Vanhatupa, M. Hännikäinen and T.D. Hämäläinen, "Genetic Algorithm to Optimize Node Placement and Configuration for WLAN Planning", In Proc. of 4th International Symposium on Wireless Communication Systems, pp. 612-616, 2007.
6. A. Lim, B. Rodrigues, F. Wang and Zh. Xua, "$k-$Center Problems with Minimum Coverage", Theoretical Computer Science, Vol. 332, No. 1-3, pp. 1-17, 2005.
7. J. Wang, B. Xie, K. Cai and D. P. Agrawal, "Efficient Mesh Router Placement in Wireless Mesh Networks", MASS, Pisa, Italy, pp. 9-11, 2007.
8. T. Oda, A. Barolli, F. Xhafa, L. Barolli, M. Ikeda, M. Takizawa, "WMN-GA: A Simulation System for WMNs and Its Evaluation Considering Selection Operators", Journal of Ambient

Intelligence and Humanized Computing (JAIHC), Springer, Vol. 4, No. 3, pp. 323-330, June 2013

9. M. Ikeda, T. Oda, E. Kulla, M. Hiyama, L. Barolli and M. Younas, “Performance Evaluation of WMN Considering Number of Connections Using NS-3 Simulator”, The Third International Workshop on Methods, Analysis and Protocols for Wireless Communication (MAPWC 2012), pp. 498-502, Victoria, Canada, November 12-14, 2012.
10. T. Oda, D. Elmazi, A. Barolli, S. Sakamoto, L. Barolli, F. Xhafa, “A Genetic Algorithm Based System for Wireless Mesh Networks: Analysis of System Data Considering Different Routing Protocols and Architectures”, Journal of Soft Computing (SOCO), Springer, Published online: 31 March 2015, DOI: 10.1007/s00500-015-1663-z, pp. 1-14, 2015.
11. T. Oda, Y. Liu, S. Sakamoto, D. Elmazi, L. Barolli, F. Xhafa, “Analysis of Mesh Router Placement in Wireless Mesh Networks Using Friedman Test Considering Different Metaheuristics”, International Journal of Communication Networks and Distributed Systems (IJCNDS), Inderscience, Vol. 15, No. 1, pp. 84-106, 2015.
12. F. Xhafa, C. Sanchez, and L. Barolli, “Locals Search Algorithms for Efficient Router Nodes Placement in Wireless Mesh Networks”, in International Conference on Network-Based Information Systems (NBiS), pp. 572-579, 2009.
13. T. Oda, A. Barolli, E. Spaho, L. Barolli, F. Xhafa, “Analysis of Mesh Router Placement in Wireless Mesh Networks Using Friedman Test”, Proc. of The 28th IEEE International Conference on Advanced Information Networking and Applications (IEEE AINA), pp. 289-296, Victoria, Canada, May 2014,
14. T. Oda, S. Sakamoto, A. Barolli, M. Ikeda, L. Barolli, F. Xhafa, “A GA-Based Simulation System for WMNs: Performance Analysis for Different WMN Architectures Considering TCP”, 2014 Eighth International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA), pp. 120-126, Guangzhou, China, November 2014.
15. T. Oda, A. Barolli, E. Spaho, F. Xhafa, L. Barolli, M. Takizawa, ”Evaluation of WMN-GA for Different Mutation Operators”, International Journal of Space-Based and Situated Computing (IJSSC), Inderscience, Vol. 2. No. 3, pp. 149-157, 2012.
16. “ns-3”, https://www.nsnam.org/.
17. “The Network Simulator-ns-2”, http://www.isi.edu/nsnam/ns/.
18. IEEE 802.11, “Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications”, IEEE Computer Society Std., June 2007. [Online]. Available: http://standards.ieee.org/getieee802/download/802.11-2007.pdf
19. T. Clausen and P. Jacquet, “Optimized Link State Routing Protocol (OLSR)”, RFC 3626 (Experimental), 2003.

Effect of Node Density and Node Movement Model on Performance of a VDTN

Kevin Bylykbashi, Evjola Spaho, Leonard Barolli and Makoto Takizawa

Abstract In this paper, we evaluate the effect of node density and node movement model in a many-to-one communication in a Vehicular Delay Tolerant Network (VDTN). Seven groups with three stationary sensor nodes sense the temperature, humidity and wind speed and send these data to a stationary destination node that collect them for statistical and data analysis purposes. Vehicles moving in Tirana city roads during the opportunistic contacts will exchange the sensed data to destination node. The simulations are conducted with the Opportunistic Network Environment (ONE) simulator. For the simulations we considered two different scenarios where the distance of the source nodes from the destination is short and long. The performance is analyzed for three routing protocols for delivery probability and average latency metrics. For both scenarios the effect of node density and node movement model is evaluated. The simulation results show that the increase of node density increases the delivery probability for all protocols and both scenarios, and better results are achieved when shortest-path map-based movement model is used.

Kevin Bylykbashi
Faculty of Information Technology, Polytechnic University of Tirana, Mother Teresa Square, No. 4, Tirana, Albania, e-mail: kevin.bylykbashi@fti.edu.al

Evjola Spaho
Department of Electronics and Telecommunication, Faculty of Information Technology, Polytechnic University of Tirana, Mother Teresa Square, No. 4, Tirana, Albania, e-mail: evjolaspaho@hotmail.com

Leonard Barolli
Department of Information and Communication Engineering, Fukuoka Institute of Technology (FIT), 3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811-0295, Japan, e-mail: barolli@fit.ac.jp

Makoto Takizawa
Department of Advanced Sciences, Hosei University, 3-7-2, Kajino-cho, Koganei-shi, Tokyo 184-8584, Japan, e-mail: makoto.takizawa@computer.org

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_14

1 Introduction

Vehicular networks are characterized by the lack of an end-to-end multi-hop path most of the time, which is caused by a highly dynamic network topology and network partitioning due to low node density and large distances. In such network environments, a complete path from source to destination does not exist most of the time.

In order to deal with these connectivity constrains, opportunistic networks, DTNs (Delay Tolerant Networks) and VDTNs (Vehicular Delay Tolerant Networks) are introduced. DTNs are a class of networks that enable communication where connectivity issues like sparse connectivity, intermittent connectivity, high latency, long delay, high error rates, asymmetric data rate, and even no end-to-end connectivity exists. VDTNs are a particular application of DTNs where vehicles are exploited to offer a message relaying service.

In order to handle disconnections and long delays in vehicular network scenarios, VDTN uses store-carry-and-forward approach. VDTNs have the potential to interconnect Vehicles in regions that current networking technology cannot reach.

VDTNs aim to support a class of vehicular network applications characterized by delay tolerant and asynchronous data traffic. VDTN is supposed to be one of the effective methods to transmit significant data even under poor network conditions. In VDTN the communication is asynchronous, bundle-oriented, and a store-carry-and-forward routing paradigm is used. Instead of working end-to-end, in VDTNs, a message-oriented overlay layer called Bundle layer employs a store-carry-and-forward message switching paradigm that moves messages from node to node, along a path that eventually reaches the destination.

In this paper, we evaluate the effect of node density and node movement model in a VDTN. We compare the performance of three different routing protocols in a many-to-one communication network. Two scenarios were designed with short and long distance of source nodes from the destination. For the simulations we use the Opportunistic Network Environment (ONE) [1] simulator.

ONE is a simulation environment, capable of generating node movement using different movement models. ONE offers various DTN routing algorithms for routing messages between nodes. Its graphical user interface visualize both mobility and message passing in real time. ONE can import mobility data from real-world traces or other mobility generators. It can also produce a variety of reports from node movement to message passing and general statistics.

Performance evaluation results, based on simulation, show that the increase of node density increases the delivery probability for all protocols and both scenarios, and better results are achieved when shortest-path map-based movement model is used.

The remainder of the paper is organized as follows. Section 2 introduces DTN and routing protocols. The simulation system design and description is presented in Section 3. In Section 4 are shown the simulation results. Finally, the conclusions and future work are presented in Section 5.

2 DTNs and Routing Protocols

2.1 DTN Overview

DTN are occasionally connected networks, characterized by the absence of a continuous path between the source and destination [2], [3]. The data can be transmitted by storing them at nodes and forwarding them later when a link is established. This technique is called message switching. Eventually the data will be relayed to the destination. DTN is the "challenged computer network" approach that is originally designed from the Interplanetary Internet, and the data transmission is based upon the store-carry-and-forward protocol for the sake of carrying data packets under a poor network environment such as space [2]. Different copies of the same bundle can be routed independently to increase security and robustness, thus improving the delivery probability and reducing the delivery delay. However, such approach increases the contention for network resources (e.g., bandwidth and storage), potentially leading to poor overall network performance.

In [4], authors have studied this model and found that it can provide substantial capacity at little cost, and that the use of a DTN model often doubles that capacity compared with a traditional end-to-end model. The main assumption in the Internet that DTNs seek to relax is that an end-to-end path between a source and a destination exists for the entire duration of a communication session. When this is not the case, the normal Internet protocols fail. DTNs get around the lack of end-to-end connectivity with an architecture that is based on message switching. It is also intended to tolerate links with low reliability and large delays. The architecture is specified in RFC 4838 [5].

Bundle protocol has been designed as an implementation of the DTN architecture. A bundle is a basic data unit of the DTN bundle protocol. Each bundle comprises a sequence of two or more blocks of protocol data, which serve for various purposes. In poor conditions, bundle protocol works on the application layer of some number of constituent Internet, forming a store-and-forward overlay network to provide its services. The bundle protocol is specified in RFC 5050. It is responsible for accepting messages from the application and sending them as one or more bundles via store-carry-and-forward operations to the destination DTN node. The bundle protocol runs above the TCP/IP level.

2.2 Routing Protocols

In order to handle disconnections and long delays in sparse opportunistic vehicular network scenarios, VDTN uses store-carry-and-forward approach. A network node stores a bundle and waits for a future opportunistic connection. When the connection is established, the bundle is forwarded to an intermediate node, according to a hop-by-hop forwarding/routing scheme. This process is repeated and the bundle will be

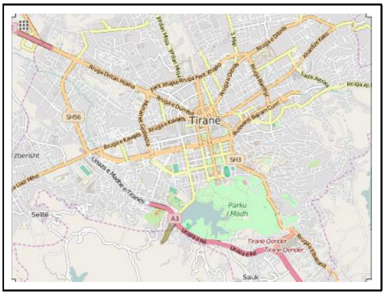


Fig. 1 Tirana city map imported from osm.

relayed hop-by-hop until reaching the destination node. In [6], [7], [8], [9], [10], [11], [12], [16], [13] authors deal with routing in DTNs.

In this work, we will use three widely applicable DTN routing protocols Spray and Wait [14], Maxprop [15] and Prophet [16].

Spray and Wait routing protocol: Spray and Wait [14], is a routing protocol that attempts to gain the delivery ratio benefits of replication-based routing as well as the low resource utilization benefits of forwarding-based routing. The Spray and Wait protocol is composed of two phases: the spray phase and the wait phase. When a new message is created in the system, a number L is attached to that message indicating the maximum allowable copies of the message in the network. During the spray phase, the source of the message is responsible for ”spraying”, or delivery, one copy to L distinct “relays”. When a relay receives the copy, it enters the wait phase, where the relay simply holds that particular message until the destination is encountered directly.

Maxprop routing protocol: Maxprop [15], is based on prioritizing both the schedule of packets transmitted to other peers and the schedule of packets to be dropped. These priorities are based on the path likelihoods to peers according to historical data and also on several complementary mechanisms, including acknowledgments, a head-start for new packets, and lists of previous intermediaries.

Prophet routing protocol: Prophet (Probabilistic Routing Protocol using History of Encounters and Transitivity) [16] is a variant of the epidemic routing protocol for intermittently connected networks that operates by pruning the epidemic distribution tree to minimize resource usage while still attempting to achieve the best case routing capabilities of epidemic routing. It uses a probabilistic metric: delivery predictability, that attempts to estimate, based on node encounter history, which node has the higher probability of successful delivery of a message to the final destination. When two nodes are in communication range, a new message copy is transferred only if the other node has a better probability of delivering it to the destination.

3 Simulation system and design

The network scenario is based on the map-based model of a part of Tirana city (Albania). The map was imported from Open Street Map [17] (see Fig. 1). Simulations are carried out using the ONE simulator. We simulated an urban scenario where vehicles move on the map roads with a speed between 10-50 km/h during 4 hours period of time. There are 7 groups with 3 nodes each group that sends data to a destination node (many-to-one communication). These 21 nodes can be sensors that gather information about temperature, humidity and wind speed and sends their data to a node that collects these data for statistical purposes. Source and destination nodes are stationary and have a 100 MB buffer. Other nodes are vehicles equipped with a 100 MB buffer. We considered two scenarios: in the first one the distance of the source nodes is between 400 m - 1000 m and the second one where the distance is 800 m - 2000 m. The initial position of all the nodes for the first scenario and the second scenario are shown in Fig. 2.

All network nodes use a WiFi link connection with a transmission data rate of 250 KBps and the transmission range is considered 20 m. We use map-based movement model and shortest-path map-based movement model for the vehicles. Shortest-path map-based movement model, initially places the nodes in random places, but selects a certain destination in the map for all nodes and uses Dijkstra‘s shortest path algorithm to find the shortest path to the destination.

The event generator is responsible for generating bundles with sizes uniformly distributed in the ranges [10kB, 50kB]. A bundle is created every 30 s and data bundles ttl is 30 min. The simulation parameters are shown in Table 1.

We evaluate the performance of the system for 3 different routing protocols: Spray and Wait, Maxprop, Prophet for different node densities and two different movement models.

We use the following metrics to measure the performance of different routing protocols: delivery probability and average latency.

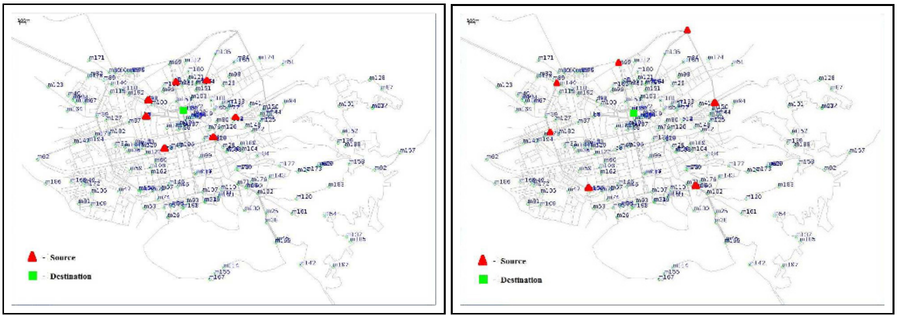
- **Delivery probability** is the ratio of number of delivered messages to that of created messages.
- **Average latency** is the average time elapsed from the creation of the messages at source to their successful delivery to the destination.

4 Simulation Results

In this section, we present the simulation results of the above described routing protocols. In Fig. 3 are shown the simulation results of number of nodes vs. delivery probability for all considered routing protocols when map-based movement model is used. For both scenarios, the increase of node density increases the delivery probability for all protocols. This is related with the increase of the number of opportunistic contacts between nodes. Better results are achieved in the first scenario

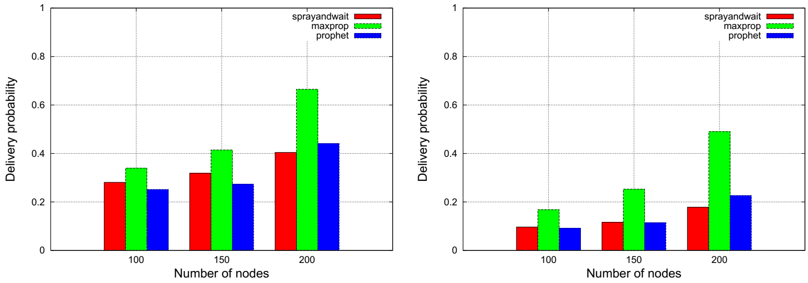
Table 1 Simulation Parameters and their values.

Parameters	Values
Number of nodes	200
Simulation time	14400 s
Map size	5 km *x* 3.5 km
Movement Model	Map-based, Shortest-path map-based
Buffer size	100 MB
Interface type	WiFi
Interface Transmission Speed	250 MBps
Interface Transmission Range	20 m
Message TTL	30 min
Vehicles speed	10-50 km/h
Message size	10k, 50k
Warm up time	100 s
Events interval	30 s



(a) First scenario (b) Second scenario

Fig. 2 Nodes initial positions for both scenarios.

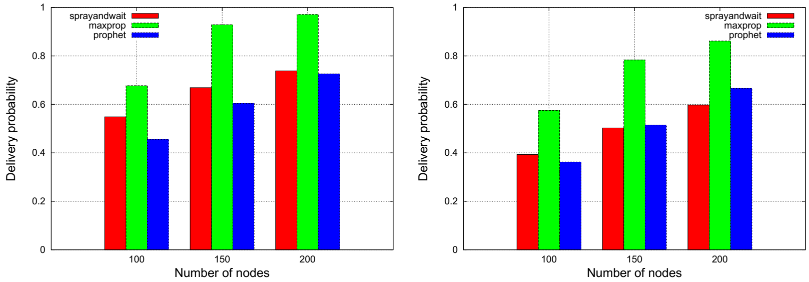


(a) First scenario (b) Second scenario

Fig. 3 Results of number of nodes vs. delivery probability for map based movement.

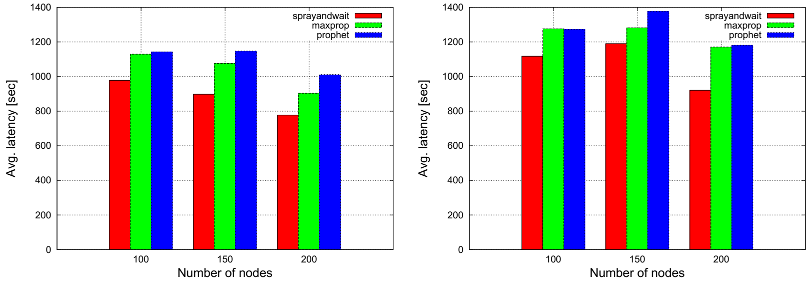
where the distance between sources and the destination is short. The simulation results show that best performance is achieved for Maxprop routing protocol.

The simulation results of number of nodes vs. delivery probability for shortest-path map-based movement are presented in Fig. 4. Shortest-path map-based movement model is a more sophisticated model compared with map-based movement model and for both scenarios it achieves better delivery probability. Maxprop per-



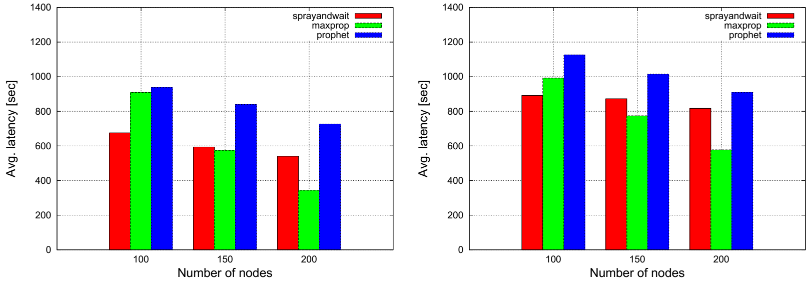
(a) First scenario (b) Second scenario

Fig. 4 Results of number of nodes vs. delivery probability for shortest path map based movement.



(a) First scenario (b) Second scenario

Fig. 5 Results of number of nodes vs. avg. latency for map based movement.



(a) First scenario (b) First scenario

Fig. 6 Results of number of nodes vs. avg. latency for shortest path map based movement.

formance is higher compared with sprayandwait and prophet. Best results are for dense network with 200 nodes where delivery probability is 98%.

In Fig. 5 are presented the results for avg. latency when map-based movement model is used. From the figure it can be noticed that the avg. latency is shorter for the first scenario because the distance between the communicating nodes is shorter.

Spray and wait results in lower latency than other protocols. Low avg. latency results are achieved for dense networks for all protocols. In dense mobile networks the probability of nodes to encounter other nodes is high and opportunistic contact happen often.

The evaluation results of number of nodes vs. avg. latency when the shortest-path map-based movement is used are shown in Fig. 6. The avg. latency of all protocols is lower compared with map-based movement model because the Dijkstra algorithm finds the shortest path to the destination. In both scenarios, best results are for dense network where maxprop is used.

5 Conclusions

In this work, we evaluated the effect of node density and node movement model on the performance of three routing protocols (maxprop, sprayandwait and prophet) in a many-to-one communication opportunistic network scenario for short and long distances between source and destination nodes. For evaluation, we considered delivery probability and avg. latency metrics.

The performance study showed the following results.

- For both scenarios, the increase of node density increases the delivery probability for all protocols.
- Better results of delivery probability are achieved in the first scenario where the distance between sources and the destination is short.
- In both scenarios, best performance in terms of delivery probability is achieved for Maxprop routing protocol.
- Shortest-path map-based movement model is a more sophisticated model compared with map-based movement model and for both scenarios it achieves better delivery probability.
- In both scenarios the avg. latency is smaller for dense networks because the probability of nodes to encounter other nodes is high and opportunistic contact happen often.
- The avg. latency of all protocols in shortest-path map-based movement is lower compared with map-based movement model because the Dijkstra algorithm finds the shortest path to the destination.

In this work, we considered a communication network with multiple sources and a single destination node. In the future, we would like to consider multiple sources and destinations and make extensive simulations to evaluate the performance of different routing protocols considering different scenarios and parameters.

References

1. A. Keranen, J. Ott, and T. Karkkainen, “The ONE Simulator for DTN Protocol Evaluation,” in Proceedings of the 2-nd International Conference on Simulation Tools and Techniques (SIMUTools-2009), 2009, http://www.netlab.tkk.fi/tutkimus/ dtn/theone/pub/the one simutools.pdf.
2. K. Fall, “A delay-tolerant network architecture for challenged Internets,” in Proceedings of the International Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, ser. SIGCOMM 03, 2003, pp. 2734.
3. “Delay- and disruption-tolerant networks (DTNs) tutorial,” NASA/JPLs Interplanetary Internet (IPN) Project, 2012, http: //www.warthman.com/images/DTN Tutorial v2.0.pdf.
4. N. Laoutaris, G. Smaragdakis, P. Rodriguez, and R. Sundaram, “Delay tolerant bulk data transfers on the Internet,” in Proceedings of the 11-th International Joint Conference on Measurement and Modeling of Computer Systems (SIGMETRICS09), 2009, pp. 22923.
5. V. Cerf, S. Burleigh, A. Hooke, L. Torgerson, R. Durst, K. Scott, K. Fall, and H. Weiss, “Delay-tolerant networking architecture,” IETF RFC 4838 (Informational), April 2007.
6. K. Massri, A. Vernata, A. Vitaletti, “Routing Protocols for Delay Tolerant Networks: a Quantitative Evaluation,” In Proceedings of ACM workshop PM2HW2N’12, pp.107-114, 2012.
7. S. Ishikawa, T. Honda, M. Ikeda, and L. Barolli, “Performance analysis of vehicular DTN routing under urban environment,” in Proceedings of CISIS-2014, July 2014.
8. M. Demmer, K. Fall. “DTLSR: Delay Tolerant Routing for Developing Regions,” in Proceedings of the 2007 ACM workshop on Networked systems for developing regions, 6 pages, 2007.
9. A. A. Ilham, M. Niswar, Agussalim, “Evaluated and Optimized of Routing Model on Delay Tolerant Network (DTN) for Data Transmission to Remote Area” In Proceedings of FORTEI, Indonesia University Jakarta, pp.24-28, 2012.
10. S. Jain, K. Fall, and R. Patra, “Routing in a Delay Tolerant Network,” in Proceedings of ACM SIGCOMM 2004 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication, Portland, Oregon, USA, August 30-September 3, 2004, pp. 145-158, 2004.
11. Z. Zhang, “Routing In Intermittently Connected Mobile Ad Hoc Networks And Delay,” Communications Surveys & Tutorials, IEEE (Volume:8, Issue:1), pp. 24-37, January 2006.
12. V. N. G. J. Soares, J. J. P. C. Rodrigues, and F. Farahmand, “GeoSpray: a geographic routing protocol for vehicular delay-tolerant networks,” Information Fusion, vol. 15, no. 1, pp. 102-113, 2014.
13. A. Vahdat, D. Becker “Epidemic Routing for Partially Connected Ad Hoc Networks,” Technical Report CS-200006, Duke University, April 2000.
14. T. Spyropoulos, K. Psounis, C.S. Raghavendra, “Spray and Wait: an efficient routing scheme for intermittently connected mobile networks,” In Proceedings of ACM SIGCOMM 2005 Workshop on Delay Tolerant Networking and Related Networks (WDTN-05), Philadelphia, PA, USA, pp. 252-259, 2005.
15. J. Burgess, B. Gallagher, D. Jensen, and B. N. Levine, “Maxprop: Routing for Vehicle-Based Disruption-Tolerant Networks”in Proceedings of the IEEE Infocom, April 2006.
16. A. Lindgren, A. Doria, E. Davies, S. Grasic, “Probabilistic Routing Protocol for Intermittently Connected Networks,” draft-irtf-dtnrg-prophet-09. (http:// tools.ietf.org/html/draft-irtf-dtnrg-prophet-09).
17. “Open street map,” http://www.openstreetmap.org/.

A Fuzzy-Based Simulation System for Actor Selection in Wireless Sensor and Actor Networks Considering as a New Parameter Density of Actor Nodes

Donald Elmazi, Tetsuya Oda, Evjola Spaho, Elis Kulla, Makoto Ikeda, Leonard Barolli

Abstract Wireless Sensor and Actor Networks (WSANs), refers to a group of sensors and actors that get the information about the physical environment and perform appropriate actions. In order to provide effective sensing and acting, a distributed local coordination mechanism is necessary among sensors and actors. In this work, we propose a fuzzy-based system for selection in WSANs. Our system uses four input parameters. Different from our previous work, we consider also the Density of Actor (DOA) parameter. The system output is Actor Selection Decision (ASD). The simulation results show that the proposed system has a good behaviour and makes a proper selection of actor nodes.

Donald Elmazi
Graduate School of Engineering,
Fukuoka Institute of Technology (FIT),
3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811-0295, Japan,
e-mail: shinji.t.sakamoto@gmail.com

Tetsuya Oda, Makoto Ikeda, Leonard Barolli
Department of Information and Communication Engineering,
Fukuoka Institute of Technology (FIT)
3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811-0295, Japan,
e-mail: oda.tetsuya.fit@gmail.com, makoto.ikd@acm.org, barolli@fit.ac.jp

Evjola Spaho
Department of Electronics and Telecommunication,
Polytechnic University of Tirana
Bul. Deshmoret e Kombit, Mother Theresa Square, Nr. 4, Tirana, Albania
e-mail: evjolaspaho@hotmail.com

Elis Kulla
Department of Information and Computer Engineering,
Okayama University of Science
1-1 Ridai-cho, Kita-Ku, Okayama 700-0005, Japan
e-mail: kulla@ice.ous.ac.jp

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_15

1 Introduction

Wireless Sensor and Actor Networks (WSANs), have emerged as a variation of WSNs. WSNs can be defined as a collection of wireless self-configuring programmable multi-hop tiny devices, which can bind to each other in an arbitrary manner, without the aid of any centralized administration, thereby dynamically sending the sensed data to the intended recipient about the monitored phenomenon [1].

WSANs are capable of monitoring physical phenomenons, processing sensed data, making decisions based on the sensed data and completing appropriate tasks when needed [2]. For example, in the case of a fire, sensors relay the exact origin and intensity of the fire to actors so that they can extinguish it before spreading in the whole building or in a more complex scenario, to save people who may be trapped by fire.

Unlike WSNs, where the sensor nodes tend to communicate all the sensed data to the sink by sensor-sensor communication, in WSANs, two new communication types may take place. They are called sensor-actor and actor-actor communications. Sensed data is sent to the actors in the network through sensor-actor communication. After the actors analyse the data, they communicate with each other in order to assign and complete tasks. To provide effective operation of WSAN, is very important that sensors and actors coordinate in what are called sensor-actor and actor-actor coordination. Coordination is not only important during task conduction, but also during network's self-improvement operations, i.e. connectivity restoration [3, 4], reliable service [5], Quality of Service (QoS) [6, 7] and so on.

Sensor-Actor (SA) coordination defines the way sensors communicate with actors, which actor is accessed by each sensor and which route should be selected to transmit data packets. Among other challenges, when designing SA coordination, the energy minimization should be considered. On the other hand, by Actor-Actor (AA) coordination can be selected which actor will lead performing the task (actor selection), how many actors should perform and how they will perform. Actor selection is not a trivial task, because it needs to be solved in real time, considering different factors. It becomes more complicated when the actors are moving, due to dynamic topology of the network.

In this paper, different from our previous work [8], we propose and implement a simulation system which considers also the Density of Actor nodes (DOA) parameter.

The system is based on fuzzy logic and considers four input parameters for actor selection. We show the simulation results for different values of parameters.

The remainder of the paper is organized as follows. In Section 2, we describe the basics of WSANs including research challenges and architecture. In Section 3, we describe the system model and its implementation. Simulation results are shown in Section 4. Finally, conclusions and future work are given in Section 5.

2 WSAN

2.1 WSAN Challenges

Some of the key challenges in WSAN are related to the presence of actors and their functionalities.

- *Deployment and Positioning:* At the moment of node deployment, algorithms must consider to optimize the number of sensors and actors and their initial positions based on applications [9, 10].
- *Architecture:* When important data has to be transmitted (an event occurred), sensors may transmit their data back to the sink, which will control the actors' tasks from distance or transmit their data to actors, which can perform actions independently from the sink node [11].
- *Real-Time:* There are a lot of applications that have strict real-time requirements. In order to fulfill them, real-time limitations must be clearly defined for each application and system [12].
- *Coordination:* In order to provide effective sensing and acting, a distributed local coordination mechanism is necessary among sensors and actors [11].
- *Power Management:* WSAN protocols should be designed with minimized energy consumption for both sensors and actors [13].
- *Mobility:* Protocols developed for WSANs should support the mobility of nodes [4, 14], where dynamic topology changes, unstable routes and network isolations are present.
- *Scalability:* Smart Cities are emerging fast and WSAN, as a key technology will continue to grow together with cities. In order to keep the functionality of WSAN applicable, scalability should be considered when designing WSAN protocols and algorithms [10, 14].

2.2 WSAN Architecture

A WSAN is shown in Fig. 1. The main functionality of WSANs is to make actors perform appropriate actions in the environment, based on the data sensed from sensors and actors. When important data has to be transmitted (an event occurred), sensors may transmit their data back to the sink, which will control the actors' tasks from distance, or transmit their data to actors, which can perform actions independently from the sink node. Here, the former scheme is called Semi-Automated Architecture and the latter one Fully-Automated Architecture (see Fig. 2). Obviously, both architectures can be used in different applications. In the Fully-Automated Architecture are needed new sophisticated algorithms in order to provide appropriate coordination between nodes of WSAN. On the other hand, it has advantages, such as *low latency*, *low energy consumption*, *long network lifetime* [2], *higher local position accuracy*, *higher reliability* and so on.

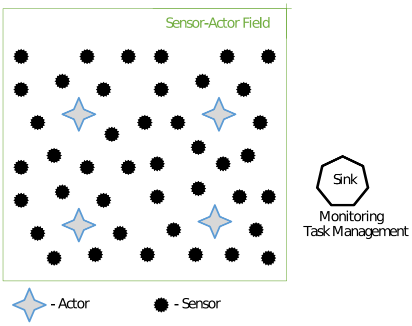
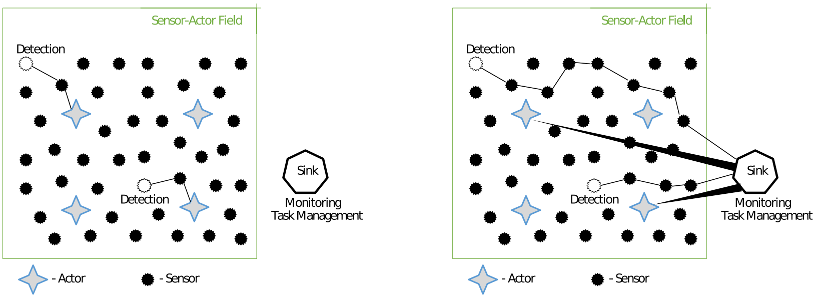


Fig. 1 Wireless Sensor Actor Network (WSAN).



(a) Fully-Automated

(b) Semi-Automated

Fig. 2 WSAN architectures.

3 Proposed System Model

3.1 Problem Description

After data has been sensed from sensors, they are collected to the sink for semi-automated architecture or spread to the actors for fully-automated architecture. Then a task is assigned to actors. In general, one or more actors take responsibility and perform appropriate actions. Different actors may be chosen for acting, depending on their characteristics and conditions. For example, if an intervention is required in a building, a flying robot can go there faster and easier. While, if a kid is inside a room in fire, it is better to send a small robot. The issue here is which of the actors will be selected to respond to critical data collected from the field (actor selection). If WSAN uses semi-automated architecture, the sinks are used to collect data and control the actors. They may be supplied with detailed information about actors characteristics (size, ability etc.). If fully-automated architecture is being used, the collected data are processed only by actors, so they first have to decide whether they have the proper ability and right conditions to perform. Soon after that, actors coor-

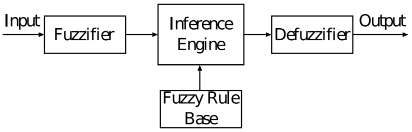


Fig. 3 FLC structure.

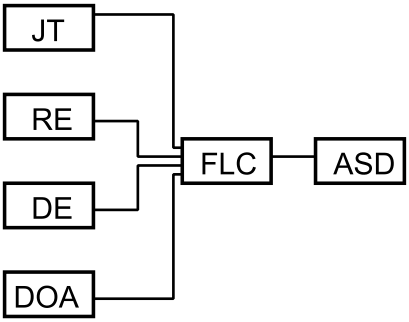


Fig. 4 Proposed System.

dinate with each-other, to decide more complicated procedures like acting multiple actors, or choosing the most appropriate one from several candidates. In this work, we propose a fuzzy-based system in order to select an appropriate actor node for a required task.

3.2 System Parameters

Based on WSAN characteristics and challenges, we consider the following parameters for implementation of our proposed system.

Job Type (JT): A sensed event may be triggered by various causes, such as when water level passed a certain height of the dam. Similarly, for solving a problem, actors need to perform actions of different types. Actions may be classified regarding time duration, complexity, working force required etc., and then assign a priority to them, which will guide actors to make their decisions. In our system, JT is defined by five levels of difficulty. The hardest the task, the more likely an actor is to be selected.

Distance to Event (DE): The number of actors in a WSAN is smaller than the number of sensors. Thus, when an actor is called for action near an event, the distance from the actor to the event is different for different actors and events. Depending on three distance levels, our system takes decisions on the availability of the actor node.

Remaining Energy (RE): As actors are active in the monitored field, they perform tasks and exchange data in different ways from each other. Consequently, also based on their characteristics, some actors may have a lot of power remaining and

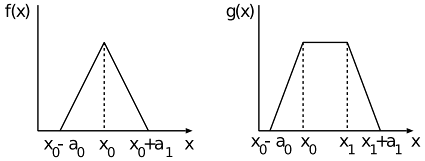


Fig. 5 Triangular and trapezoidal membership functions.

Table 1 Parameters and their term sets for FLC.

Parameters	Term Sets
Job Type (JT)	Easy (Ea), Medium (Me), Hard (Ha)
Distance to Event (DE)	Near (Ne), Middle (Mi), Far (Fa)
Remaining Energy (RE)	Low (L), Middle (M), High (H)
Density of Actors (DOA)	Spare (SP), Normal (Nrm), Dense (DN)
Actor Selection Decision (ASD)	VLSP, LSP, MSP, HSP, VHSP

other may have very little, when an event occurs. We consider three levels of RP for actor selection.

Density of Actors (DOA): The number of actor nodes can be different in various areas. When in an area we have spare actors, the probability to select an actor node is very high, otherwise if it is dense it has a low probability to be selected for carring out the task.

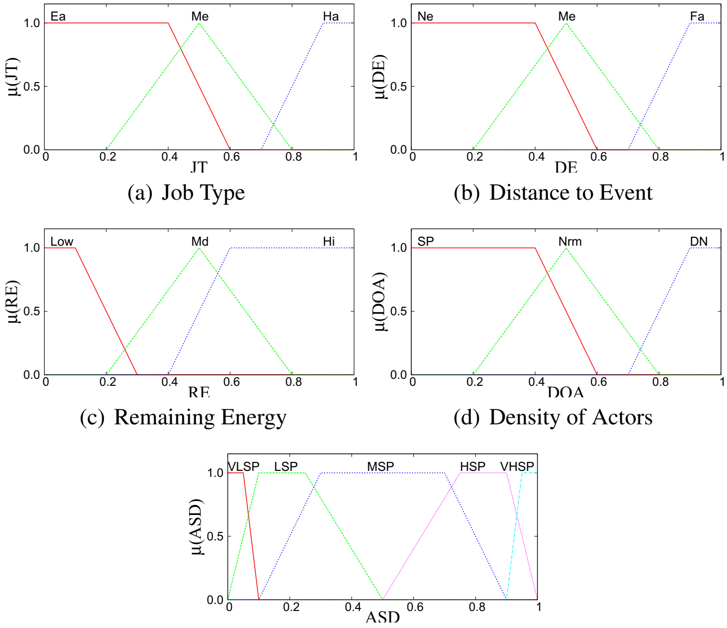
Actor Selection Decision (ASD): Our system is able to decide the willingness of an actor to be assigned a certain task at a certain time. The actors respond in five different levels, which can be interpreted as:

- Very Low Selection Possibility (VLSP) - It is not worth assigning the task to this actor.
- Low Selection Possibility (LSP) - There might be other actors which can do the job better.
- Middle Selection Possibility (MSP) - The Actor is ready to be assigned a task, but is not the “chosen” one.
- High Selection Possibility (HSP) - The actor takes responsibility of completing the task.
- Very High Selection Possibility (VHSP) - Actor has almost all required information and potential and takes full responsibility.

3.3 System Implementation

Fuzzy sets and fuzzy logic have been developed to manage vagueness and uncertainty in a reasoning process of an intelligent system such as a knowledge based system, an expert system or a logic control system [15–29]. In this work, we use fuzzy logic to implement the proposed system.

The structure of the proposed system is shown in Fig. 4. It consists of one Fuzzy Logic Controller (FLC), which is the main part of our system and its basic elements



(a) Job Type

(b) Distance to Event

(c) Remaining Energy

(d) Density of Actors

(e) Actor Selection Decision

Fig. 6 Fuzzy membership functions.

are shown in Fig. 3. They are the fuzzifier, inference engine, Fuzzy Rule Base (FRB) and defuzzifier.

As shown in Fig. 5, we use triangular and trapezoidal membership functions for FLC, because they are suitable for real-time operation [30]. The x_0 in $f(x)$ is the center of triangular function, $x_0(x_1)$ in $g(x)$ is the left (right) edge of trapezoidal function, and $a_0(a_1)$ is the left (right) width of the triangular or trapezoidal function. We explain in details the design of FLC in following.

3.4 Description of FLC

We use four input parameters for FLC:

- Job Type (JT);
- Distance to Event (DE);
- Remaining Energy (RE);
- Density of Actors (DOA);

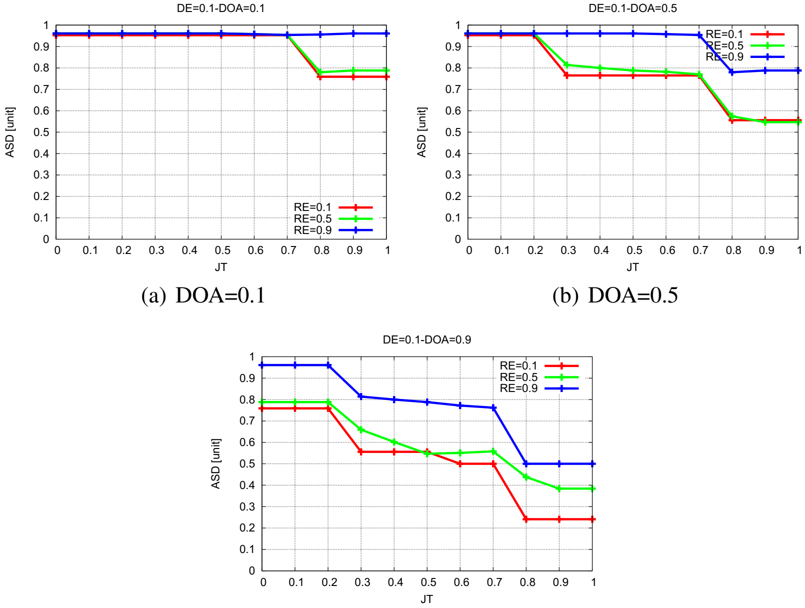
The term sets for each input linguistic parameter are defined respectively as shown in Table 1.

The output linguistic parameter is the Actor Selection Decision (ASD).

Table 2 FRB of proposed fuzzy-based system.

No.	JT	DE	RE	DOA	ASD	No.	JT	DE	RE	DOA	ASD
1	Ea	Ne	L	DN	VLSP	41	Me	Mi	M	Nrm	MSP
2	Ea	Ne	L	Nrm	LSP	42	Me	Mi	M	SP	MSP
3	Ea	Ne	L	SP	LSP	43	Me	Mi	H	DN	HSP
4	Ea	Ne	M	DN	LSP	44	Me	Mi	H	Nrm	HSP
5	Ea	Ne	M	Nrm	MSP	45	Me	Mi	H	SP	HSP
6	Ea	Ne	M	SP	MSP	46	Me	Fa	L	DN	VLSP
7	Ea	Ne	H	DN	MSP	47	Me	Fa	L	Nrm	VLSP
8	Ea	Ne	H	Nrm	HSP	48	Me	Fa	L	SP	LSP
9	Ea	Ne	H	SP	HSP	49	Me	Fa	M	DN	LSP
10	Ea	Mi	L	DN	VLSP	50	Me	Fa	M	Nrm	LSP
11	Ea	Mi	L	Nrm	VLSP	51	Me	Fa	M	SP	MSP
12	Ea	Mi	L	SP	LSP	52	Me	Fa	H	DN	MSP
13	Ea	Mi	M	DN	LSP	53	Me	Fa	H	Nrm	MSP
14	Ea	Mi	M	Nrm	LSP	54	Me	Fa	H	SP	HSP
15	Ea	Mi	M	SP	MSP	55	Ha	Ne	L	DN	MSP
16	Ea	Mi	H	DN	MSP	56	Ha	Ne	L	Nrm	MSP
17	Ea	Mi	H	Nrm	MSP	57	Ha	Ne	L	SP	MSP
18	Ea	Mi	H	SP	HSP	58	Ha	Ne	M	DN	HSP
19	Ea	Fa	L	DN	VLSP	59	Ha	Ne	M	Nrm	HSP
20	Ea	Fa	L	Nrm	VLSP	60	Ha	Ne	M	SP	HSP
21	Ea	Fa	L	SP	VLSP	61	Ha	Ne	H	DN	VHSP
22	Ea	Fa	M	DN	VLSP	62	Ha	Ne	H	Nrm	VHSP
23	Ea	Fa	M	Nrm	LSP	63	Ha	Ne	H	SP	VHSP
24	Ea	Fa	M	SP	LSP	64	Ha	Mi	L	DN	LSP
25	Ea	Fa	H	DN	LSP	65	Ha	Mi	L	Nrm	MSP
26	Ea	Fa	H	Nrm	MSP	66	Ha	Mi	L	SP	MSP
27	Ea	Fa	H	SP	MSP	67	Ha	Mi	M	DN	MSP
28	Me	Ne	L	DN	LSP	68	Ha	Mi	M	Nrm	HSP
29	Me	Ne	L	Nrm	LSP	69	Ha	Mi	M	SP	HSP
30	Me	Ne	L	SP	MSP	70	Ha	Mi	H	DN	HSP
31	Me	Ne	M	DN	MSP	71	Ha	Mi	H	Nrm	VHSP
32	Me	Ne	M	Nrm	MSP	72	Ha	Mi	H	SP	VHSP
33	Me	Ne	M	SP	HSP	73	Ha	Fa	L	DN	LSP
34	Me	Ne	H	DN	HSP	74	Ha	Fa	L	Nrm	LSP
35	Me	Ne	H	Nrm	HSP	75	Ha	Fa	L	SP	LSP
36	Me	Ne	H	SP	VHSP	76	Ha	Fa	M	DN	MSP
37	Me	Mi	L	DN	LSP	77	Ha	Fa	M	Nrm	MSP
38	Me	Mi	L	Nrm	LSP	78	Ha	Fa	M	Sp	MSP
39	Me	Mi	L	SP	LSP	79	Ha	Fa	H	DN	HSP
40	Me	Mi	M	DN	MSP	80	Ha	Fa	H	Nrm	HSP
						81	Ha	Fa	H	SP	HSP

The membership functions are shown in Fig. 6 and the Fuzzy Rule Base (FRB) is shown in Table 2. The FRB forms a fuzzy set of dimensions $|T(JT)| \times |T(DE)| \times |T(RE)| \times |T(DOA)|$, where $|T(x)|$ is the number of terms on $T(x)$. The FRB has 81 rules. The control rules have the form: IF “conditions” THEN “control action”.



(a) DOA=0.1

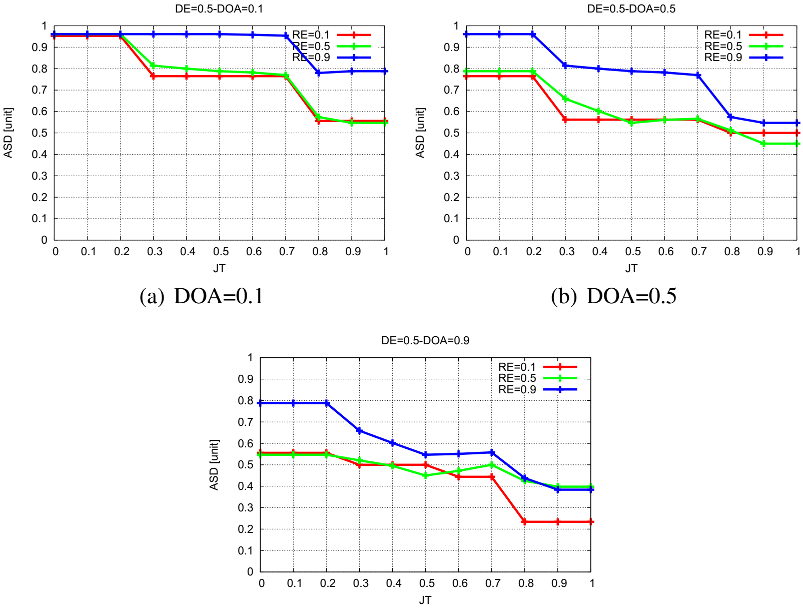
(b) DOA=0.5

(c) DOA=0.9

Fig. 7 Results for $DE = 0.1$.

4 Simulation Results

The simulation results are presented in Fig. 7, Fig. 8 and Fig. 9. From results, we found that as JT becomes difficult the ASD becomes higher because actors are programmed for different jobs. As we can see the performance is constant from 0 to 0.7 unit and after that is decrased for different values of RE. When the number of actor nodes in an area is small our system selects the present the best actor node to perform the task. When there are many actors in the area, the present actor is not selected and the energy can be saved. In Fig. 8, we can see that the performance is lower than in the previous graphics beacuse of the increase of DE and DOA parameters. Furthermore in Fig. 9 we can see that the performance is the lowest because DE and DOA have maximum value and affect the system in a negative way. The DE defines the distance of the actor from the job place, so when DE is small, the ASD is higher. The actors closest to the job place use less energy to reach the job position. When RE is increased, the ASD is increased. However, when DOA is increased, the actor node is not selected for the required job.



(a) DOA=0.1

(b) DOA=0.5

(c) DOA=0.9

Fig. 8 Results for $DE = 0.5$.

5 Conclusions and Future Work

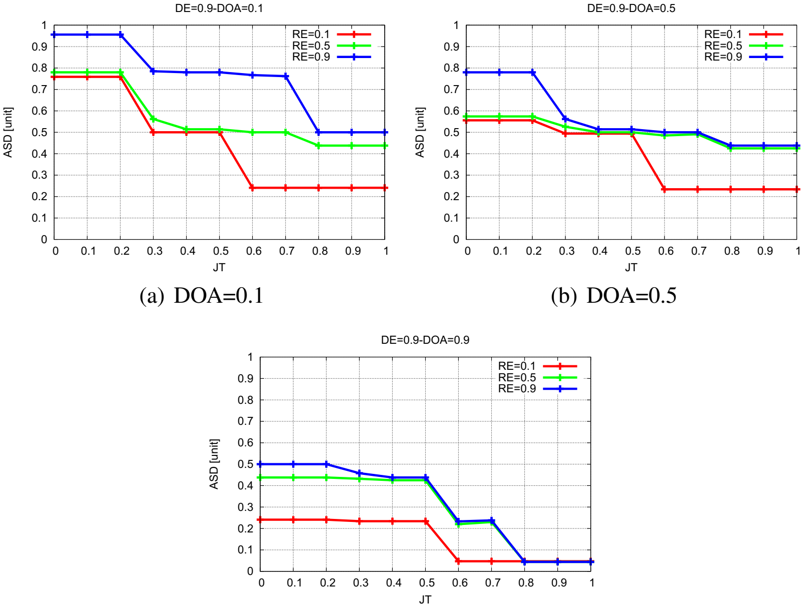
In this paper, we proposed and implemented a fuzzy-based simulation system for WSAN, which takes into account four input parameters, including DOA and decides the actor selection for a required task in the network.

The simulation results show that our system has a good performance.

In the future work, we will consider also other parameters for actor selection and make extensive simulations to evaluate the proposed system.

References

1. I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks (Elsevier)*, vol. 38, no. 4, pp. 393–422, 2002.
2. I. F. Akyildiz and I. H. Kasimoglu, "Wireless sensor and actor networks: Research challenges," *Ad Hoc Networks Journal (Elsevier)*, vol. 2, no. 4, pp. 351–367, October 2004.
3. N. Haider, M. Imran, N. Saad, and M. Zakariya, "Performance analysis of reactive connectivity restoration algorithms for wireless sensor and actor networks," in *Communications (MICC-2013), IEEE Malaysia International Conference on*, Nov 2013, pp. 490–495.
4. A. Abbasi, M. Younis, and K. Akkaya, "Movement-assisted connectivity restoration in wireless sensor and actor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 20, no. 9, pp. 1366–1379, Sept 2009.



(a) DOA=0.1

(b) DOA=0.5

(c) DOA=0.9

Fig. 9 Results for $DE = 0.9$.

5. X. Li, X. Liang, R. Lu, S. He, J. Chen, and X. Shen, "Toward reliable actor services in wireless sensor and actor networks," in *Mobile Adhoc and Sensor Systems (MASS), 2011 IEEE 8th International Conference on*, Oct 2011, pp. 351–360.
6. K. Akkaya and M. Younis, "Cola: A coverage and latency aware actor placement for wireless sensor and actor networks," in *Vehicular Technology (VTC-2006) Fall, IEEE 64th Conference on*, Sept 2006, pp. 1–5.
7. J. Kakarla and B. Majhi, "A new optimal delay and energy efficient coordination algorithm for wsan," in *Advanced Networks and Telecommuncations Systems (ANTS), 2013 IEEE International Conference on*, Dec 2013, pp. 1–6.
8. E. Kulla, M. Ikeda, and B. Leonard, "A fuzzy approach to actor selection in wireless sensor and actor networks," in *The 17-th International Conference on Network-Based Information Systems (NBiS-2014)*, Salerno, Italy, September 2014, pp. 244–248.
9. M. Akbas and D. Turgut, "Apawsan: Actor positioning for aerial wireless sensor and actor networks," in *Local Computer Networks (LCN), 2011 IEEE 36th Conference on*, Oct 2011, pp. 563–570.
10. M. Akbas, M. Brust, and D. Turgut, "Local positioning for environmental monitoring in wireless sensor and actor networks," in *Local Computer Networks (LCN), 2010 IEEE 35th Conference on*, Oct 2010, pp. 806–813.
11. T. Melodia, D. Pompili, V. Gungor, and I. AkyildizZX, "Communication and coordination in wireless sensor and actor networks," *IEEE Transactions on Mobile Computing*, vol. 6, no. 10, pp. 1126–1129, October 2007.
12. V. Gungor, O. Akan, and I. Akyildiz, "A real-time and reliable transport (rt2) protocol for wireless sensor and actor networks," *Networking, IEEE/ACM Transactions on*, vol. 16, no. 2, pp. 359–370, April 2008.

13. K. Selvaradjou, N. Handigol, A. Franklin, and C. Murthy, “Energy-efficient directional routing between partitioned actors in wireless sensor and actor networks,” *Communications, IET*, vol. 4, no. 1, pp. 102–115, January 2010.
14. H. Nakayama, Z. Fadlullah, N. Ansari, and N. Kato, “A novel scheme for wsan sink mobility based on clustering and set packing techniques,” *Automatic Control, IEEE Transactions on*, vol. 56, no. 10, pp. 2381–2389, Oct 2011.
15. T. Inaba, S. Sakamoto, V. Kolici, G. Mino, and L. Barolli, “A CAC Scheme Based on Fuzzy Logic for Cellular Networks Considering Security and Priority Parameters,” *The 9-th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA-2014)*, pp. 340–346, 2014.
16. E. Spaho, S. Sakamoto, L. Barolli, F. Xhafa, V. Barolli, and J. Iwashige, “A Fuzzy-Based System for Peer Reliability in JXTA-Overlay P2P Considering Number of Interactions,” *The 16th International Conference on Network-Based Information Systems (NBiS-2013)*, pp. 156–161, 2013.
17. K. Matsuo, D. Elmazi, Y. Liu, S. Sakamoto, G. Mino, and L. Barolli, “FACS-MP: A Fuzzy Admission Control System with Many Priorities for Wireless Cellular Networks and Its Performance Evaluation,” *Journal of High Speed Networks*, vol. 21, no. 1, pp. 1–14, 2015.
18. Y. Liu, S. Sakamoto, K. Matsuo, M. Ikeda, L. Barolli, and F. Xhafa, “Improving Reliability of JXTA-Overlay P2P Platform: A Comparison Study for Two Fuzzy-based Systems,” *Journal of High Speed Networks*, vol. 21, no. 1, pp. 27–45, 2015.
19. M. Grabisch, “The Application of Fuzzy Integrals in Multicriteria Decision Making,” *European journal of operational research*, vol. 89, no. 3, pp. 445–456, 1996.
20. T. Inaba, D. Elmazi, Y. Liu, S. Sakamoto, L. Barolli, and K. Uchida, “Integrating Wireless Cellular and Ad-Hoc Networks Using Fuzzy Logic Considering Node Mobility and Security,” *The 29th IEEE International Conference on Advanced Information Networking and Applications Workshops (WAINA-2015)*, pp. 54–60, 2015.
21. E. Kulla, G. Mino, S. Sakamoto, M. Ikeda, S. Caballé, and L. Barolli, “FBMIS: A Fuzzy-Based Multi-interface System for Cellular and Ad Hoc Networks,” *International Conference on Advanced Information Networking and Applications (AINA-2014)*, pp. 180–185, 2014.
22. D. Elmazi, E. Kulla, T. Oda, E. Spaho, S. Sakamoto, and L. Barolli, “A Comparison Study of Two Fuzzy-based Systems for Selection of Actor Node in Wireless Sensor Actor Networks,” *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–11, 2015.
23. L. Zadeh, “Fuzzy logic, neural networks, and soft computing,” *ACM Communications*, pp. 77–84, 1994.
24. E. Spaho, S. Sakamoto, L. Barolli, F. Xhafa, and M. Ikeda, “Trustworthiness in P2P: Performance Behaviour of Two Fuzzy-based Systems for JXTA-overlay Platform,” *Soft Computing*, vol. 18, no. 9, pp. 1783–1793, 2014.
25. T. Inaba, S. Sakamoto, E. Kulla, S. Caballe, M. Ikeda, and L. Barolli, “An Integrated System for Wireless Cellular and Ad-Hoc Networks Using Fuzzy Logic,” *International Conference on Intelligent Networking and Collaborative Systems (INCoS-2014)*, pp. 157–162, 2014.
26. K. Matsuo, D. Elmazi, Y. Liu, S. Sakamoto, and L. Barolli, “A Multi-modal Simulation System for Wireless Sensor Networks: A Comparison Study Considering Stationary and Mobile Sink and Event,” *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–11, 2015.
27. V. Kolici, T. Inaba, A. Lala, G. Mino, S. Sakamoto, and L. Barolli, “A Fuzzy-Based CAC Scheme for Cellular Networks Considering Security,” *International Conference on Network-Based Information Systems (NBiS-2014)*, pp. 368–373, 2014.
28. Y. Liu, S. Sakamoto, K. Matsuo, M. Ikeda, L. Barolli, and F. Xhafa, “A Comparison Study for Two Fuzzy-based Systems: Improving Reliability and Security of JXTA-overlay P2P Platform,” *Soft Computing*, pp. 1–11, 2015.
29. K. Matsuo, D. Elmazi, Y. Liu, S. Sakamoto, G. Mino, and L. Barolli, “FACS-MP: A Fuzzy Admission Control System with Many Priorities for Wireless Cellular Networks and Its Peforemance Evaluation,” *Journal of High Speed Network*, vol. 21, no. 1, pp. 1–14, 2015.
30. J. M. Mendel, “Fuzzy logic systems for engineering: a tutorial,” *Proc. of the IEEE*, vol. 83, no. 3, pp. 345–377, 1995.

A Fuzzy-based System for Qualified Voting in P2P Mobile Collaborative Team

Yi Liu, Tetsuya Oda, Keita Matsuo, Leonard Barolli and Fatos Xhafa

Abstract Mobile computing has many application domains. One important domain is that of mobile applications supporting collaborative work, such as, eLearning and eHealth. In such applications, a team of people collaborate online using smartphones to accomplish a common goal, such as a project development in e-Business. Often, however, the members of the team has to take decision or solve conflicts in project development (such as delays, changes in project schedule, task asignment, etc.) and therefore members have to vote. Voting can be done in many ways, and in most works in the literature consider majority voting, in which every member of the team accounts on for a vote. In this work, we consider a more realistic case where a vote does not account equal for every member, but accounts on according to member's active involvement and reliability in the groupwork. We present a voting model, that we call qualified voting, in which every member has a voting score according to three parameters. Then, we use fuzzy based model to compute a voting score for the member. This model is useful to implement in a P2P mobile collaborative team in replacement to majority voting as it gives more realistic view of the collaborative activity and better decisions for the groupwork, while encouraging peers to increase their reliability in order to increase their voting score.

Yi Liu
Graduate School of Engineering,
Fukuoka Institute of Technology (FIT),
3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811-0295, Japan,
e-mail: shinji.t.sakamoto@gmail.com

Tetsuya Oda, Keita Matsuo, Leonard Barolli
Department of Information and Communication Engineering,
Fukuoka Institute of Technology (FIT)
3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811-0295, Japan,
e-mail: oda.tetsuya.fit@gmail.com, makoto.ikd@acm.org, barolli@fit.ac.jp

Fatos Xhafa
Department of Languages and Informatics Systems,
Technical University of Catalonia
C/Jordi Girona 1-3, 08034 Barcelona, Spain,
e-mail: fatos@lsi.upc.edu

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_16

1 Introduction

P2P technologies has been among most disruptive technologies after Internet. Indeed, the emergence of the P2P technologies changed drastically the concepts, paradigms and protocols of sharing and communication in large scale distributed systems. As pointed out since early 2000 years [1], the nature of the sharing and the direct communication among peers in the system, being these machines or people, makes possible to overcome the limitations of the flat communications through email, newsgroups and other forum-based communication forms.

The usefulness of P2P technologies on one hand has been shown for the development of stand alone applications. On the other hand, P2P technologies, paradigms and protocols have penetrated other large scale distributed systems such as Mobile Adhoc Networks (MANETs), Groupware systems, Mobile Systems to achieve efficient sharing, communication, coordination, replication, awareness and synchronization. In fact, for every new form of Internet-based distributed systems, we are seeing how P2P concepts and paradigms again play an important role to enhance the efficiency and effectiveness of such systems or to enhance information sharing and online collaborative activities of groups of people. We briefly introduce below some common application scenarios that can benefit from P2P communications.

With the fast development in mobile technologies we are witnessing how the mobile devices are widely used for supporting collaborative team work. Indeed, by using mobile devices (such as PDAs, smartphones, etc.) members of a team can not only be geographically distributed, they can also be supported on the move, when network connection can change over time. In this paper, we propose a fuzzy-based system for qualified voting in P2P mobile collaborative team.

Fuzzy Logic (FL) is the logic underlying modes of reasoning which are approximate rather then exact. The importance of FL derives from the fact that most modes of human reasoning and especially common sense reasoning are approximate in nature. FL uses linguistic variables to describe the control parameters. By using relatively simple linguistic expressions it is possible to describe and grasp very complex problems. A very important property of the linguistic variables is the capability of describing imprecise parameters.

The concept of a fuzzy set deals with the representation of classes whose boundaries are not determined. It uses a characteristic function, taking values usually in the interval [0, 1]. The fuzzy sets are used for representing linguistic labels. This can be viewed as expressing an uncertainty about the clear-cut meaning of the label. But important point is that the valuation set is supposed to be common to the various linguistic labels that are involved in the given problem.

The fuzzy set theory uses the membership function to encode a preference among the possible interpretations of the corresponding label. A fuzzy set can be defined by examplification, ranking elements according to their typicality with respect to the concept underlying the fuzzy set [2].

The proposed fuzzy-based peer voting score system considers three parameters: Numbers of Activities the Member Participates (NAMP), Number of Activities the Member has Successfully Finished, Number of Online Discussions the Member has

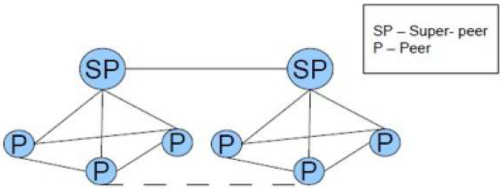


Fig. 1 Super-peer P2P group netwok.

Participated (NODMP) to decide the Voting Score (VS). We evaluated the proposed system by simulations. The simulation results show that with increasing of NAMP, NAMSF, and NODMP, the VS is increasing. Thus, the proposed system can choose reliable peers with good voting score in P2P mobile collaborative team.

The structure of this paper is as follows. In Section 2,we introduce the scenarios of collaborative teamwork. In Section 3, we introduce the vote weights and voting score. In Section 4, we introduce FL used for control. In Section 5, we present the proposed fuzzy-based system. In Section 6, we discuss the simulation results. Finally, conclusions and future work are given in Section 7.

2 Scenarios of Collaborative Teamwork

In this section, we describe and analyse some main scenarios of collaborative teamwork for which P2P technologies can support efficient system design.

2.1 Collaborative Teamwork and Virtual Campuses

Collaborative work through virtual teams is a significant way of collaborating in modern businesses, online learning, etc. Collaboration in virtual teams requires efficient sharing of information (both data sharing among the group members as well as sharing of group processes) and efficient communication among members of the team. Additionally, coordination and interaction are crucial for accomplishing common tasks through a shared workspace environment. P2P systems can enable fully decentralized collaborative systems by efficiently supporting different forms of collaboration [3]. One such form is using P2P networks, with super-peer structure as show in Fig. 1.

During the last two decades, online learning has become very popular and there is a widespread of virtual campuses or combinations of face-to-face with semi-open teaching and learning. Virtual campuses are now looking at ways to effectively support learners, especially for online courses implemented as PBL-Project Based Learning or SBL Scenario Based Learning there is an increasing need to develop mobile applications that support these online groupwork learning paradigms [4]. In such setting, P2P technologies offer interesting solutions for (a) decentralizing the virtual campuses, which tend to grow and get further centralized with the increase of number of students enrolled, new degrees, and increase in academic activity; (b) in taking advantage of resources of students and developing volunteerbased computing systems as part of virtual campuses and (c) alleviating the communication burden for efficient collaborative teamwork. The use of P2P libraries such as JXTA have been investigated to design P2P middleware for P2P eLearning applications . Also, the use of P2P technologies in such setting is used for P2P video synchronization in a collaborative virtual environment [5]. Recently, virtual campuses are also introducing social networking among their students to enhance the learning activities through social support and scaffolding. Again the P2P solutions are sought in this context [6] in combination with social networking features to enhance especially the interaction among learners sharing similar objectives and interest or accomplishing a common project.

2.2 Mobile Ad hoc Networks (MANETs)

Mobile ad-hoc networks are among most interesting infrastructureless network of mobile devices connected by wireless having self-configuring properties. The lack of fixed infrastructure and of a centralized administration makes the building and operation in MANETS challenging. P2P networks and mobile ad hoc networks (MANETs) follow the same idea of creating a network without a central entity. All nodes (peers) must collaborate together to make possible the proper functioning of the network by forwarding information on behalf of others in the network [7]. P2P and MANETs share many key characteristics such as self-organization and decentralization due to the common nature of their distributed components. Both MANETs and P2P networks follow a P2P paradigm characterized by the lack of a central node or peer acting as a managing server, all participants having therefore to collaborate in order for the whole system to work. A key issue in both networks is the process of discovering the requested data or route efficiently in a decentralized manner. Recently, new P2P applications which uses wireless communication and integrates mobile devices such as PDA and mobile phones is emerging. Several P2P-based protocols can be used for MANETs such as Mobile P2P Protocol (MPP), which is based on Dynamic Source Routing (DSR), JXTA prtotocols, and MANET Anonymous Peer-to-peer Communication Protocol (MAPCP), which serves as an efficient anonymous communication protocol for P2P applications over MANET.

3 Vote Weights

3.1 Votes with Embedded Weight

The weights can be included in voting bulletins distributed to voters, which would then be copied into the votes sent to Counters. But this approach requires a strong assumption: the voters' application must be trusted not to forge weights. Since the voters' application may be tampered in some scenarios, namely when "voting anywhere" is considered, the voters' side cannot be trusted to give the correct input for the system when weights are considered.

The simple copy/paste of weights could be strengthened by adding a cleartext value of the weight when submitting a blinded vote digest for getting a signature from an Administrator. Then, the weight, checked and signed by all the required Administrators, could be added to the final vote submitted to Counters. A bit commitment value should also be added to the weight to prevent stolen, signed weights, to be used by other voters. The drawback of this approach is that protocol messages from voters to Administrators and from voters to Counters would increase in size, namely would double in size. This collides with the requirement of keeping the performance of system close to the performance of the initial version of REVS(Robust Electronic Voting System [8]).

3.2 Voting Score

Score voting (sometimes called range voting) is a single-winner voting system where voters rate candidates on a scale. The candidate with the highest rating wins. For comparison, consider ratings systems from site like: Internet Movie Database, Amazon, Yelp, and Hot or Not. Variations of score voting can use a score-style ballot to elect multiple candidates simultaneously.

Simplified forms of score voting automatically give skipped candidates the lowest possible score for the ballot they were skipped. Other forms have those ballots not affect the candidate's rating at all. Those forms not affecting the candidates rating frequently make use of quotas. Quotas demand a minimum proportion of voters rate that candidate in some way before that candidate is eligible to win [9].

4 Application of Fuzzy Logic for Control

The ability of fuzzy sets and possibility theory to model gradual properties or soft constraints whose satisfaction is matter of degree, as well as information pervaded with imprecision and uncertainty, makes them useful in a great variety of applications.

The most popular area of application is Fuzzy Control (FC), since the appearance, especially in Japan, of industrial applications in domestic appliances, process control, and automotive systems, among many other fields.

4.1 FC

In the FC systems, expert knowledge is encoded in the form of fuzzy rules, which describe recommended actions for different classes of situations represented by fuzzy sets.

In fact, any kind of control law can be modeled by the FC methodology, provided that this law is expressible in terms of "if ... then ..." rules, just like in the case of expert systems. However, FL diverges from the standard expert system approach by providing an interpolation mechanism from several rules. In the contents of complex processes, it may turn out to be more practical to get knowledge from an expert operator than to calculate an optimal control, due to modeling costs or because a model is out of reach.

4.2 Linguistic Variables

A concept that plays a central role in the application of FL is that of a linguistic variable. The linguistic variables may be viewed as a form of data compression. One linguistic variable may represent many numerical variables. It is suggestive to refer to this form of data compression as granulation [10].

The same effect can be achieved by conventional quantization, but in the case of quantization, the values are intervals, whereas in the case of granulation the values are overlapping fuzzy sets. The advantages of granulation over quantization are as follows:

- it is more general;
- it mimics the way in which humans interpret linguistic values;
- the transition from one linguistic value to a contiguous linguistic value is gradual rather than abrupt, resulting in continuity and robustness.

4.3 FC Rules

FC describes the algorithm for process control as a fuzzy relation between information about the conditions of the process to be controlled, x and y, and the output for the process z. The control algorithm is given in "if ... then ..." expression, such as:

If x is small and y is big, then z is medium;
If x is big and y is medium, then z is big.

These rules are called *FC rules*. The "if" clause of the rules is called the antecedent and the "then" clause is called consequent. In general, variables x and y are called the input and z the output. The "small" and "big" are fuzzy values for x and y, and they are expressed by fuzzy sets.

Fuzzy controllers are constructed of groups of these FC rules, and when an actual input is given, the output is calculated by means of fuzzy inference.

4.4 Control Knowledge Base

There are two main tasks in designing the control knowledge base. First, a set of linguistic variables must be selected which describe the values of the main control parameters of the process. Both the input and output parameters must be linguistically defined in this stage using proper term sets. The selection of the level of granularity of a term set for an input variable or an output variable plays an important role in the smoothness of control. Second, a control knowledge base must be developed which uses the above linguistic description of the input and output parameters. Four methods [11–14] have been suggested for doing this:

- expert's experience and knowledge;
- modelling the operator's control action;
- modelling a process;
- self organization.

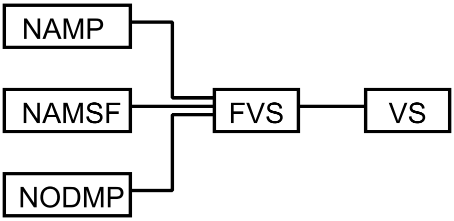
Among the above methods, the first one is the most widely used. In the modeling of the human expert operator's knowledge, fuzzy rules of the form "If Error is small and Change-in-error is small then the Force is small" have been used in several studies [15, 16]. This method is effective when expert human operators can express the heuristics or the knowledge that they use in controlling a process in terms of rules of the above form.

4.5 Defuzzification Methods

The defuzzification operation produces a non-FC action that best represent the membership function of an inferred FC action. Several defuzzification methods have been suggested in literature. Among them, four methods which have been applied most often are:

- Tsukamoto's Defuzzification Method;
- The Center of Area (COA) Method;
- The Mean of Maximum (MOM) Method;

Fig. 2 Stucture of System



- Defuzzification when Output of Rules are Function of Their Inputs.

5 Proposed Fuzzy-based Peer Voting Score System

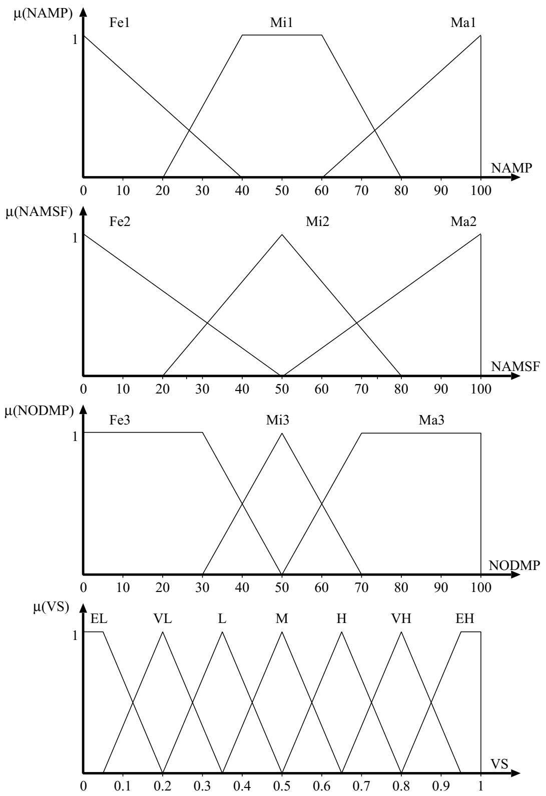
In this work, we consider there parameters: Numbers of Activities the Member Participates (NAMP), Number of Activities the Member has Successfully Finished, Number of Online Discussions the Member has Participated (NODMP) to decide the Voting Score (VS). The structure of this system called Fuzzy-based Vote System (FVS) is shown in Fig. 2. These three parameters are fuzzified using fuzzy system, and based on the decision of fuzzy system a voting score is calculated. The membership functions for our system are shown in Fig. 3. In Table 1, we show the Fuzzy Rule Base (FRB) of our proposed system, which consists of 27 rules.

The input parameters for FVS are: NAMP, NAMSF, NODMP and the output linguistic parameter is VS. The term sets of *NAMP*, *NAMSF* and *NODMP* are defined respectively as:

$$\begin{aligned}
NAMP &= \{Few1, Middle1, Many1\} \\
&= \{Fe1, Mi1, Ma1\}; \\
NAMSF &= \{Few2, Middle2, Many2\} \\
&= \{Fe2, Mi2, Ma2\}; \\
NODMP &= \{Few3, Middle3, Many3\} \\
&= \{Fe3, Mi3, Ma3\}.
\end{aligned}$$

and the term set for the output *VS* is defined as:

Fig. 3 Membership functions.



$$VS = \begin{pmatrix} Extremely\ Low \\ Very\ Low \\ Low \\ Middle \\ High \\ Very\ High \\ Extremely\ High \end{pmatrix} = \begin{pmatrix} EL \\ VL \\ L \\ M \\ H \\ VH \\ EH \end{pmatrix}$$

6 Simulation Results

In this section, we present the simulation results for our proposed system. In our system, we decided the number of term sets by carrying out many simulations. These simulation results were carried out in MATLAB.

From Fig. 4 to Fig. 6, we show the relation between NAMP, NAMSF, NODMP and VS. In this simulation, we consider the NODMP as a constant parameter.

In Fig. 4, we consider the NODMP value 0 unit. When the NAMP increases, the VS is increased. Also, when the NAMSF increases, the VS is increased.

Table 1 FRB.

Rule	NAMP	NAMF	NODMP	VS
1	Fe1	Fe2	Fe3	EL
2	Fe1	Fe2	Mi3	EL
3	Fe1	Fe2	Ma3	L
4	Fe1	Mi2	Fe3	EL
5	Fe1	Mi2	Mi3	VL
6	Fe1	Mi2	Ma3	M
7	Fe1	Ma2	Fe3	VL
8	Fe1	Ma2	Mi3	L
9	Fe1	Ma2	Ma3	H
10	Mi1	Fe2	Fe3	EL
11	Mi1	Fe2	Mi3	L
12	Mi1	Fe2	Ma3	M
13	Mi1	Mi2	Fe3	VL
14	Mi1	Mi2	Mi3	M
15	Mi1	Mi2	Ma3	H
16	Mi1	Ma2	Fe3	L
17	Mi1	Ma2	Mi3	H
18	Mi1	Ma2	Ma3	VH
19	Ma1	Fe2	Fe3	VL
20	Ma1	Fe2	Mi3	M
21	Ma1	Fe2	Ma3	VH
22	Ma1	Mi2	Fe3	L
23	Ma1	Mi2	Mi3	H
24	Ma1	Mi2	Ma3	VH
25	Ma1	Ma2	Fe3	M
26	Ma1	Ma2	Mi3	VH
27	Ma1	Ma2	Ma3	VVH

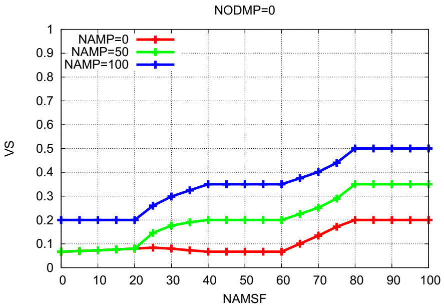


Fig. 4 Voting Score for NODMP=0.

In Fig. 5 and Fig. 6, we increase the NODMP values to 50 and 100 units, respectively. We see that, when the NODMP increases, the VS is increased.

Fig. 5 Voting score for NODMP=50.

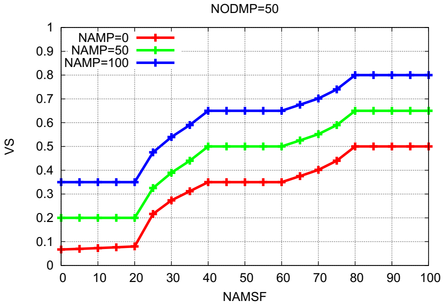
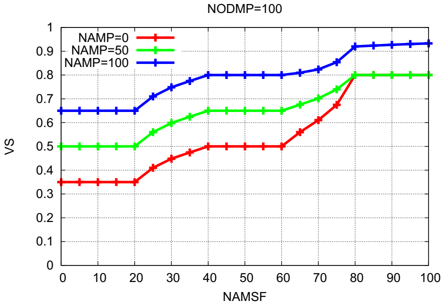


Fig. 6 Voting Score for NODMP=100.



7 Conclusions and Future Work

In this paper, we proposed a fuzzy-based system to decide the VS. We took into consideration three parameters: NAMP, NAMSF, and NODMP. We evaluated the performance of proposed system by computer simulations. From the simulations results, we conclude that with increasing of NAMP, NAMSF, and NODMP, the VS is increasing. Thus, the proposed system can choose reliable peers with good voting score in P2P mobile collaborative team.

In the future, we would like to make extensive simulations to evaluate the proposed system and compare the performance of our proposed system with other systems.

References

1. Oram, A. (Ed.). Peer-to-Peer: Harnessing the power of disruptive technologies. CA: O'Reilly and Associates, 2001.
2. T. Terano, K. Asai, and M. Sugeno, "Fuzzy Systems Theory And Its Applications", Academic Press, INC. Harcourt Brace Jovanovich, Publishers, 1992.
3. F. Xhafa, and A. Poulovassilis, "Requirements for Distributed Event-Based Awareness in P2P Groupware Systems", Proc. of AINA 2010 , pp. 220-225, April 2010.
4. F.Xhafa, L.Barolli, S.Caball e, and R.Fernandez," Supporting Scenario-Based Online Learning with P2P Group-Based Systems", Proc of NBiS 2010 , pp.173-180, September 2010
5. S.Gupta and G. Kaiser, "P2P video synchronization in a collaborative virtual environment", In Proceedings of the 4th international conference on Advances in Web-Based Learning (ICWL' 05), pp.86-98, 2005.

6. Ana M. Martnez-Alemn, A. M. and Wartman, K.L, “Online Social Networking on Campus Understanding What Matters in Student Culture”, Taylor and Francis, Routledge 2008.
7. E. Spaho, E. Kulla, F. Xhafa and L. Barolli, “P2P Solutions to Efficient Mobile Peer Collaboration in MANETs”, Proc. of 3PGCIC 2012 , pp.379-383, November 2012.
8. Rui Joaquim, Andr e Z uquete and Paulo Ferreira. “REVS A Robust Electronic Voting System”, IADIS Int. Journal of WWW/Internet, 1(2), December 2003.
9. https://electology.org/score-voting.
10. A. Kandel, “Fuzzy Expert Systems”, CRC Press, 1992.
11. H. J. Zimmermann, “Fuzzy Set Theory and Its Applications”, Kluwer Academic Publishers, Second Revised Edition, 1991.
12. F. M. McNeill, and E. Thro, “Fuzzy Logic. A Practical Approach”, Academic Press, Inc., 1994.
13. L. A. Zadeh, J. Kacprzyk, “Fuzzy Logic For The Management of Uncertainty”, John Wiley & Sons, Inc., 1992.
14. T. J. Procyk and E. H. Mamdani, “A Linguistic Self-organizing Process Controller”, Automatica, Vol. 15, No. 1, pp. 15-30, 1979.
15. G. J. Klir, and T. A. Folger, “Fuzzy Sets, Uncertainty, And Information”, Prentice Hall, Englewood Cliffs, 1988.
16. T. Munakata, and Y. Jani, “Fuzzy Systems: An Overview”, Commun. of ACM, Vol. 37, No. 3, pp. 69-76, March 1994.

An Integrated Simulation System Considering WMN-PSO Simulation System and Network Simulator 3

Shinji Sakamoto, Tetsuya Oda, Makoto Ikeda, Leonard Barolli and Fatos Xhafa

Abstract With the fast development of wireless technologies, Wireless Mesh Networks (WMNs) are becoming an important networking infrastructure due to their low cost and increased high speed wireless Internet connectivity. In our previous work, we implemented a simulation system based on Particle Swam Optimization for solving node placement problem in wireless mesh networks, called WMN-PSO. In this paper, we implement an integrated system considering WMN-PSO and network simulator 3 (ns-3). For simulation, we consider two WMN architectures. From simulation results, we found that the total received throughput of I/B WMN is higher than Hybrid WMN and the delay of I/B WMN is lower than Hybrid WMN.

1 Introduction

The wireless networks and devises are becoming increasingly popular and they provide users access to information and communication anytime and anywhere [1, 4,

Shinji Sakamoto
Graduate School of Engineering,
Fukuoka Institute of Technology (FIT),
3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811-0295, Japan,
e-mail: shinji.t.sakamoto@gmail.com

Tetsuya Oda, Makoto Ikeda, Leonard Barolli
Department of Information and Communication Engineering,
Fukuoka Institute of Technology (FIT)
3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811-0295, Japan,
e-mail: oda.tetsuya.fit@gmail.com, makoto.ikd@acm.org, barolli@fit.ac.jp

Fatos Xhafa
Department of Languages and Informatics Systems,
Technical University of Catalonia
C/Jordi Girona 1-3, 08034 Barcelona, Spain,
e-mail: fatos@lsi.upc.edu

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_17

8, 9, 10, 12, 13, 14, 15, 17, 35].Wireless Mesh Networks (WMNs) are gaining a lot of attention because of their low cost nature that makes them attractive for providing wireless Internet connectivity. A WMN is dynamically self-organized and self-configured, with the nodes in the network automatically establishing and maintaining mesh connectivity among them-selves (creating, in effect, an ad hoc network). This feature brings many advantages to WMNs such as low up-front cost, easy network maintenance, robustness and reliable service coverage [2]. Moreover, such infrastructure can be used to deploy community networks, metropolitan area networks, municipal and corporative networks, and to support applications for urban areas, medical, transport and surveillance systems.

Mesh node placement in WMN can be seen as a family of problems, which are shown (through graph theoretic approaches or placement problems, e.g. [6, 20]) to be computationally hard to solve for most of the formulations [32]. In fact, the node placement problem considered here is even more challenging due to two additional characteristics: (a) locations of mesh router nodes are not pre-determined (any available position in the considered area can be used for deploying the mesh routers) and (b) routers are assumed to have their own radio coverage area. Here, we consider the version of the mesh router nodes placement problem in which we are given a grid area where to deploy a number of mesh router nodes and a number of mesh client nodes of fixed positions (of an arbitrary distribution) in the grid area. The objective is to find a location assignment for the mesh routers to the cells of the grid area that maximizes the network connectivity and client coverage.

Node placement problems are known to be computationally hard to solve [18, 19, 33]. In some previous works, intelligent algorithms have been recently investigated [3, 7, 11, 21, 22, 24, 25, 26, 27].

In our previous work, we implemented a simulation system based on Particle Swam Optimization for solving node placement problem in wireless mesh networks, called WMN-PSO. In this paper, we implement an integrated system considering WMN-PSO and network simulator 3 (ns-3). For WMN-PSO, the metrics used for optimization are the Size of Giant Component (SGC) and the Number of Covered Mesh Clients (NCMC).

The rest of the paper is organized as follows. The Architectures of WMNs are described in Section 2. The mesh router nodes placement problem is defined in Section 3. We present our proposed and implemented WMN-PSO simulation system in Section 4. The ns-3 is explained in Section 5. The simulation results are given in Section 6. Finally, we give conclusions and future work in Section 7.

2 Architectures of WMNs

In this Section, we describe the architectures of WMNs. Node architectures for WMNs can be classified according to the functionalities they offer as follows:

Infrastructure/Backbone (I/B) WMNs:

This type of architecture is the most used and consists of a grid of mesh routers which are connected to different clients. In addition, routers have gateway functionality thus allowing Internet access to clients. This architecture enables integration with other existing wireless networks and is widely used in neighboring communities.

Client WMNs:

This type of architecture provides a peer to peer based communication network over clients devices. In this architecture, there is no definition for mesh router. In this case, we have a network of mesh nodes which provide routing functionality and configuration as well as end-user applications, so that when a packet is sent from one node to another, the packet will be hopped from node to node in the mesh nodes in order to reach destination.

Hybrid WMNs:

This architecture combines two previous architectures. Mesh clients are able to access the network through mesh routers as well as through direct connection with other mesh clients. Benefiting from the advantages of the two architectures, Hybrid WMNs can connect to other networks (Internet, Wi-Fi, and sensor networks) and enhance the connectivity and coverage due to the fact that the mesh clients can act as mesh routers.

3 Node Placement Problem in WMNs

For this problem, we have a grid area arranged in cells we want to find where to distribute a number of mesh router nodes and a number of mesh client nodes of fixed positions (of an arbitrary distribution) in the grid area. The objective is to find a location assignment for the mesh routers to the area that maximizes the network connectivity and client coverage. Network connectivity is measured by SGC of the resulting WMN graph, while the user coverage is simply the number of mesh client nodes that fall within the radio coverage of at least one mesh router node and is measured by NCMC.

An instance of the problem consists as follows.

- N mesh router nodes, each having its own radio coverage, defining thus a vector of routers.
- An area $W \times H$ where to distribute N mesh routers. Positions of mesh routers are not pre-determined and are to be computed.
- M client mesh nodes located in arbitrary points of the considered area, defining a matrix of clients.

It should be noted that network connectivity and user coverage are among most important metrics in WMNs and directly affect the network performance.

In this work, we have considered a bi-objective optimization in which we first maximize the network connectivity of the WMN (through the maximization of the SGC) and then, the maximization of the NCMC.

In fact, we can formalize an instance of the problem by constructing an adjacency matrix of the WMN graph, whose nodes are router nodes and client nodes and whose edges are links between nodes in the mesh network. Each mesh node in the graph is a triple $v =< x, y, r >$ representing the 2D location point and r is the radius of the transmission range. There is an arc between two nodes u and v, if v is within the transmission circular area of u.

4 Proposed WMN-PSO System

4.1 PSO

In PSO a number of simple entities (the particles) are placed in the search space of some problem or function and each evaluates the objective function at its current location. The objective function is often minimized and the exploration of the search space is not through evolution [23]. However, following a widespread practice of borrowing from the evolutionary computation field, in this work, we consider the bi-objective function and fitness function interchangeably. Each particle then determines its movement through the search space by combining some aspect of the history of its own current and best (best-fitness) locations with those of one or more members of the swarm, with some random perturbations. The next iteration takes place after all particles have been moved. Eventually the swarm as a whole, like a flock of birds collectively foraging for food, is likely to move close to an optimum of the fitness function.

Each individual in the particle swarm is composed of three $\mathscr{D}$-dimensional vectors, where $\mathscr{D}$ is the dimensionality of the search space. These are the current position $\mathbf{x}_i$, the previous best position $\mathbf{p}_i$ and the velocity $\mathbf{v}_i$.

The particle swarm is more than just a collection of particles. A particle by itself has almost no power to solve any problem; progress occurs only when the particles interact. Problem solving is a population-wide phenomenon, emerging from the individual behaviors of the particles through their interactions. In any case, populations are organized according to some sort of communication structure or topology, often thought of as a social network. The topology typically consists of bidirectional edges connecting pairs of particles, so that if j is in i's neighborhood, i is also in j's. Each particle communicates with some other particles and is affected by the best point found by any member of its topological neighborhood. This is just the vector $\mathbf{p}_i$ for that best neighbor, which we will denote with $\mathbf{p}_g$. The potential kinds of population "social networks" are hugely varied, but in practice certain types have been used more frequently.

In the PSO process, the velocity of each particle is iteratively adjusted so that the particle stochastically oscillates around $\mathbf{p}_i$ and $\mathbf{p}_g$ locations.

Algorithm 1 Pseudo code of PSO.

```
/* Generate the initial solutions and parameters */
Computation maxtime:= $T_{max}$, $t = 0$;
Number of particle-patterns:= $m$, $2 \le m \in R^1$;
Particle-patterns initial solution:= $P_i^0$;
Global initial solution:= $G^0$;
Particle-patterns initial position:= $x_{ij}^0$;
Particles initial velocity:= $v_{ij}^0$;
PSO parameter:= $\omega$, $0 < \omega \in R^1$;
PSO parameter:= $C_1$, $0 < C_1 \in R^1$;
PSO parameter:= $C_2$, $0 < C_2 \in R^1$;
/* Start PSO */
Evaluate($G^0, P^0$);
/* "Evaluate" does calculate present fitness value of each Particle-patterns. */
while $t < T_{max}$ do
    /* Update velocities and positions */
    $v_{ij}^{t+1} = \omega \cdot v_{ij}^t$
        $+C_1 \cdot \mathrm{rand}() \cdot (best(P_{ij}^t) - x_{ij}^t)$
        $+C_2 \cdot \mathrm{rand}() \cdot (best(G^t) - x_{ij}^t)$;
    $x_{ij}^{t+1} = x_{ij}^t + v_{ij}^{t+1}$;
    Update_Solutions($G^t, P^t$);
    /* "Update_Solutions" compares and updates the Particle-pattern's best solutions and the
    global best solutions if their fitness value is better than previous. */
    Evaluate($G^{(t+1)}, P^{(t+1)}$);
    $t = t + 1$;
end while
Update_Solutions($G^t, P^t$);
return Best found pattern of particles as solution;
```

4.2 WMN-PSO System for Mesh Router Node Placement

We propose and implement a new simulator that uses PSO algorithm to solve the node placement problem in WMNs. We call this simulator WMN-PSO. Our system can generate instances of the problem using different iterations of clients and mesh routers.

We present here the particularization of the PSO algorithm (see Algorithm 1) for the mesh router node placement problem in WMNs.

Initialization

Our proposed system starts by generating an initial solution randomly, by *ad hoc* methods [34]. We decide the velocity of particles by a random process considering the area size. For instance, when the area size is $W \times H$, the velocity is decided randomly from $-\sqrt{W^2 + H^2}$ to $\sqrt{W^2 + H^2}$.

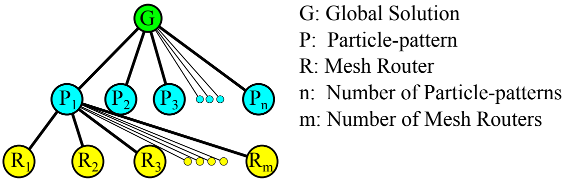


Fig. 1 Relationship among global solution, particle-patterns and mesh routers.

Particle-pattern

A particle is a mesh router. A fitness value of a particle-pattern is computed by combination of mesh routers and mesh clients positions. In other words, each particle-pattern is a solution as shown is Fig. 1. Therefore, the number of particle-patterns is a number of solutions.

Fitness function

One of most important thing in PSO algorithm is to decide the determination of an appropriate objective function and its encoding. In our case, each particle-pattern has an own fitness value and compares other particle-pattern's fitness value in order to share information of global solution. The fitness function follows a hierarchical approach in which the main objective is to maximize the SGC in WMN. The fitness function of this scenario is considered as Where α and β are weight-coefficient of SGC and NCMC, respectively.

Routers replacement method

A mesh router has x, y positions and velocity. Mesh routers are moved based on velocities. There are many moving methods in PSO field, such as:

Constriction Method (CM)
: CM is a method which PSO parameters are set to a week stable region ($\omega = 0.729$, $C_1 = C2 = 1.4955$) based on analysis of PSO by M. Clerc et. al. [5, 30].

Random Inertia Weight Method (RIWM)
: In RIWM, the ω parameter is changing ramdomly from 0.5 to 1.0. The C_1 and C_2 are kept 2.0. The ω can be estimated by the week stable region. The average of ω is 0.75 [30].

Linearly Decreasing Inertia Weight Method (LDIWM)
: In LDIWM, C_1 and C_2 are set to 2.0, constantly. On the other hand, the ω parameter is changed linearly from unstable region ($\omega = 0.9$) to stable region ($\omega = 0.4$) with increasing of iterations of computations [30, 31].

Linearly Decreasing Vmax Method (LDVM)
: In LDVM, PSO parameters are set to unstable region ($\omega = 0.9$, $C_1 = C_2 = 2.0$). A value of V_{max} which is maximum velocity of particles is considered. With

Table 1 Simulation parameters for WMN-PSO.

Parameters	Values
Clients distribution	Normal distribution
Grid size	32×32
Number of mesh routers	16
Number of mesh clients	48
Total iterations	12800
Iteration per phase	64
Number of particle-patterns	20
Radius of a mesh router	2.0
Independent runs	10
Movement method	RDVM

Table 2 Simulation parameters for ns-3.

Parameters	Values
Clients distribution	Normal distribution
Area size	640m × 640m
Number of mesh routers	16
Number of mesh clients	48
MAC	IEEE 802.11s
Propagation loss model	Log distance path-loss model
Propagation delay model	Constant speed model
Routing protocol	HWMP
Transport protocol	UDP
Application type	CBR
Packet size	1024 [bytes]
Number of source nodes	10
Number of destination nodes	1
Simulation time	600 [sec]

increasing of iteration of computations, the V_{max} is kept decreasing linearly [29]. In this work, we apply this method to optimize the weight-coefficients of SGC and NCMC.

Rational Decrement of Vmax Method (RDVM)

In RDVM, PSO parameters are set to unstable region ($\omega = 0.9$, $C_1 = C_2 = 2.0$). The V_{max} is kept decreasing with the increasing of iterations as

$$V_{max}(x) = \sqrt{W^2 + H^2} \times \frac{T - x}{x}. \quad (1)$$

Where, W and H are the width and the height of the considered area, respectively. Also, T and x are the total number of iterations and a current number of iteration, respectively [28].

5 ns-3

The network simulator 3 (ns-3) is a free software written in C++ programming language. The ns-3 architecture is similar to Linux computers, with internal interface and application interfaces such as, network interfaces, device drivers and sockets. The goals of ns-3 are set very high: to create a new network simulator aligned with modern research needs and develop it in an open source community. Users of ns-3 are free to write their simulation scenarios by C++ or Python. The ns-3's low-level API is oriented towards the power-user but more accessible "helper" APIs are overlaid on top of the low-level API.

In order to archive scalability of a very large number of simulated network elements, the ns-3 tools also support standardized output formats for trace-data, such as the pcap format used by network packet analyzing tools such as wireshark, tcpdump, and standardized input format such as importing mobility trace file from ns-3.

The ns-3 simulator has models for all network elements that comprise a computer network. For example, network devices represent the physical device that connects a node to the communication channel. This might be a simple Ethernet network interface card or a more complex wireless IEEE 802.11 device.

The ns-3 is intended as an eventual replacement for popular ns-2. The ns-3's wifi models a wireless network interface controller based on IEE802.11 standard [16]. The ns-3 provides models for these aspects of IEEE 802.11:

- Basic IEEE 802.11 DCF with infrastructure and ad-hoc modes.
- IEEE 802.11a, IEEE 802.11b, IEEE 802.11g, and IEEE 802.11s physical layers.
- QoS based EDCA and queuing extensions of IEEE 802.11e.
- Various propagation loss models including Nakagami, Rayleigh, Friis, LogDistance, FixedRSS, and so on.
- Two propagation delay models, distance-based and Random-model.
- Various rate control algorithms including Aarf, Arf, Cara, Onoe, Rraa, ConstantRate, and Minstrel.

6 Simulation

In this section, we show simulation results using WMN-PSO simulation system and ns-3. In WMN-PSO, the area size is considered 32 × 32. The simulation parameters for WMN-PSO are shown in Table 1. We conducted simulations 10 times, in order to avoid the effect randomness and create a general view of results. In Table 2, we show parameters for ns-3.

We show the simulation results from Fig. 2 to Fig. 3. In Fig. 2, we show the optimized placement of mesh routers by using WMN-PSO system. Then, we evaluate the optimized placement by using ns-3. In Fig. 3, we show the total received throughput. For I/B WMN, the total received throughput is higher than Hybrid WMN. In Fig. 4, we show the delay for both I/B WMN and Hybrid WMN ar-

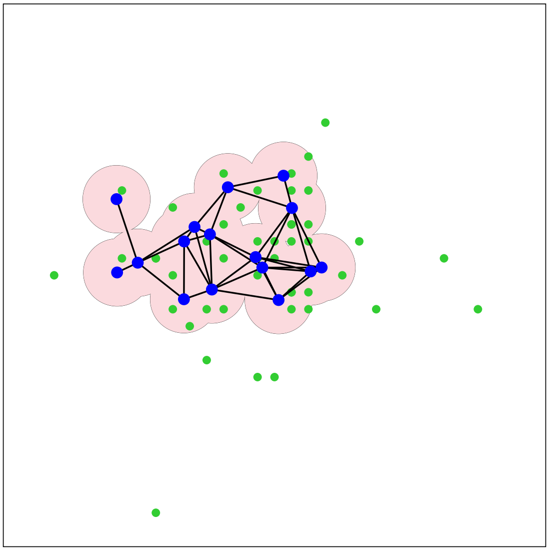


Fig. 2 Placement of mesh routers by WMN-PSO.

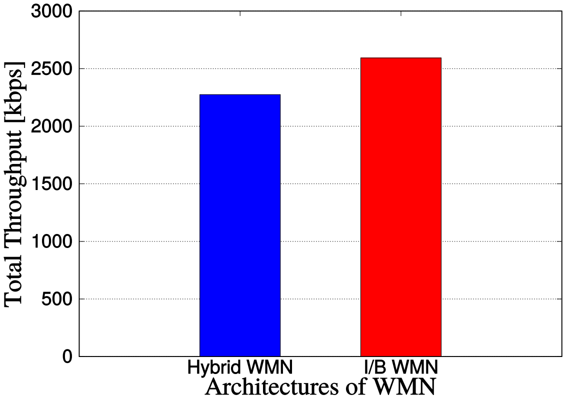


Fig. 3 Simulation results: I/B WMN v.s. Hybrid WMN for total receive throughput.

chitectures. When the architecture of WMN is I/B, the delay is lower than Hybrid WMN.

7 Conclusions

In this paper, we implemented an integrated system considering WMN-PSO and network simulator 3 (ns-3). For simulation, we considered two WMN architectures. From simulation results, we found that the total received throughput of I/B WMN is higher than Hybrid WMN and the delay of I/B WMN is lower than Hybrid WMN.

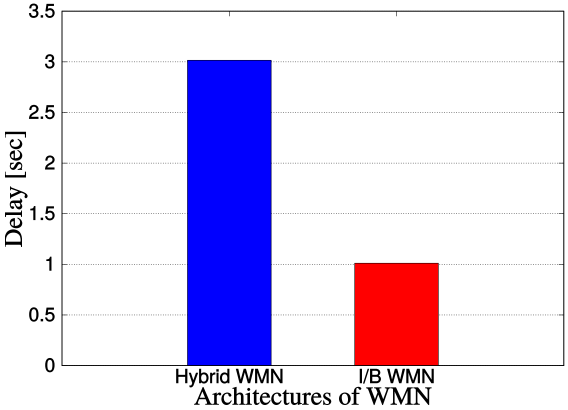


Fig. 4 Simulation results: I/B WMN v.s. Hybrid WMN for delay.

In our future work, we would like to evaluate the performance of the proposed system for different parameters and patterns. Moreover, we would like to compare its performance with other algorithms.

Acknowledgement

This work is supported by a Grant-in-Aid for Scientific Research from Japanese Society for the Promotion of Science (JSPS KAKENHI Grant Number 15J12086). The authors would like to thank JSPS for the financial support.

References

[1] Aikebaier A, Enokido T, Takizawa M (2011) TMPR-scheme for Reliably Broadcast Messages Among Peer Processes. International Journal of Grid and Utility Computing 2(3):175–182

[2] Akyildiz IF, Wang X, Wang W (2005) Wireless Mesh Networks: A Survey. Computer Networks 47(4):445–487

[3] Amaldi E, Capone A, Cesana M, Filippini I, Malucelli F (2008) Optimization Models and Methods for Planning Wireless Mesh Networks. Computer Networks 52(11):2159–2171

[4] Boyinbode O, Le H, Takizawa M (2011) A Survey on Clustering Algorithms for Wireless Sensor Networks. International Journal of Space-Based and Situated Computing 1(2):130–136

[5] Clerc M, Kennedy J (2002) The Particle Swarm-Explosion, Stability, and Convergence in a Multidimensional Complex Space. IEEE Transactions on Evolutionary Computation 6(1):58–73

[6] Franklin AA, Murthy CSR (2007) Node Placement Algorithm for Deployment of Two-tier Wireless Mesh Networks. Proc of Global Telecommunications Conference pp 4823–4827

[7] Girgis MR, Mahmoud TM, Abdullatif BA, Rabie AM (2014) Solving the Wireless Mesh Network Design Problem using Genetic Algorithm and Simulated Annealing Optimization Methods. International Journal of Computer Applications 96(11):1–10

[8] Goto K, Sasaki Y, Hara T, Nishio S (2013) Data Gathering using Mobile Agents for Reducing Traffic in Dense Mobile Wireless Sensor Networks. Mobile Information Systems 9(4):295–314

[9] Hiyama M, Kulla E, Ikeda M, Barolli L (2012) Evaluation of MANET Protocols for Different Indoor Environments: Results from a Real MANET Testbed. International Journal of Space-Based and Situated Computing 2(2):71–82

[10] Hiyama M, Sakamoto S, Kulla E, Ikeda M, Barolli L (2013) Experimental Results of a MANET Testbed for Different Settings of HELLO Packets of OLSR Protocol. Journal of Mobile Multimedia 9(1-2):27–38

[11] Hoshi T, Kumata Y, Koyama A (2013) A Proposal and Evaluation of Access Point Allocation Algorithm for Wireless Mesh Networks. International Conference on Network-Based Information Systems (NBiS-2013) pp 389–394

[12] Ikeda M (2012) Analysis of Mobile Ad-hoc Network Routing Protocols using Shadowing Propagation Model. International Journal of Space-Based and Situated Computing 2(3):139–148

[13] Ikeda M (2012) End-to-End Single and Multiple Flows Fairness in Mobile Ad-hoc Networks. Journal of Mobile Multimedia 8(3):204–224

[14] Ikeda M, Honda T, Barolli L (2015) Performance of Optimized Link State Routing Protocol for Video Streaming Application in Vehicular Ad-hoc networks Cloud Computing. Concurrency and Computation: Practice and Experience 27(8):2054–2063

[15] Inaba T, Sakamoto S, Kulla E, Caballe S, Ikeda M, Barolli L (2014) An Integrated System for Wireless Cellular and Ad-Hoc Networks Using Fuzzy Logic. International Conference on Intelligent Networking and Collaborative Systems (INCoS-2014) pp 157–162

[16] Jonsson A, Akerman D, Fitzgerald E, Nyberg C, Priyanto BE, Agardh K (2016) Modeling, implementation and evaluation of ieee 802.11ac in ns-3 for enterprise networks. Wireless Days (WD-2016) pp 1–6, DOI 10.1109/WD.2016.7461452

[17] Kulla E, Mino G, Sakamoto S, Ikeda M, Caballé S, Barolli L (2014) FBMIS: A Fuzzy-Based Multi-interface System for Cellular and Ad Hoc Networks. IEEE International Conference on Advanced Information Networking and Applications (AINA-2014) pp 180–185

[18] Lim A, Rodrigues B, Wang F, Xu Z (2004) k-Center Problems with Minimum Coverage. Computing and Combinatorics pp 349–359

[19] Maolin T, et al (2009) Gateways Placement in Backbone Wireless Mesh Networks. International Journal of Communications, Network and System Sciences 2(1):44

[20] Muthaiah SN, Rosenberg CP (2008) Single Gateway Placement in Wireless Mesh Networks. Proc of 8th International IEEE Symposium on Computer Networks pp 4754–4759

[21] Oda T, Barolli A, Spaho E, Xhafa F, Barolli L, Takizawa M (2012) Evaluation of WMN-GA for Different Mutation Operators. International Journal of Space-Based and Situated Computing 2(3):149–157

[22] Oda T, Barolli A, Xhafa F, Barolli L, Ikeda M, Takizawa M (2012) Performance Evaluation of WMN-GA for Different Mutation and Crossover Rates Considering Number of Covered Users Parameter. Mobile Information Systems 8(1):1–16

[23] Poli R, Kennedy J, Blackwell T (2007) Particle Swarm Optimization. Swarm intelligence 1(1):33–57

[24] Sakamoto S, Kulla E, Oda T, Ikeda M, Barolli L, Xhafa F (2013) A Comparison Study of Simulated Annealing and Genetic Algorithm for Node Placement Problem in Wireless Mesh Networks. Journal of Mobile Multimedia 9(1-2):101–110

[25] Sakamoto S, Kulla E, Oda T, Ikeda M, Barolli L, Xhafa F (2014) A Comparison Study of Hill Climbing, Simulated Annealing and Genetic Algorithm for Node Placement Problem in WMNs. Journal of High Speed Networks 20(1):55–66

[26] Sakamoto S, Kulla E, Oda T, Ikeda M, Barolli L, Xhafa F (2014) Performance Evaluation Considering Iterations per Phase and SA Temperature in WMN-SA System. Mobile Information Systems 10(3):321–330

[27] Sakamoto S, Lala A, Oda T, Kolici V, Barolli L, Xhafa F (2014) Application of WMN-SA Simulation System for Node Placement in Wireless Mesh Networks: A Case Study for a Realistic Scenario. International Journal of Mobile Computing and Multimedia Communications (IJMCMC) 6(2):13–21

[28] Sakamoto S, Oda T, Ikeda M, Barolli L, Xhafa F (2016) Implementation of a New Replacement Method in WMN-PSO Simulation System and Its Performance Evaluation. The 30th IEEE International Conference on Advanced Information Networking and Applications (AINA-2016) pp 206–211, DOI 10.1109/AINA.2016.42

[29] Schutte JF, Groenwold AA (2005) A Study of Global Optimization using Particle Swarms. Journal of Global Optimization 31(1):93–108

[30] Shi Y (2004) Particle Swarm Optimization. IEEE Connections 2(1):8–13

[31] Shi Y, Eberhart RC (1998) Parameter Selection in Particle Swarm Optimization. Evolutionary programming VII pp 591–600

[32] Vanhatupa T, Hannikainen M, Hamalainen T (2007) Genetic Algorithm to Optimize Node Placement and Configuration for WLAN Planning. Proc of 4th IEEE International Symposium on Wireless Communication Systems pp 612–616

[33] Wang J, Xie B, Cai K, Agrawal DP (2007) Efficient Mesh Router Placement in Wireless Mesh Networks. Proc of IEEE Internatonal Conference on Mobile Adhoc and Sensor Systems (MASS-2007) pp 1–9

[34] Xhafa F, Sanchez C, Barolli L (2009) Ad hoc and Neighborhood Search Methods for Placement of Mesh Routers in Wireless Mesh Networks. Proc of 29th IEEE International Conference on Distributed Computing Systems Workshops (ICDCS-2009) pp 400–405

[35] Xhafa F, Sun J, Barolli A, Biberaj A, Barolli L (2012) Genetic Algorithms for Satellite Scheduling Problems. Mobile Information Systems 8(4):351–377

Impact of Delayed Acknowledgment for Message Suppression in Vehicular-DTN

Daichi Koga, Makoto Ikeda and Leonard Barolli

Abstract In our previous work, we proposed a method of Enhanced Message Suppression Controller (EMSC) considering delayed acknowledgment for vehicular Delay/Disruption Tolerant Networking (DTN). In this paper, we evaluate the performance of our proposed method for message suppression in Vehicular-DTN. The simulation platform based on Scenargie simulator has been developed in order to evaluate the impact of persist timer to control the delayed acknowledgment. We consider 802.11p standard and send bundle messages in a Manhattan grid scenario. From the simulation results, we observed that our proposed method can increase efficiency with less network resource consumption and higher packet delivery ratio.

Key words: Vehicular-DTN, Enhanced Message Suppression Controller, Delayed Acknowledgment

1 Introduction

The Delay/Disruption Tolerant Networking (DTN) aims to provide inter-operable communications with wide range of networks which have poor performance characteristics. DTN is an end-to-end architecture providing communications in and/or through highly stressed environments [3]. Stressed networking environments in-

Daichi Koga
Graduate School of Engineering, Fukuoka Institute of Technology (FIT),
3-30-1 Wajiro-higashi, Higashi-ku, Fukuoka 811-0295, Japan
e-mail: daichi.kg@outlook.com

Makoto Ikeda and Leonard Barolli
Department of Information and Communication Engineering,
Fukuoka Institute of Technology,
3-30-1 Wajiro-higashi, Higashi-ku, Fukuoka 811-0295, Japan
e-mail: makoto.ikd@acm.org, barolli@fit.ac.jp

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_18

clude those with intermittent connectivity, large and/or variable delays, and high bit error rates.

The DTN architecture has been applied for vehicular networks called Vehicular-DTN. It has the potential to interconnect vehicles in regions that current networking protocol cannot reach the destination. Based on the wireless network technologies, Vehicular-DTNs comprise Vehicle-to-Vehicle (V2V), Vehicle-to-Pedestrian (V2P), Vehicle-to-Infrastructure (V2I) and Vehicle-to-X (V2X) communications [5, 8, 2, 20, 14, 15, 6].

In [9], we proposed a Message Suppression Controller (MSC) for V2V and V2I communications. The MSC was an expanded version of MS method [10]. We have proposed some parameters to control the message suppression dynamically. The simulations were conducted in urban environment considering obstacles (buildings), where MSC embedded in Road-side units (MSCRs) were present and not present.

Then, we proposed Enhanced Message Suppression Controller (EMSC) for Vehicular-DTN [11]. The EMSC is an expanded version of MSC [9] and can be used for various network conditions. But, the number of control packets were increased. To solve the overhead problem, we proposed an improved message suppression controller by considering delayed acknowledgment [12]. In [12], we only considered MS-ACK as evaluation metric.

In this paper, we evaluate the performance of proposed EMSC considering delayed acknowledgment for Vehicular-DTN. We use number of bundle messages and number of MS-ACKs as evaluation metrics. We developed the simulation platform based on Scenargie [17] network simulator for this work.

The structure of the paper is as follows. In Section 2, we give an overview of DTN. Section 3 provides a detailed description of our proposed EMSC considering delayed acknowledgment. The simulation system design is shown in Section 4. In Section 5, we show the simulation results. Finally, conclusions and future work are given in Section 6.

2 Delay/Disruption Tolerant Networking

DTN are occasionally connected networks, characterized by the absence of a continuous path between the source and destination [7, 1]. The data can be transmitted by storing them at nodes and forwarding them later when there is a working link. This technique is called message switching. Eventually the data will be relayed to the destination. The inspiration for DTNs came from an unlikely source: efforts to send packets in space. Space networks must deal with intermittent communication and very long delays [19]. In [7], the author observed the possibility to apply these ideas for other applications.

Laoutaris et al. [13] have studied this model and find that it can provide substantial capacity at little cost, and that the use of a DTN model often doubles that capacity compared with a traditional end-to-end model.

The main assumption in the Internet that DTNs seek to relax is that an end-to-end path between a source and a destination exists for the entire duration of a communication session. When this is not the case, the normal Internet protocols fail. DTNs get around the lack of end-to-end connectivity with an architecture that is based on message switching. It is also intended to tolerate links with low reliability and large delays. The architecture is specified in RFC 4838 [4].

Epidemic is flooding-based DTN routing protocol [16, 21]. Vehicles continuously replicate and send messages to newly discovered nodes that do not already possess a copy of the message. In the most simple case, Epidemic routing is flooding, but more elaborated techniques can be used to limit the number of message transfers.

Bundle protocol has been designed as an implementation of the DTN architecture. A bundle is a basic data unit of the DTN bundle protocol. Each bundle comprises a sequence of two or more blocks of protocol data, which serve for various purposes. In poor conditions, bundle protocol works on the application layer of some number of constituent Internet, forming a store-and-forward overlay network to provide its services. The bundle protocol is specified in RFC 5050 [18]. It is responsible for accepting messages from the application and sending them as one or more bundles via store-carry-forward operations to the destination DTN node. The bundle protocol runs above the TCP/IP level. In other words, TCP/IP may be used over each contact to move bundles between DTN nodes. This positioning raises the issue of whether the bundle protocol is a transport layer protocol or an application layer protocol. The bundle protocol provides a transport service for many different applications.

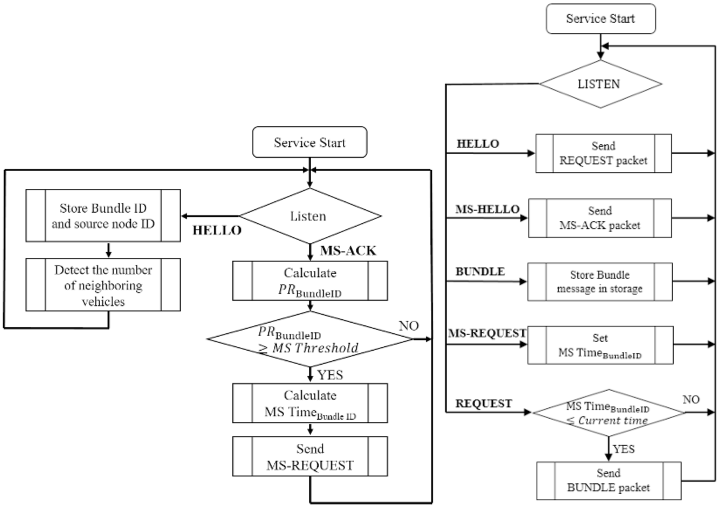
3 EMSC Considering Delayed Acknowledgment

The EMSC algorithm is based on Epidemic [16, 21] routing protocol. We present the flowchart of EMSC algorithms for two states of vehicles as shown in Fig. 1(a) and Fig. 1(b). EMSC is embedded in Vehicles (EMSCV).

When EMSCV receives HELLO message from other vehicles, the EMSCV stores the bundle ID and source node ID in the memory (see Fig. 1(a)). In addition, EMSCV detects the number of neighboring vehicles N to calculate the duration of message suppression.

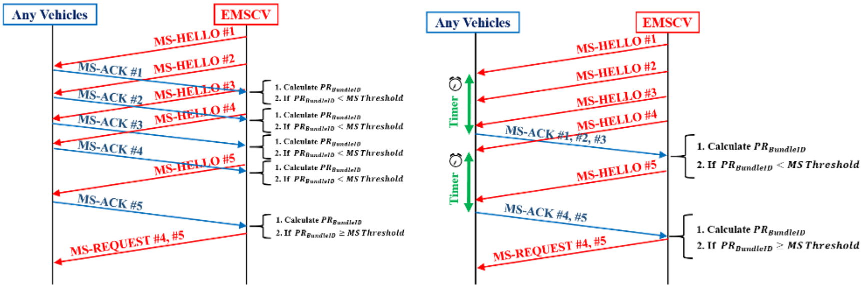
In our previous work, EMSCV sends the MS-HELLO packet to vehicles in intervals of 1 second (see Fig. 2(a)). When vehicle receives the MS-HELLO from EMSCV, the vehicle sends a MS-ACK to the EMSCV as shown in Fig. 2(a). In this case, vehicle receives many MS-HELLOs in short time, thus increasing processing overhead.

In this paper, we evaluate the performance of EMSC considering delayed acknowledgment to solve the problem. The method will be increasing efficiency by sending fewer packets than the conventional EMSC. The delayed acknowledgment has a persist timer as shown is Fig. 2(b).


Service Start
Listen
HELLO
Store Bundle ID and source node ID
Detect the number of neighboring vehicles
MS-ACK
Calculate $PR_{BundleID}$
$PR_{BundleID} \geq MS\ Threshold$
NO
YES
Calculate MS $Time_{Bundle\ ID}$
Send MS-REQUEST
Service Start
LISTEN
HELLO
Send REQUEST packet
MS-HELLO
Send MS-ACK packet
BUNDLE
Store Bundle message in storage
MS-REQUEST
Set MS $Time_{BundleID}$
REQUEST
MS $Time_{BundleID} \leq Current\ time$
NO
YES
Send BUNDLE packet

(a) Case 1 (b) Case 2

Fig. 1 Flowcharts of EMSCV functions.



(a) Conventional EMSC (b) EMSC cosidering delayed acknowlegment

Fig. 2 Sequence chart of different enhanced message suppression controller.

When the EMSCV receives the MS-ACK from a vehicle, the EMSCV calculates Possession Rate of the Bundle ID ($PR_{BundleID}$). The formula of $PR_{BundleID}$ is:

$$PR_{BundleID} = \frac{\text{Num. of detected same Bundle IDs (NB)}}{N_{new}}, \tag{1}$$

where N_{new} indicates the number of newly discovered vehicles from the EMSCV, which calculates from number of received HELLO packets. If $PR_{BundleID} \geq MSThreshold$, EMSCV sends MS-REQUEST to the vehicle. $MSThreshould$ (MST) is threshold value. In this paper, we set the MST 1.

Then, we used following steps to calculates the Message Suppression Time (MS Time) for each MS-REQUEST. Before EMSCV sends MS-REQUEST, the EMSCV calculates the MS Time based on NB, N_{new} and N values as shown in Eq. (2) and Eq. (3).

$$\text{MS Time}_{BundleID} = \text{Current time} + (\text{NB} \times N_{new} \times R_{new}) \qquad (2)$$

$$R_{new} = \frac{N_{new}}{N} \qquad (3)$$

MS $\text{Time}_{BundleID}$ will be used in MS-REQUEST to suppress the bundle message of its bundle ID. After that, the vehicle does not send the bundle message to other vehicles until the MS $\text{Time}_{BundleID}$ will be expired. As an exception, when vehicle moves to near end-points, the vehicle sends the bundle message to end-point even if MS $\text{Time}_{BundleID}$ is not expired.

R_{new} indicates the ratio of the newly discovered vehicles at each EMSCV. The stored N and N_{new} will be reset every second. In order to reduce the storage usage in the network, we use five functions as shown in Fig. 1(b). When a vehicle receives HELLO packet from other vehicles, the other vehicles send REQUEST packet to the vehicle. Then, if Current time is greater than MS $\text{Time}_{BundleID}$, the vehicle sends the bundle message to other vehicles. When vehicle receives bundle message, the vehicle stores the bundle message in the memory. In this way, the amount of traffic can be reduced.

4 Simulation System Design

4.1 Scenario Settings

In this work, we consider two urban area (Manhattan grid) scenarios with 100 and 150 vehicles, respectively. In the first scenario, we consider the obstacles (buildings), where 10 EMSCVs are present which are located on the road for forwarding the bundle messages considering the EMSC with delayed acknowledgment algorithms to all vehicles. In Fig. 3 is shown the evaluation scenario called VDTN.

In the second scenario, 15 EMSCVs are present, which are located at road for forwarding the bundle messages considering the EMSC with delayed acknowledgment algorithms to all vehicles.

The start-point sends a bundle message to four end-points during simulation. Both start-point and end-points are static. The other vehicles and EMSCVs move ac-

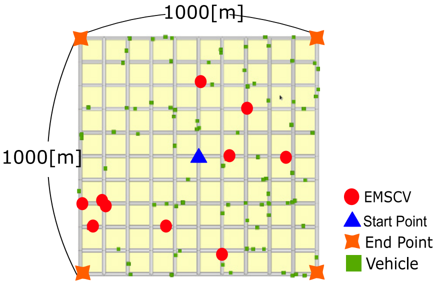


Fig. 3 VDTN scenario.

cording to Geographic Information System (GIS)-based Random Waypoint (RWP) mobility model.

EMSCVs send the MS-HELLO to all near vehicles according to EMSC algorithm for one second (see Fig. 2).

In simulations, every vehicle has an on-board unit for V2V communication and to display the traffic information.

4.2 IEEE802.11p

In our simulations, we considered IEEE 802.11p standard. The Scenargie network simulator has implemented the 802.11p standard. IEEE 802.11p is an approved amendment to the IEEE 802.11 standard to add Wireless Access in Vehicular Environments (WAVE). It defines enhancements to 802.11 required to support ITS applications. The 802.11p standard is based on the 802.11 architecture, but version "p" is aimed at communications between V2V or between V2I. This new technology uses the 5.9 GHz band in various propagation environments: vehicle, open, urban, and so on. This standard defines the WAVE as the signaling technique and interface functions that are controlled by the physical layer (MAC) devices where the physical layer properties change rapidly and where the exchanges of information have a short duration. The purpose of this standard is to provide a set of specifications to ensure interoperability between wireless devices trying to communicate in rapidly changing environments and in particular time periods.

4.3 Mobility Model for Vehicular-DTN

RWP mobility model is commonly used in most simulations. In RWP mobility model all nodes are considered mobile. In this model, the nodes move from one waypoint to another independently from each other. Every node in the network follows the following directions:

1. Selects a random starting position in the simulation area ($X \times Y$).
2. Selects a random location in the area as the next waypoint (W).
3. Starts moving towards the destination W, with a randomly chosen speed between $V_{\min}$ and $V_{\max}$.
4. After reaching destination W, the node pauses for a randomly chosen T_p seconds.
5. Then it repeats steps 2-4, until the simulation time $T_{\max}$ finishes.

GIS-based RWP mobility model is implemented for being used in V-DTN and constrains vehicle movement to streets defined by map data for real cities and limits their mobility according to vehicular congestion and simplified traffic control mechanisms. This mobility model provides reasonable run-times and memory consumption that scales fairly well with the size of the simulated network.

4.4 Application Settings

We model the application (for broadcasting bundle information) with message size of 500 bytes. Simulation parameters are shown in Table 1. We evaluate the performance considering number of sent bundles and number of sent MS-ACKs for different timers. The persist timer of delayed acknowledgment is set from 1 to 5 seconds.

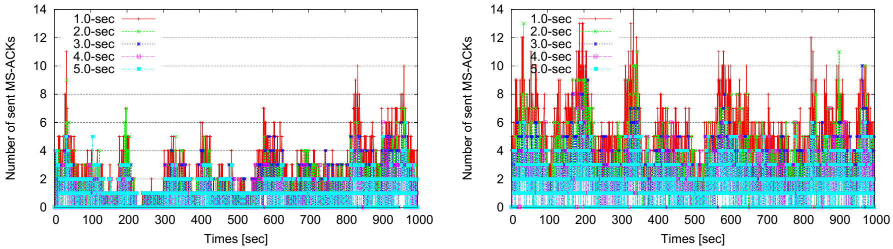
Table 1 Simulation parameters.

Parameter	Value
Simulation Time	1000 sec
Area Dimensions	1000 m ×1000 m
Number of Vehicles	100, 150
Number of EMSCVs	10, 15
Mobility Model	GIS-Based RWP
Minimum Speed ($V_{\min}$)	8.333 m/s
Maximum Speed ($V_{\max}$)	16.666 m/s
Routing Algorithm	EMSC
EMSC: Timer of Delayed Ack	1 - 5 sec
EMSC: MS Threshold	1
EMSC: Refresh Interval	1 sec
Application	Bundle Message
Number of Start Point	1 (Fixed)
Number of End Points	4 (Fixed)
Bundle: Start and End Time	1 - 1000 sec
Bundle: Message Sent Interval	10 sec
Bundle: Message Size	500 bytes
PHY Model	IEEE 802.11p
Frequency	5.9 GHz
Propagation Model	ITU-R P.1411
Antenna Model	Omni-directional
Antenna Height	1.5 meters

5 Simulation Results

We evaluated the network performance of EMSC considering delayed acknowledgment for different timers. In Fig. 4 are shown the results of number of sent MS-ACKs for different duration of timers. For both scenarios, the number of sent MS-ACKs decreased with increase of timer duration. For 150 vehicles, the difference of performance is big compared with first scenario.

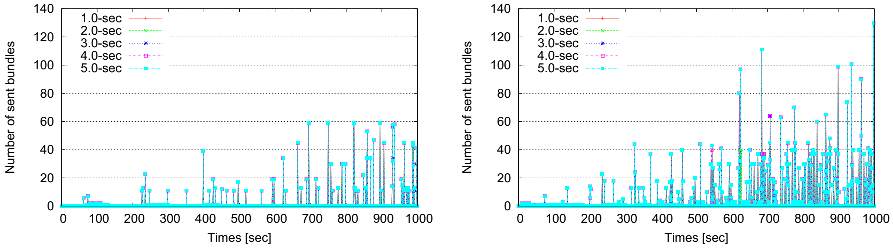
In Fig. 5 are shown the results of number of sent bundle messages for different duration of timers. For both scenarios, the difference of performance is small, even if when duration of timer was 5 seconds. From these results, we observed that our proposed method increased the efficiency by sending fewer control packets than the conventional EMSC.



(a) 100 vehicles

(b) 150 vehicles

Fig. 4 Results of sent MS-ACKs for different timers.



(a) 100 vehicles

(b) 150 vehicles

Fig. 5 Results of sent bundles for different timers.

6 Conclusions

In this paper, we evaluated the performance of EMSC considering delayed acknowledgment for Vehicular-DTN. We considered number of sent MS-ACKs and number of sent bundles for different duration of persist timers. We consider 802.11p standard and send bundle messages in a Manhattan grid scenario. From simulations, we observed that our proposed EMSC which considers delayed acknowledgment can increase the efficiency with less network resource consumption.

In the future, we would like to make extensive simulations to evaluate the proposed method and compare the performance with other DTN protocols.

References

1. Delay- and disruption-tolerant networks (DTNs) tutorial. NASA/JPL's Interplanetary Internet (IPN) Project (2012), `http://www.warthman.com/images/DTN_Tutorial_v2.0.pdf`
2. Araniti, G., Campolo, C., Condoluci, M., Iera, A., Molinaro, A.: Lte for vehicular networking: a survey. IEEE Communications Magazine 21(5), 148–157 (May 2013)
3. Burleigh, S., Hooke, A., Torgerson, L., Fall, K., Cerf, V., Durst, B., Scott, K., Weiss, H.: Delay-tolerant networking: an approach to interplanetary internet. IEEE Communications Magazine 41(6), 128–136 (2003)
4. Cerf, V., Burleigh, S., Hooke, A., Torgerson, L., Durst, R., Scott, K., Fall, K., Weiss, H.: Delay-tolerant networking architecture. IETF RFC 4838 (Informational) (April 2007)
5. Cheng, X., Yao, Q., Wen, M., Wang, C.X., Song, L.Y., Jiao, B.L.: Wideband channel modeling and intercarrier interference cancellation for vehicle-to-vehicle communication systems. IEEE Journal on Selected Areas in Communications 31(9), 434–448 (August 2013)
6. Dias, J.A.F.F., Rodrigues, J.J.P.C., Xia, F., Mavromoustakis, C.X.: A cooperative watchdog system to detect misbehavior nodes in vehicular delay-tolerant networks. IEEE Transactions on Industrial Electronics 62(12), 7929–7937 (December 2015)
7. Fall, K.: A delay-tolerant network architecture for challenged Internets. In: Proceedings of the International Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications. pp. 27–34. SIGCOMM '03 (2003)
8. Grassi, G., Pesavento, D., Pau, G., Vuyyuru, R., Wakikawa, R., Zhang, L.: VANET via named data networking. In: Proceedings of the IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS 2014). pp. 410–415 (April 2014)
9. Honda, T., Ikeda, M., Ishikawa, S., Barolli, L.: A message suppression controller for vehicular delay tolerant networking. In: Proceedings of the 29th IEEE International Conference on Advanced Information Networking and Applications (IEEE AINA-2015). pp. 754–760 (March 2015)
10. Honda, T., Ishikawa, S., Ikeda, M., Barolli, L.: A message suppression method for vehicular delay tolerant networking. In: Proceedings of the 5th International Workshop on Methods, Analysis and Protocols for Wireless Communication (MAPWC-2014). pp. 351–356 (November 2014)
11. Ikeda, M., Ishikawa, S., Barolli, L.: An enhanced message suppression controller for vehicular-delay tolerant networks. In: Proceedings of the 30th IEEE International Conference on Advanced Information Networking and Applications (IEEE AINA-2016). pp. 573–579 (March 2016)
12. Koga, D., Ikeda, M., Barolli, L.: An improved message suppression controller considering delayed acknowledgment for vanets (September 2016), accepted, to appear in Proceedings of the 19th International Conference on Network-Based Information Systems (NBiS-2016)

13. Laoutaris, N., Smaragdakis, G., Rodriguez, P., Sundaram, R.: Delay tolerant bulk data transfers on the Internet. In: Proceedings of the 11th International Joint Conference on Measurement and Modeling of Computer Systems (SIGMETRICS '09). pp. 229–238 (2009)
14. Mahmoud, A., Noureldin, A., Hassanein, H.S.: VANETs positioning in urban environments: A novel cooperative approach. In: Proceedings of the IEEE 82nd Vehicular Technology Conference (VTC-2015 Fall). pp. 1–7 (September 2015)
15. Ohn-Bar, E., Trivedi, M.M.: Learning to detect vehicles by clustering appearance patterns. IEEE Transactions on Intelligent Transportation Systems 16(5), 2511–2521 (2015)
16. Ramanathan, R., Hansen, R., Basu, P., Hain, R.R., Krishnan, R.: Prioritized epidemic routing for opportunistic networks. In: Proceedings of the 1st International MobiSys Workshop on Mobile Opportunistic Networking (MobiOpp '07). pp. 62–66 (2007)
17. Scenargie: Space-time engineering, LLC, `http://www.spacetime-eng.com/`
18. Scott, K., Burleigh, S.: Bundle protocol specification. IETF RFC 5050 (Experimental) (November 2007)
19. Tanenbaum, A.S., Wetherall, D.J.: Computer Networks Fifth Edition. Pearson Education, Inc., Prentice Hall (2011)
20. Theodoropoulos, T., Damousis, Y., Amditis, A.: A load balancing control algorithm for EV static and dynamic wireless charging. In: Proceedings of the IEEE 81st Vehicular Technology Conference (VTC-2015 Spring). pp. 1–5 (May 2015)
21. Vahdat, A., Becker, D.: Epidemic routing for partially-connected ad hoc networks. Tech. rep., Duke University (2000)

Reputation based Access Control in Social Networks for Persona Management

Keonsoo Lee[1], Yunyoung Nam[2]
[1] Medical Information Communication Technology, Soonchunhyang University, Asan, Republic of Korea
keonsoo@sch.ac.kr
[2] Dept of Computer Science and Engineering Soonchunhyang University, Asan, Republic of Korea
ynam@sch.ac.kr

Abstract. Social network is a way of representing the personality. Postings which is written by an owner, and comments which are produced by friends are the significant elements in social network. Unintended postings are not made but unsolicited comments can be produced and influence the online persona of the owner. Therefore a method of preventing such comments is needed. In this paper, we propose a reputation based access control for persona management. This method is to allow the right of making comments to only the deserved friends. This qualification is determined by comparing the reputation of each friend and the characteristic of each posting. The reputation of each social network user is made from the behaviors and their social evaluation. With this method, unnecessary fray and controversy are avoided without provoking social alienation.

1 Introduction

Social networking service (SNS) provides a chance of communicating with various people overcoming the restriction of time and space. Through SNS, various ideas and thoughts are shared with various peoples. Social exchanges are increased not only in quantity but also in quality with the help of SNS. However, the conflicts and private issues are also increased. As the desire of being exposed to the public increases, the desire of keeping privacy, which is easily violated in online, grows. The delicate difference between being exposed and being taken off, lay in the identity of the audience. As long as we control the audience by allowing only the authorized viewer to watch the approved article, the privacy can be protected. But in real environment, it is almost impossible to manage the viewers' permission for the articles properly. Therefore, an automated permission management method is required.

In order to determine who will be allowed for which article, reputation is used. For example, when articles about travels are provided to those who do not agree with the worth of trips, arguments occur. This is not what we expect in SNS. We use SNSs to be happy, to feel the sense of belonging, to be connected, and to win approvals [1]. Using reputation, the reaction for articles can be predicted. Therefore, conflicts and violations can be evaded by allowing articles only to viewers who can accept and

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_19

empathize with the articles. In this paper, we propose a method of using a reputation to manage the permission for articles.

2 Background

2.1 Access Control

Access control is a method of protecting security especially in computer systems [2]. The foundation of access control is that only a user who are authorized for the resource can access the resource. As the number of users and resources increases, the computing load becomes heavier. In order to resolve this problem, Role-Base Access Control (RBAC) is proposed [3]. RBAC employs roles to enhance the efficiency of managing each user's right of accessing to resources. Each role has a permission for a set of resources and a role is assigned to each user. Via the assigned role, users have permissions for resources. Therefore, it is possible to authorize various users' permission simultaneously using roles.

The advantages of RBAC are sustainable as long as the number of roles is smaller than the number of users. In worst case, when the number of roles is equal to the number of users, RBAC becomes useless. In modernistic computing environment where customized and personalized services are provided, roles are not sufficient for managing each user's access control [4]. The proposed method in this paper is not to replace RBAC but to enhance it by providing additional functionality which automatically modifies the set of permissions for resources in the system. The automatic modification is executed based on the reputation of a user.

2.2 Reputation

Reputation is defined as a common opinion that people have about someone or something [5]. When we have a belief that we know a person well enough, we do not need his/her reputation. However, when we believe that we do not know the person enough, his/her reputation becomes important. In short, using reputation is a way of getting help from others by borrowing their experiences. In open collaboration system where various 3rd party computing objects are collaborated to achieve a shared objective, recruiting reliable partners is one of the most important issues. As it is impossible to evaluate all the possible computing objects, reputations are employed [6]. The opinion of others who have experienced the computing object can used to make a better decision.

However, reputation is dynamically changing and affected by various elements in the environment. For example, a person who is regarded as good can be regarded as bad in other situation. Therefore, the reliability of given reputation should be considered. In multi-agent systems, directory facility (DF) and Agent Management System (AMS) control all the agents in the given environment. The reliability of computing objects is guaranteed by the system [7]. However, in open system where the sources of reputation is trustless, a method of managing the uncertainty of reputation is required.

3 Proposed Method

3.1 A Method of Calculating Reputation

Reputation is calculated from the behaviors of a person in social networks. The behavior is classified into two types. One is the expression of his/her own thought. The other is the expression of his/her feeling for others' thought. The element of the first type is an article or a post. The elements of the second type are replies for others' articles or posts. Therefore, reputation of a person is a set of what s/he thinks and how s/he responses to others' thoughts. Every writing has two main properties which are a topic and a standpoint for the topic. For example, Malcolm X and Martin Luther King Jr. were interested in the same topic. But their standpoints were different. The standpoint is represented in friendliness. If a user agrees a specific topic with highly aggressively, his/her friendliness for the topic will be a positive maximum value. In this paper, we set the range of friendliness property from positive 5 to negative 5.

The thought of users are expressed in text or emoji. The difference between text and emoji is the simplicity. The feeling expressed in emoji is more explicit and clear. Even though, emoji is a convenient tool for presenting emotion, the strength of the emotion presented in emoji is weaker than that presented in sentences. Because emoji is used as a cliché, when serious feelings need to be presented, characters are preferred in formal form.

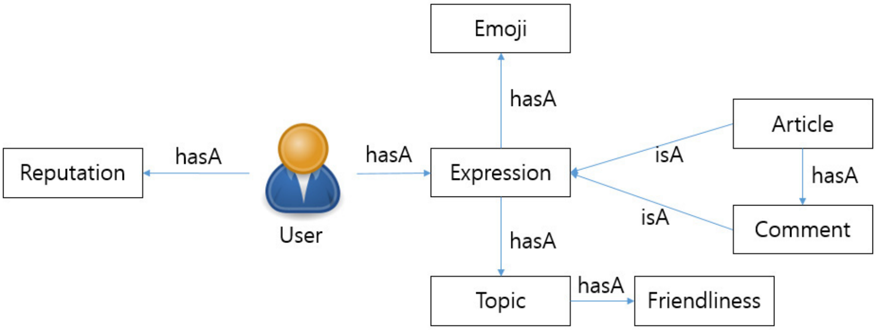


Fig 1. The relationship among elements for representing reputation.

Figure 1 shows the relationship among elements for representing a user's reputation in SNS. Table 1 shows descriptions for each element.

Table 1. Five elements that are used to calculate the reputation of users form their behaviours in SNS.

Element	Description
Article	Article is a post that is written or copied by the host to represent his/her own thought.
Comment (Reply)	Comment is a small set of sentences written by readers to represent their impressions for the article. Therefore, a comment should have an article that it is attached to.
Topic	All articles and comments have a topic which is the theme of the writing. When a comment has a topic which is different from the topic of the article, which the comment is attached, it should be ignored and the reputation of the owner of such comments will be dropped.
Friendliness	Articles which have the same topic, may have different attitude for the topic. Friendliness of an article represent the pros and cons for the topic and the level of aggressiveness.
Emoji	Emoticons are widely used to assist the expression of feelings. As every emoji has explicit meaning, it can send a clear message than words which can be misunderstood and misinterpreted.

Reputation consists of two sub-elements which are general reputation and reputations for each topic. A gentleman for a topic can be a gangster for another topic. Psychological complex is one of the reasons for such inconsistency in human mind [8]. Even though, general consistency can be found for tastes. The trend of response for various issues becomes the general reputation. Specific reputation for each topic is the attitude for the given topic. From this reputation, a user who has the possibility of conflicts is prevented from the article. For an article whose topic has not been exposed to a user, the user's general reputation is used for determining the right of accessing.

3.2 Rules of Reputation based Access Control

Reputation consists of two sub-elements. General reputation is the average ratio of friendliness. From the general reputation, generosity of the personality is assumed. Specific reputation is a set of friendliness for each topic. The riskiness of a quarrel increases when those who have different friendliness for the same topic are gathered in the same place. As both extremes of friendliness have higher possibility of arguments, these two ideas are used for managing specific reputation. The first idea is not to allow extremists to be gathered. The second idea is to find a broad-minded users who can be friends with those who have fiery temperaments.

The heuristic instantiated rules from the ideas are shown in Table 2. When general reputation and specific reputation are collided, specific reputation is preferred. Therefore, for a topic for which the user's history is empty, general reputation is used to authorize the right to access. For topics for which the user's history is filled, specific reputation for each topic is used.

Table 2. Heuristic rules for managing reputation.

Rules	Description
Definition of GR	General Reputation (GR) is made by averaging the set of specific reputations : GR = max(types of friendliness)
Definition of SR	Specific Reputation (SR) is a set of friendliness for each topic : SR = <Friendliness$_1$, Friendliness$_2$, … , Friendliness$_x$>
Authorizing Permissions	Access is allowed when the reputation of a guest is similar to the reputation of the resource owner for the topic of the resources
Priority of Order	GR is used only for the resources which have no SR

4 Simulation and Results

In order to evaluate the proposed method, we collect 300 broken cases of social relationship from twitter[9]. Twitter is one of the most popular SNSs. The social relation is made by selecting to follow others. Following is an action of subscribing other's post which is represented within 140 characters. Those who follow are called follower. This relation between follower and a user of this service is broken when the user blocks his/her follower. Unlike facebook[10], making reply for others' posts is not allowed but making new post with reference for others' posts is possible.

From the collected cases, the causes of the brokenness is found by analyzing the text with TweeboParser [11]. Causes of brokenness are categorized into 'Unknown', 'Disagreed Opinion', and 'Violation'. The analyzed results of 300 posts are shown in Table 3. In twitter, relations are easily connected and disconnected. However, as shown in the results, reputation based access control can prevent 64% of brokenness of relationship by allowing only the acceptable expressions to be shared.

Table 3. Results of post analysis from twitter.

Cause		Ratio	Description
Unknown		24%	There is no clear reason why they are blocked
Disagreed Opinion	Implicit Disagreement	27%	They represent some opinion and then blocked
	Explicit Argument	12%	They represent opinions and argument occurs. Then they are blocked
Violation	Service Policy	12%	They violate the contract of services and are blocked
	Teasing	22%	They make fun of others and are blocked
	Cursing	3%	They call names and are blocked

5 Conclusion

The main objective of reputation based access control in social networks is to prevent hooligans or troublemakers from vandalizing the other's personal relationship. From the articles, existing friends, and attitude to others' article, the reputation is estimated. By using this reputation, only the users who have acceptable attitudes and opinions are allowed to access the article in social networks. This access control can reduce the possibility of online conflicts. However, it is not a good solution to keep away from persons who are not similar to us for enhancing ourselves and enlightening our contemporaries. Tolerance is one of the most important properties of civilized democratic society. And it can be learned by communicating with others who have different viewpoints. Restricted social boundary may provide convenience and comfort but progress cannot be achieved in such coziness. Using reputation for social networks is a kind of double-edged sword. The responsibility of correct usage is on each user.

Acknowledgments. This research was financially supported by the “ICT Convergence Smart Rehabilitation Industrial Education Program” through the Ministry of Trade, Industry&Energy(MOTIE) and Korea Institute for Advancement of Technology(KIAT) and was also supported by the Bio & Medical Technology Development Program of the NRF funded by the Korean government, MSIP(NRF-2015M3A9D7067219).

References

1. “The 10 Top Reasons Why We Use Social Networks,” WeRSM | We Are Social Media, 06-May-2015
2. R. S. Sandhu and P. Samarati, “Access control: principle and practice,” IEEE Communications Magazine, vol. 32, no. 9, pp. 40–48, Sep. 1994
3. D. F. Ferraiolo and D. R. Kuhn, “Role-Based Access Controls,” arXiv:0903.2171 [cs], Mar. 2009
4. L. Rostad and O. Nytro, “Personalized Access Control for a Personally Controlled Health Record,” in Proceedings of the 2Nd ACM Workshop on Computer Security Architectures, New York, NY, USA, 2008, pp. 9–16
5. “Definition of REPUTATION.” [Online]. Available: http://www.merriam-webster.com/dictionary/reputation. [Accessed: 31-Aug-2016]
6. B. T. Adler, L. de Alfaro, A. Kulshreshtha, and I. Pye, “Reputation Systems for Open Collaboration,” Commun ACM, vol. 54, no. 8, pp. 81–87, Aug. 2011
7. S. D. Ramchurn, D. Huynh, and N. R. Jennings, “Trust in multi-agent systems,” The Knowledge Engineering Review, vol. 19, no. 1, Mar. 2004
8. J. Jolande, Complex/Archetype/Symbol In The Psychology Of C G Jung. Routledge, 2013
9. “Twitter.” [Online]. Available: https://twitter.com/. [Accessed: 09-Sep-2016].
10. “Facebook.” [Online]. Available: https://www.facebook.com/. [Accessed: 09-Sep-2016].
11. “Twitter Natural Language Processing -- Noah’s ARK.” [Online]. Available: http://www.cs.cmu.edu/~ark/TweetNLP/. [Accessed: 09-Sep-2016].

The Automatic Text Summarization Using Semantic Relevance And Hierarchical Structure Of Wordnet

Jun Seok Cha[1], Pan Koo Kim[2]
[1] Dept of Software Convergence Engineering
Chosun University
qwert3750@naver.com
[2] Dept of Computer Engineering
Chosun University
pkkim@chosun.ac.kr

Abstract. **In recent years, rapid development and spread of smart devices has resulted in increase in data of online documents on the Internet day by day. Growing information overload of web texts leaves users facing difficulties in browsing and understanding huge data in web pages. Therefore, in the field of automatic document summarization, diverse studies are underway to find ways of creating summaries efficiently. This study aims to propose document summarization methods using sentence segmentation and lexical chaining to extract important sentences of a given text and make a summary by excluding unnecessary sentences. Sentences of a given text are divided by analyzing their syntactic structure or identifying parts of speech of words and phrases and clauses used in the sentences. Important sentences are extracted by means of lexical chain. Results of previous document summarization research were improved through experiment, allowing for a summary using key points of a text.**

1 Introduction

Unlike document clustering sectors include information retrieval and document classification, automatic document summarization is to create a summary that contains key points of a given text. As a result, lots of quality information on the topic of the text is necessary to enhance results of automatic document summarization. Using unique features of a text such as relations between words or sentences depending on the topic for summarization provides significant benefits. Semantic relations of constituents of a text such as words and sentences are defined differently as coherence and cohesion in text linguistics [1]. In this paper, we suggest a document summarization method based on sentence segmentation and lexical chaining using ranking rules to raise efficiency of document summarization. Generally, texts composed of long sentences are of various syntactic structures and summarized with

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_20

single document summarization or multi-document summarization; multi-document summarization extracts important words and lexicon from the original and generates a summary with new sentences using natural language processing techniques. In contrast, single document summarization picks out sentences deemed important from the original only to make a summary, which has been frequently studied. In this study, sentences are segmented with rule-based phrases and clauses for better document summarization. Important sentences are extracted from the segmented sentences, and key words are extracted from collected data before creating a summary. On the basis of extracted key words, lexical chains are generated for summary. In regards to lexical chaining, a specific weighting scale is used to choose highly weighted lexical chains, on the basis of which important sentences are extracted. This study is designed to suggest a document summarization method using lexical chaining and how lexical chains help improve document summarization results. In order to create lexical chains and investigate semantic relations between words in a text, this study hired the hierarchy system of WordNet and examined meanings of individual words in question. This study consists of five chapters. Chapter 2 reviews document summarization related literatures, chapter 3 introduces how to summarize textual documents, and chapter 4 provides results of experiment with the proposed method. Finally, chapter 5 presents conclusions and suggestions for further study.

2 Related Work

There are two types of document summarization methods to extract representative sentences, remove unnecessary ones, and finally reduce sentences of a text: multi-document summarization and single document summarization. First, multi-document summarization has been proposed to extract important key words of sentences based on the topic of a given text and summarize the entire text in alignment method [2]. This method is used to extract the topic of sentences via prior processing and create a summary. However, it requires huge calculation cost for topic search processing. Another approach is WordNet thesaurus-based query method where core sentences closest in relation to queries of a given sentence are extracted using WordNet thesaurus and then supplementary sections that will support the summary are extracted from the remaining sentences [3]. In addition, there is also a method in which text and query similarity and similarity between the current paragraph and the previous paragraph are calculated for summarization. As this method uses simply statistical processing, errors may happen, compared to complex natural language processing or information extraction [4]. Besides, research has been in progress to analyze sentences and words in a long text and weigh representative and meaningful sentences in the text to extract representative sentences on the basis of high weighted values [5]. Finally, in document clusters, document summarization method was proposed in which topics of a text are grouped with classification algorithm like k-means, relations between sentences and clusters are analyzed with link analysis algorithm and Markov algorithm, and then texts are summarized by giving scores to sentences [6]. Secondly, when it comes to single document summarization, research focuses on automatic summarization of sentences once by reducing sentences into concise summary upon input. Among several approaches, a method was proposed to summarize a text using important sentences and user-intervened query expansion [7]. Another approach suggested is to use information on the topic included in syntactic structure of the text and frequency of terms for document summarization [8], and another research based on sentence query is also underway to come up with query-specific document summarization method by applying complex topics extracted from the text using sentences closest to sentence query and semantic relations [9]. This study examines how to improve existing English text analysis research based on ranking rules [10].

3 LEXICAL CHAIN-BASED AUTOMATIC DOCUMENT SUMMARIZATION

Sentence segmentation comes before summarization. Sentences are segmented in ranking rules which allow for segmentation of collected document data into phrases and clauses, and new sentences are generated in the context of segmented sentences. In this case, if any of the new sentences are not essential to summarize, then they are removed at the time of summarizing the text, and representative sentences are only taken into account. Fig 1. shows diagram of the system described in this study.

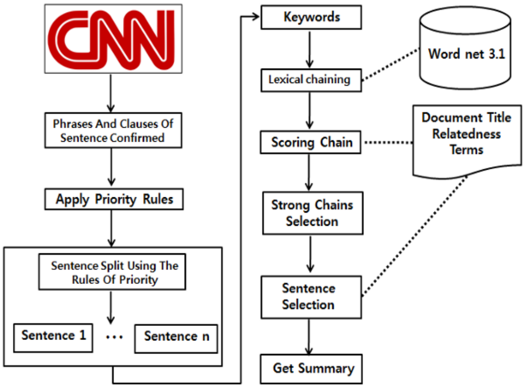


Figure 1. Extracting Summary Sentences from Lexical Chains

A. Sentence segmentation for Document Summarization

In this chapter, sentences are segmented before document summarization. In order to segment sentences, phrases and clauses need to be identified, with sentences segmented in the ranking process. First, data were collected from CNN News website through crawling, followed by identification process of phrases and clauses. Table 1 shows contexts of 11 possible locations for sentence segmentation and locations of sentence segmentation. In Table 1, sentence segmentation rankings were identified with consideration for in which locations sentences should be segmented, and sentences are segmented into phrases and clauses in a parallel structure for ranking-based sentence segmentation. In order to extract representative sentences from the segmented sentences and produce a summary, segmentation should be carried out in the correct syntactic structure. If conditions for a phrase as listed in Table 1 are met, it is ranked ‘0.’ Since phrases have easier conditions to form and appear more often in a sentence than clauses, they are ranked ‘0.’ Clauses are ranked ‘I.’ Under ranking rules as shown in Table 1, ‘0’ is the highest rank, and bigger numbers means lower ranks.

Sentences at rank ‘0’ are segmented if there are no subjects and verbs and they consist of two or more words and have words with meanings as a part of speech. Moreover, Sentences are at rank ‘1’ if they have subjects and verbs and words starting a clause. Finally, words starting a predicational clause in a sentence are ranked ‘2’, the lowest rank. According to these ranking rules, important sentences are extracted from sentences segmented into phrases and clauses to summarize a given text.

Table I. Priority rules for sentence Split

Type word	**Split Location**	**POS patterns**	**Priority**
NP	Two or more nouns meaning sentences having a word	“POS”+“NNS“+“NN“	0
		“POS”+“NN”+“NNS“	
		“POS”+“NNP“+“NN“	
AP	Two or more words and an adjective meaning sentences having	“POS”+“JJ“+“JJ“	0
AP	The position of the word in the statement that two or more of the role of the words and adverbs	“POS”+“ADV“+“ADV“	0
VP	Two or more words verb meaning sentences having	“POS”+“VBD“+“VBG“	0
		“POS”+“VBZ“+“VBG“	
		“POS”+“VBZ“+“VBD“	
		“POS”+“VBD“+“VBN“	
PP	Two or more words and prepositional sense sentences having	“POS”+“PRE“+“PRE“	0
SCONJ	The position of the word the beginning of the subordinating conjunction	“IN”+“VB”+“NN”	1
P_PRON	The position of the word of the relative pronoun of the Interrogatives preposition	“IN”+“VB”+“PRE“+”NNP	1
		“IN”+“VB”+“PRE“+”WP”	
IROGATIVE	The Position of words that begin with interrogative	“IN”+“VBG”+“WP”	1
VERB	The position of words that begin with the verb clause	“IN”+“POS“+”VB”	2
AUXVERB	The position of the word the beginning of the AUXVERB	“IN”+“POS“+”MD”	2

B. Lexical Chain-based Document Summarization

In sentence segmentation step, a document consisting of long texts are divided into multiple sentences. From the divided sentences, keyword extraction should be made to generate lexical chains for document summarization. To extract keywords, prior processing of data of news articles collected from the CNN website takes place. Prior processing goes through elimination of stop words, extraction of word stems, identification of parts of speech, and extraction of nouns. Titles of collected documents also are processed for keyword extraction. In regards to lexical chains, their efficiency increases as key words of a document grow in number. Stop words are eliminated using a list of stop words appearing in segmented sentences at the stage of stop word elimination as part of prior processing. Stop words refer to words not used often in search via the web. Secondly, word stem extraction is a step that extracts unnecessary steps like 'es' and 'ed.' At the third stage of part of speech identification before noun extraction, Java stanford-postagger is used to identify parts of speech of words such as noun, verb, and proper noun. Finally, noun extraction is designed to extract nouns only among other parts of speech of words whose part of speech is defined. Table 2 shows how key words are extracted

Table II. Keyword extraction process

Preprocessing	**CNN News Document**
Document title	Asia shares continue global rebound
Document	Asian stock markets have recorded more gains, continuing the positive lead set by the US and Europe on Tuesday. Wall Street and bourses across Europe have been recovering some. the ground since the UK voted last week to leave the European Union.
StopWords Remove	Asian stock markets **have** recorded more gains, continuing the positive lead **set by the** US **and** Europe **on** Tuesday. **Wall** Street **and** bourses across Europe **have been** recovering **some**. **the** ground **since the** UK voted **last** week **to** leave **the** European Union.
Stemming extract	Asian stock market**s** recorded gain**s**, continuing the positive lead US Europe Tuesday. Street bours**es** across Europe recovering. ground UK voted week leave European Union. Japan, benchmark Nikkei 225 index finished 1.6% higher 15,566.83.
Parts of Speech	Asian/JJ, stock/NN, markets/NNS, recorded/VBN, gains/NNS, continuing/VBG, positive/JJ, lead/NN, US/NNP, Europe/NNP, Tuesday/NNP, Street/NNP, bourses/NN, across/IN, Europe/NNP,ground/NNP, UK/NNP,
Noun extract	stock/NN, markets/NNS, gains/NNS, lead/NN, US/NNP, Europe/NNP, Tuesday/NNP, Street/NNP, bourses/NN, Europe/NNP,ground/NNP, UK/NNP, week/NN, Shares/NN, carmaker/NN, Toyota/NNP, irbags/NN

Lexical chains should be generated from the extracted key words, and thereby important sentences should be extracted from a given text. Lexical chain is a system that analyzes a text using lexical meanings only except grammatical devices and groups words according to their semantic relations. The document summarization method proposed in this study summarizes a text with lexical chains. To this end, hierarchies of WordNet are used; regarding nouns extracted via prior processing, synonyms, hypernyms, hyponyms, and antonyms of WordNet are searched to find two close words which have common features and appear commonly in hierarchy. Fig 2 illustrates lexical chains to investigate semantic relations between two words

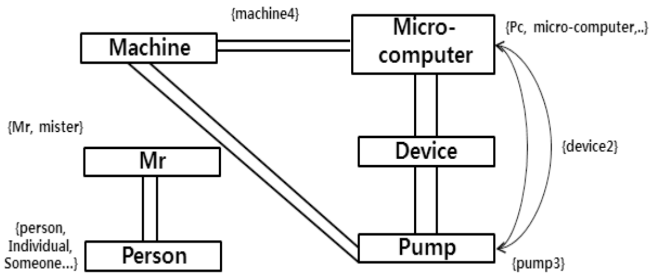


Figure 2. Lexical chain

As shown in Fig 2, if Machine is used as a meaning of “a person working mechanically or mechanical person,” there is a semantic relation between Person and Machine, resulting in a lexical chain. However, if Machine is used to mean “system or tool,” there is no semantic relation between “Person and Machine,’ creating no lexical chain. However, semantic relations are created among ‘Machine, Micro-computer, Device, and Pump”, giving birth to a lexical chain. Fig 3 shows a hierarchy of WordNet indicating semantic relation between two words Scoring Chain based on such lexical chains is required to summarize documents With a view to extracting key sentences of collected data of news articles, lexical chains with semantic relations are used to assign weighted values to segmented sentences. Lexical chains are created with nouns having defined meanings and weighted values are given to each lexical chain and segmented sentence in the following manner. Scores of nouns in lexical chains are determined by relations with other nouns. In this case, relations which nouns have in lexical chains are four types: synonyms, hypernyms, hyponyms, and antonyms. Weighted values in lexical chains are calculated in the following. equation:

$$\mathrm{Score}(\mathrm{chain}) > Average(Score) + 2 * StandardDeviation(Score) \quad (1)$$

In Equation 1, Average (Score) is given by means of lexical chain of each sentence. Even if there are no semantic relations between two words or there are words in a document from which important sentences are extracted, weighted values are given. As shown in Fig 2, no lexical chain is created between “Person and Machine” as there is no semantic relation between them, but if they appear in a sentence, 0.5 point is given as a weighted value. In addition, if there is a lexical chain, but they are not present in a sentence, 0 point is assigned. Finally, if “Machine and Micro-computer“ with a lexical chain are present in the sentence in question, weighted value is 1. Weighted values are calculated by sentence cluster of the document before computing standard deviation (SD) of weighted values as StandardDeviation (Scores) in Equation 1. On the basis of averages of lexical chains from SDs and weighted values, important sentences are extracted. Table 3 documents resulting lexical chains and weighted values.

Table III. Keyword extraction and lexical creation process chain

	processing
Document	Asian stock markets have recorded more gains, continuing the positive lead set by the US and Europe on Tuesday. Wall ….

	processing
Apply priotity rules	**[Asian stock markets have recorded more gains], [continuing the positive lead]** set by the US and Europe on Tuesday. Wall Street and bourses across Europe have been recovering some. …….
Split sentence	- Asian stock markets have recorded more gains - continuing the positive lead - set by the US and Europe on Tuesday.…….
keyword (Preprocessing)	stock/NN, markets/NNS, gains/NNS, lead/NN, US/NNP, Europe/NNP, Tuesday/NNP, Street/NNP, bourses/NN, …….
Lexical Chain	['Union'], ['share'], ['share', 'Shares'], ['shares', 'share'], ['vote'], ['ground'], ['Europe'], ['index', 'index'], ['Australia', 'Australia'], ['lead'], ['referendum', 'vote'], ['Tuesday', 'week'], ['markets'] …….

If union, share, and ground in Table 3 are present in a sentence of a given text, 0.5 point is given, and if lexical chains based on semantic relations such as “referendum, vote” or “Tuesday', week” appear in the sentence, 1 point is given. Sentences with an average of 2.1 or over as shown in weighted value averages in Table 2 are extracted as key sentences. Sentences 1 and 3 in Table 3 become key sentences, and sentence 2 is eliminated because its score is below average.

4 Experiment

With the aim to summarize a document using the method proposed in this study, data were collected from CNN news articles in the fields of politics, economy, and general issues. Using the data, sentences were segmented in the proposed method, and key words were extracted from the segmented sentences and document titles of the data. These keywords were used to create lexical chains, on the basis of which the document was summarized. Table 4 shows the result of lexical chain-based document summarization.

Table IV. Text Summary

	Summary document processing
Document	Asian stock markets have recorded more gains, continuing the positive lead set by the US and Europe on Tuesday. Wall Street and bourses across Europe have been recovering some …….
Lexical Chain	['Union'], ['share'], ['share', 'Shares'], ['shares', 'share'], ['vote'], ['ground'], ['Europe'], ['index', 'index'], ['Australia', 'Australia'], ['lead'], …….

	Summary document processing
Summary Results	Asian stock markets have recorded more gains, the ground since the UK voted last week to leave the European Union. In Japan the benchmark Nikkei 225 share index finished 1.6 higher at 15 566.83. Shares in commodity giants Rio Tinto and BHP Billiton were both up by almost.

5 Conclusions and suggestions for further work

The proposed automatic document summarization segments sentences on the basis of collected data in accordance with ranking principles, extracts key words of document titles and contents and thereby generates lexical chains. Upon generated, the lexical chains is used to summarize the segmented sentences. In order to generate lexical chains, common semantic relations between two words are examined with consideration for synonyms, hypernyms, hyponyms, and antonyms present in hierarchies in WordNet, and weighted values are given to sentences where each word appears. With these weighted values in mind, important sentences are extracted from the document, which makes it possible to summarize long texts. For enhanced key word extraction for document summarization, subsequent studies are expected to focus on N-gram based methods.

Acknowledgment

This research was supported by SW Master's course of hiring contract Program grant funded by the Ministry of Science, ICT and Future Planning(H0116-16-1013) and This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) (No. NRF-2016R1A2B4012638).

References

[1] YongdoKim, "Text binds Theory". Pusan University of Foreign Studies Press. 1996

[2] Harabagiu, S. Finley L. "Topic Themes for Multi Document Summarization," In proceeding of ACM SIGIR, 202-209, 2005.

[3] Sakurai, T., Utsumi, A. "Query-based Multi document Summarization for Information Retrieval," The Proceeding of NTCIR, 2004.

[4] Goldstein. J., Mittal. V. Carbonell. J., Callan. J.,"Creating and Evaluating Multi-Document Sentence Extract Summaries," The Proceeding of CIKM, 165-172, 2000.

[5] Inderjeet Mani, Automatic Summarization, Kohn Benjamins publishing Co., 2001.

[6] Xiaojun Wan, Jianwu Yang "Multi-Doucument Summarization Using Cluster-based Ling Analysis". Proceedings of the International Conference SiGIR'08, 2008. PP. 299-306

[7] Lewis, D.D, Sparck Jones, K. "Natural language processing for information retrieval," Communications of the ACM, Vol. 39, No.1, 1996, pp. 92-101

[8] Liddy, E.D, Myaeng, S.H. "DR-LINK's: linguistic-comceptual approach to document and detection," ," The First Text The First Text REtreival Conference (TREC-1), 1993, pp. 113-129.

[9] Morris, J, Hirst, G. "Lexical cohesion computed by thesaural relations as an indicator of the structure of text," Computational Linguistics, Vol.17, No.1, pp. 21-43, 1991.

[10] JunSeok Cha, Seunghyeon Bak, PanKoo Kim. "An Approach of Sentence Segmentation for Improving Parsing Effecieny" Korea MultimediA Society Vol. 19, No. 1, 2016. Pp.63-66

Korean spelling error correction using a Hangul similarity algorithm

SeungHyeon Bak[1], PanKoo Kim[2]
[1] Dept of Software Convergence Engineering, Chosun University, GwangJu, Korea
cronyandiver@gmail.com
[2] Dept of Computer Engineering, Chosun University, GwangJu, Korea
pkkim@chosun.ac.kr

Abstract. Increasingly people use computers for word processing. This helps reduce word processing time and fatigue of hands, but may increase the possibility of occurrence of spelling errors. Although spelling errors are generally easy to find and correct, it is hard to make a document totally free of spelling errors partly due to lack of knowledge of users or presence of spelling errors which are difficult to notice. Since there is no set of online word processing rules and manners in place and problems of spelling errors are not often raised, spelling errors in important documents may lead to decrease in reliability. Even experts cannot correct spelling errors perfectly, so there is a need for research to come up with spelling correction methods for the general public. This study aims to correct spelling errors using Korean alphabet similarity algorithm. To this end, words most similar to misspelled words found in a corpus containing spelling errors collected by previous research were identified to correct spelling errors by measuring frequency of simultaneous appearance with adjacent words.

1 Introduction

Long time has passed since computers which used to be a means of research were commercialized and available for the general public. People used writing instruments to write before computer was commercialized. However, today a growing number of them are using computers to write instead. Computerized word processing helps write faster and reduces fatigue of hands than writing instruments, making it better fit to making long texts. However, word processing programs are more likely to cause spelling errors by the mistake of users. Spelling errors distort the shape of words, making it easy for the writer to find and correct directly, but those caused due to users' lack of knowledge or those hard to find may make it almost impossible to produce a document free of spelling errors. Even though people often write for chatting or on their SNS pages on the Internet, there are no set of spelling rules and manners for cyberspace and as a result, serious issues have not be raised to date. However, spelling errors in important documents such as theses or business proposals may lead to falling reliability. These spelling errors should be completely avoided, but

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_21

it is not easy even for experts. Consequently, it is necessary to conduct research on high-level spelling error correction programs for the general public. This study was designed to produce a system to correct sentence-level spelling errors to normal words with Korean alphabet similarity algorithm. On the basis of findings reported in related literatures that corrected words are significantly similar to misspelled words in form, spelling errors were extracted from a corpus. Extracted corrected words were replaced with misspelled ones to correct spelling errors with spelling error detection algorithm. In this paper, chapter 2 documents previous works on spelling error correction and chapter 3 describes methods for spelling error correction study. Chapter 4 presents results of the study and finally chapter 5 discusses conclusions and directions for subsequent studies.

2 Related Work

Spelling errors refer to so-called typos, including misspellings, deletion or insertion of words which causes distortion of meaning, or replacement of words not in line with the context [1]. There are two types of spelling errors in general: simple spelling errors and context-sensitive spelling errors. The former means spelling errors which are not present in vocabulary dictionary and can be detected simply via morphological analysis, and therefore are easy to correct. In contrast, context-sensitive spelling errors refer to spelling errors which exist in vocabulary dictionary but are not in line with the context. They are not easy to detect and correct, compared to simple spelling errors and have been extensively studies up to now [2]. Context-sensitive spelling error correction methods are grouped largely into two approaches: rule-based methods and statistical methods. Rule-based methods are to correct spelling errors under man-made rules. These methods show better performance as the number of rules grows, but requires highly qualified experts and enormous costs for maintenance of rules, making the methods not much studied [3]. Statistical methods correct spelling errors using a statistical model and are subdivided primarily into n-gram linguistic model-based approaches and vocabulary pair for correction-based approaches [4]. The n-gram linguistic model-based methods work by extracting eojeol 3-gram from a large corpus and thereby measuring the probability of each sentence or partial sentence for spelling error correction. In Korea, morpheme n-grams are widely used due to the problem of postpositions and endings, but a study conducted spelling error correction using an eojeol n-gram model instead of a morpheme n-gram model with a view to using word-postposition combination data as they are [5]. Methods based on a vocabulary pair for correction use approaches such as solution of semantic ambiguity. In a vocabulary pair-based approach, words of a vocabulary pair are considered ambiguous and if the result is the same as the original words after solving ambiguity in a statistical method, the spelling is considered correct, and if different, the spelling is considered wrong, and spelling error correction takes place [6].

3 Methods

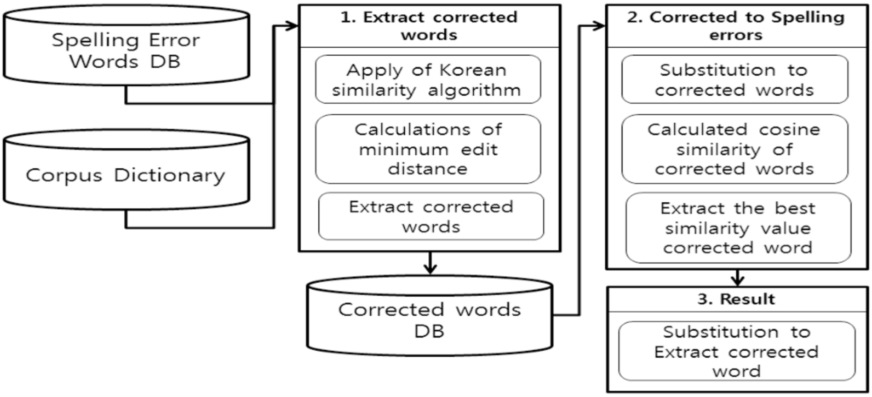

Spelling Error Words DB
Corpus Dictionary
1. Extract corrected words
Apply of Korean similarity algorithm
Calculations of minimum edit distance
Extract corrected words
Corrected words DB
2. Corrected to Spelling errors
Substitution to corrected words
Calculated cosine similarity of corrected words
Extract the best similarity value corrected word
3. Result
Substitution to Extract corrected word

Fig. 1. Korean Spelling Error Correction Process

In this section, Korean alphabet similarity algorithm-based spelling error correction is presented. First of all, it is necessary to detect spelling errors in a given document. To detect spelling errors, spelling error correction data used in previous spelling error detection research using cosine similarity-based spelling error detection algorithm [7] were hired for spelling error correction study. In this paper, Korean alphabet similarity algorithm was used to extract words similar to misspelled words from a corpus and correct spelling errors. This corpus was built in previous research for spelling error detection. Upon misspelled words replaced with extracted correct words, spelling error detection algorithm was operated to calculate cosine similarity with words in the sentence. Considering that previous research regarded words with the lowest cosine similarity as misspelled, words with the highest cosine similarity were considered as corrected right in contrast, and misspelled words were replaced with words with the highest cosine similarity to complete spelling error correction. “Fig. 1” illustrates a schematic of spelling error correction system using Korean alphabet similarity algorithm.

3.1 Korean alphabet similarity algorithm

Korean alphabet similarity algorithm is an algorithm to measure the edit distance between two Korean words. Edit distance means the minimum number of operations like insertion, deletion, and replacement of syllables or phonemes needed to change word A into word B, and shorter edit distances between two words result in higher similarity between two words. In [8], SylED (syllable-based edit distance) and PhoED (phoneme-based edit distance) algorithms were proposed to measure edit distance between two Korean words. In SylED, operation unit for edit distance is measured on the basis of syllables; edit distance of two words is calculated by comparing syllables in the same order. If edit distance δ(A, B) of words A and B in “Table 1” is calculated by SylED algorithm, '사' as the first syllable of word A needs to be replaced with

'머' as the first syllable of word B, and as the second syllable '슴' is identical in both words A and B and not replaced, making the value of δ(A, B) become '1.'

Table 1. Illustrated words for Korean similarity Algorithm.

A	사	슴
B	머	슴
C	가	슴

PhoED refers to a mechanism where operation unit to calculate edit distance of words is measured on a phoneme basis. Going beyond the SylED algorithm, this mechanism calculates edit distance by subdividing syllables in the same order into initial sound, vowel and final consonant as phonemes. Take the calculation of the edit distance δ (A, C) between words A and C in Table 1 with PhoED algorithm for example. Word A's first syllable '사' is compared and matched with '가' as the first syllable of word C, and 'ㅅ' as the initial sound of '사' needs to be replaced with 'ㄱ' as the initial sound of '가.' In addition, the second syllable '슴' is identical in both words A and C and therefore not replaced, resulting in δ(A, C) of '1.' Furthermore, to calculate δ(A, B) with PhoED algorithm, '사' as the first syllable of word A is compared and matched with '머' as the first syllable of word B, and 'ㅅ' as the initial sound of '사' is replaced with 'ㅁ' as the initial sound of '머' and vowel 'ㅏ' with vowel 'ㅓ', respectively, producing δ(A, B) of '2.'

3.2 Correct word extraction

For spelling correction purposes, words similar to misspelled words are extracted as correct words from a corpus with Korean alphabet similarity algorithm, considering the reports from previous literatures that normal words and misspelled words are similar in form. The similarity between them is found in statistical data provided in [9], which shows that spelling errors detected in the collected documents constitute around 0.5% only. This result indicates that in general, detected spelling errors are small in number because writers corrected spelling errors directly upon happening and detected spelling errors, however, are similar to the original words, making them hard to be noticed by the writer. In fact, statistical data reveal that a large number of detected spelling errors happened as consonants and vowels contained in a syllable of words are replaced. Given that normal words and misspelled words are similar in form as discussed in [9], this study extracted words similar to misspelled words in form from a corpus.

3.3 Spelling error correction

Finally, investigation is made to examine whether correct words exist among corrected words extracted. As previous studies worked on how to detect spelling errors in sentences with cosine similarity, this study used cosine similarity to detect correct words as well. Resorting to findings from previous studies, words with the highest cosine similarity among those having the minimum cosine similarity of '0.5' were considered as correct words. Equation 1 indicates how to produce cosine similarity used for spelling error detection research.

$$\mathrm{SIM}(t_i, t_j) = \frac{\sum_{k=1(k\neq i,j)}^{n} f_{ik} \times f_{jk}}{\sqrt{\sum_{k=1(k\neq i,j)}^{n} f_{ik}^2} \times \sqrt{\sum_{k=1(k\neq i,j)}^{n} f_{jk}^2}} \tag{1}$$

Using Equation (1), pairs of words in sentences are created to calculate cosine similarity based on the frequency of simultaneous appearance with other words than in such pairs, from which spelling errors were detected. Cosine similarity values range from 0 to 1, and as cosine similarity approaches 1, two words' simultaneous appearance is more likely, while as it comes closer to 0, the two words are more likely to be independent. As a consequence, lower cosine similarity is considered to lead to higher probability of spelling errors, so spelling errors with lower cosine similarity were detected. Conversely, considering that as cosine similarity comes closer to 1, simultaneous appearance of two words is more likely, higher computational results may indicate higher probability of correct words. In order to calculate cosine similarity of corrected words, misspelled words in the sentence are replaced with corrected words. Words with the highest cosine similarity among corrected words whose cosine similarity values are '0.5' or higher are regarded as correct words, and misspelled words were replaced with these words to conclude the spelling error correction process.

3.4 Conclusions and Directions for

Table 2. Cosine similarity of corrected words.

	Spelling Error Words	Extract Corrected Words	Cosine Similarity Values
1	사임	상임	0.73
		자임	0.13
		사이	0.15
2	호보	후보	0.71
		호소	0.20

		호봉	0.07
3	사울	서울	0.62
		사물	0.02
4	사랑	사람	0.58
		자랑	0.32

With the aim of exploring Korean alphabet spelling error correction method, this study employed spelling error data used in previous spelling error detection research. In methodology, over 40 news articles were collected, and misspelled words were inserted into 71 sentences of the data to examine how cosine similarity of normal words and misspelled words is calculated. When it came to spelling error, such insertion was made in the manner of replacing existing normal words with misspelled words, and 11 inserted words including '사임 (상임)', '사울 (서울)', '사랑 (사람) ', and '후보 (호보)' were replaced. The corpus was built by extracting words from the news article data, before replacing normal words with misspelled words. With Korean alphabet similarity algorithm, words similar to misspelled words in form were extracted from the corpus. In order to extract corrected words, the minimum edit distance between words in the corpus and misspelled words were measured with Korean alphabet similarity algorithm first, and words within the resulting edit distance were extracted only. Since the replaced misspelled words with edit distance of 1 were inserted only, the edit distance of extracted corrected words was 1 altogether. To select the right corrected words from the extracted corrected words, the spelling error detection algorithm applied once again to calculate cosine similarity. “Table 2” provides some of the cosine similarity values calculated by replacing extracted corrected words with misspelled words.

According to results of cosine similarity computation, cosine similarity values all of the normal words among corrected words are at the highest. Among them, ‘saram (사람)’ was found to have the lowest value at ‘0.58’, while ‘sangim (상임)’ having the highest at ‘0.73.’ Besides, average cosine similarity of extracted normal words was ‘0.67.’

Table 3. Result.

	All	Proper Calibration	Wrong Calibration
Value	71	69	2
Percent	100%	97%	3%

As a result of the spelling error correction process, a total of 69 sentences out of 71 were successfully corrected. The word contained in the two sentences which failed to be corrected in spelling was 'sarang (사랑)'; in one sentence, 'sarang (사랑)' was replaced with a wrong word 'jarang (자랑)', and the word remained not corrected in the other sentence.

3.5 Conclusions and Directions for Subsequent Studies

This study explored spelling error correction methods drawing on findings from previous research. To correct spelling errors, Korean alphabet similarity algorithm was employed to extract words similar to misspelled words in form from a corpus, and in order to select the right corrected words from the extracted words, cosine similarity algorithm used for spelling error detection was hired to extract the right corrected words. The misspelled words were those artificially inserted in the previous study for spelling error detection, and all the spelling errors inserted in this spelling error correction study were corrected. As such spelling errors were artificially inserted, although considered as a general method, and corrected words extracted with the Korean alphabet similarity algorithm numbered 2 or 3, more data is needed to gain better results. Subsequent studies will work on more data than used in this study and use data with spelling errors inserted naturally. Further studies on Korean spelling errors are considered necessary because unlike other languages, Korean words are constructed in a complex structure.

ACKNOWLEDGMENT

The This research was supported by SW Master's course of hiring contract Program grant funded by the Ministry of Science, ICT and Future Planning(H0116-16-1013) and the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) (No. NRF-2016R1A2B4012638).

References

1. 최철, 박세진, 김철중, 권규식, “Analysis of Uncorrected Typing Rate of Keyboard Design Ergonomic Keyboard Based on Qwerty Keyboard” EEromonomics Society of Korea, vol. 2000-1 no.-, pp.142-145
2. Hyunsoo Choi, Hyukchul Kwon, Aesun Yoon. “Improving Recall for Context-Sensitive Spelling Correction Rules using Conditional Probability Model with Dynamic Window Sizes” Journal of KIISE, vol.42 no.5, 2015, pp.629-636
3. Jingzhi Jin, Sungki Chio, Hyuk-chul Kwon. “Adaptive Context-Sensitive Spelling Error Correction Techniques for The Extremely Unpredictable Error Generating Language Environments”, Korea Informatio
4. Hyunsoo Choi, Aesun Yoon, Hyuk-Chul Kwom, “Improving Recall for Context-Sensitive Spelling Correction Rules Using Integrated Method”, Korea Infomation Science Society, vol.2014 no.6, 215, pp.577-579
5. Minho Kim, Hyuk-Chul Kwon, Sungki Choi. “Context-sensitive Spelling Error Correction using Eojeol N-gram”, Journal of KIISE vol.41 no.12, 2014. pp.1081-1089
6. Minho Kim, Jingzhi Jin, Hyuk-Chul Kwon, “Statistical Context-sensitive Spelling Correction using Confusion Set”, Korea Infomaton Science Society, vol.2013 no.6, 2013, pp.607-609
7. SeungHyeon Bak, JunSeok Cha, TaekEun Hong, JuHyun Shin, PanKoo Kim, “Korean Spelling Error Detection Method for Research”, Spring Conference of KISM 2016, vol.5 no.2, 2016
8. Kangho Roh, Jin Wook Kim, Eunsang Kim, Kunsoo Park, Hwan-Gue Cho. “Edit Distance Problem for the Korean Alphabet”, Journal of KIISE : Computer Systems and Theory, vol.31 no.2, 2010. pp.103-109
9. Hankyu Lim, Ungmo Kim. “A Spelling Correction System Based on Statistcal Data of Spelling Errors”, The KIPS Transactionsty, vol.2 no.6, 1995. pp.839-846

A Comprehensive Security Analysis Checksheet for OpenFlow Networks

Yoshiaki Hori[1,2], Seiichiro Mizoguchi[3], Ryosuke Miyazaki[2,4], Akira Yamada[3], Yaokai Feng[2,4], Ayumu Kubota[3], and Kouichi Sakurai[2,4]

[1] Organization for General Education, Saga University,
1 Honjo, Saga 840-8502, Japan
[2] Institute of Systems, Information Technologies and Nanotechnologies,
2-1-22 Momochihama, Sawara-ku, Fukuoka 814-0001, Japan
[3] KDDI R&D Laboratories, Inc.,
2-1-15 Ohara, Fujimino, Saitama 356-8502, Japan
[4] Faculty of Information Science and Electrical Engineering, Kyushu University,
744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan

Abstract. Software-defined networking (SDN) enables the flexible and dynamic configuration of a network, and OpenFlow is one practical SDN implementation. Although it has been widely deployed in actual environments, it can cause fatal flaws. In this paper, we consolidate the security threats to OpenFlow mentioned in previous work and introduce a new security checksheet that includes risk assessment methods. We compare the Kreutz et al. threat vectors with the SDNSecurity.org attack list to discover new threats. Our checksheet enables the security of a given OpenFlow network design to be comprehensively assessed. Furthermore, we evaluate the performance of an OpenFlow network with two attack scenarios using the checksheet and identify critical performance degradations.

Keywords: SDN, OpenFlow, system security, risk assessment

1 Introduction

Software-defined networking (SDN) is an emerging networking paradigm that is a good candidate for relieving the limitations of the current network infrastructures [1][2]. By separating the control logic (the *Control-Plane*, referred to as the *C-Plane* hereafter) of the network from data packet forwarding mechanisms (the *Data-Plane*, referred to as the *D-Plane* hereafter) such as traditional routers and switches, it enables dynamic and flexible configurations of the network in order to properly allocate network resources. When building a network infrastructure, considering the security of a network for social infrastructure to reduce its security risk is mandatory. When building a network using an SDN, the security of the SDN is one of the requirements of its system design. An SDN tends to be more complicated than traditional non-SDN networks because it consists of many components and their interfaces. Therefore, building a secure

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_22

SDN is a mandatory challenge for future various network infrastructures, from a campus network to a carrier's backbone network. We focus on the OpenFlow [3] network, which is one implementation of an SDN. OpenFlow has interface protocols between the C-Plane and D-Plane that are widely used in actual network environments and will be deployed in the future. In this paper, we deal with the threats to the OpenFlow network and their countermeasures. We classify the security threats of the OpenFlow network and make clear its security risks. Furthermore, we discuss a method for risk assessment and countermeasures for every security risk. We devise a security checksheet for the security of the SDN system. We believe our SDN security checksheet is useful for designing a secure SDN network. The contributions of this paper are as follows:

- We classify the security threats of the OpenFlow network system by consolidating the Kreutz et al. threat vectors and the SDNSecurity.org attack list, and we introduce some new significant risk items to complete our security threat list.
- We create a security checksheet that includes practical assessment methods for risks and their countermeasures. This security checksheet is useful for the risk assessment of an OpenFlow network system design and its operation.
- We evaluate two DoS (Denial of Service) risk scenarios that are included our proposed security checksheet with a given actual OpenFlow network testbed consisting of commercial OpenFlow switches and an open source OpenFlow controller implementation. As a result, we obtain quantitative conditions for the risk.

2 Organizing SDN Security Threats

In 2003, the National Institute of Standards and Technology (NIST) originally published the "Guideline on Network Security Testing" (NIST SP800-42) [4] as a guideline for security when constructing a network. In 2008, NIST also published the "Technical Guide to Information Security Testing and Assessment" (NIST SP800-115) [5], updating NIST SP800-42. Although these documents mention network security, they do not consider an OpenFlow network system. There are some existing studies that analyze the security of SDN. For example, Shin et al. presented an early discussion about attacks on SDN [9]. They briefly mention the C-Plane's resource consumption or DoS attacks, and D-Plane's resource consumption or DoS attacks. Kilöti et al. performed a security analysis of OpenFlow using STRIDE and an attack tree approach [10]. They focused on a Data Flow Diagram of the OpenFlow protocol, which does not include OpenFlow applications or the system environment. Hayward et al. recently presented a survey on security in SDN [11]. They summarized several security analysis studies. However, their work focused on specific layers and interfaces and did not provide a comprehensive security analysis. Kreutz et al. [6] and SDNSecurity.org [7] separately summarized OpenFlow's security threats in 2014 and 2015, respectively, but they do not provide assessment methods and countermeasures for a given

OpenFlow network. We consolidate the security threats of an OpenFlow network system by comparing the Kreutz et al. threat vectors and the SDNSecurity.org attack list, and we introduce some new significant risk items to create our final security threat list.

2.1 Seven Threat Vectors of Kreutz et al.

Kreutz et al. pointed out the seven main potential threat vectors in SDN [6], which are as follows: **Threat vector 1:** forged or faked traffic **Threat vector 2:** attacks on vulnerabilities in switches **Threat vector 3:** attacks on C-Plane communications **Threat vector 4:** attacks on and vulnerabilities in controllers **Threat vector 5:** lack of mechanisms to ensure trust between the controller and management applications **Threat vector 6:** attacks on and vulnerabilities in administrative stations **Threat vector 7:** lack of trusted resources for forensics and remediation

They state that threat vectors 3, 4, and 5 are specific to SDN, as they stem from the separation of the C-Plane and D-Plane, and the others are not specific. In addition, they proposed nine solutions for making control platforms dependable and secure against their threat vectors [6]: replication, diversity, self-healing mechanisms, dynamic device association, trust between devices and controllers, trust between application and controller software, security domains, secure components, and fast and reliable software update and patching. They proposed a general design for a secure and dependable control platform. However, a detailed assessment is required for the actual security design of a given SDN network.

2.2 SDNSecurity.org SDN Threat Analysis

The Network and System Security Laboratory of KAIST analyzed the threats to SDN architecture and created an “attack list” for SDN [7]. They categorized the components of an SDN by whether they reside in the Application Layer, Control Layer, Infrastructure Layer, or the Control Channel between the Control Layer and Infrastructure Layer. They then pointed out security threats for every SDN component. Figure 1 shows their list of security threats. For instance, one item on the attack list, **[A-1] packet-in flooding**, is a threat to network operating systems. These details were posted on the SDNSecurity.org site in the summer of 2015. However, this site was only partially online as of May 2016, and the attack list is no longer available.

2.3 Reported Vulnerabilities of OpenFlow

Benton et al. provided a brief overview of the vulnerabilities present in the OpenFlow protocol [8]. They highlighted the classes of vulnerabilities that emerge from the separation and centralization of the protocol plane in OpenFlow network designs. They discuss Man-in-the-middle Attacks, Listener Mode, Switch Authentication, Flow Table Verification, DoS Risks, and Controller Vulnerabilities. However, they discuss them only briefly. For OpenFlow network design

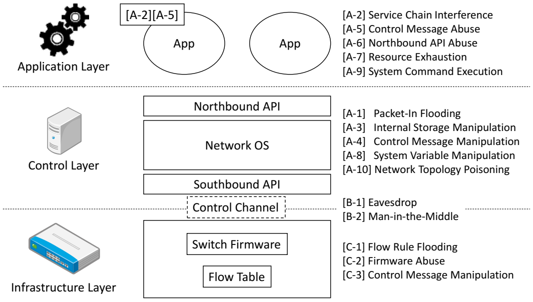

Application Layer
[A-2][A-5]
App
App
[A-2] Service Chain Interference
[A-5] Control Message Abuse
[A-6] Northbound API Abuse
[A-7] Resource Exhaustion
[A-9] System Command Execution
Control Layer
Northbound API
Network OS
Southbound API
[A-1] Packet-In Flooding
[A-3] Internal Storage Manipulation
[A-4] Control Message Manipulation
[A-8] System Variable Manipulation
[A-10] Network Topology Poisoning
Control Channel
[B-1] Eavesdrop
[B-2] Man-in-the-Middle
Infrastructure Layer
Switch Firmware
Flow Table
[C-1] Flow Rule Flooding
[C-2] Firmware Abuse
[C-3] Control Message Manipulation

Fig. 1. Threat analysis of SDNSecurity.org. The authors drew this figure based on [7].

and operation, it is important to organize the details of the OpenFlow network system vulnerabilities and discuss them.

3 Our Proposal

In order to improve the security of a given SDN system, it is important to prepare a checksheet for risk assessment. Before we provide the checksheet, we list the security threats against SDN systems.

3.1 OpenFlow Network System Security Threat List

To list the threats against SDN, we refer to the comprehensive survey by Kreutz et al. [6] and the vulnerability list by SDNSecurity.org [7]. Table 1 lists these threats.

The "Category" column represents the objects that would be damaged by the threats. There are three categories: D-Plane, C-Plane, and Others. The D-Plane includes the data path and switches, and the C-Plane includes the southbound API (Application Programming Interface), controller itself, northbound API, and applications. The Others category consists of the systems that operate administrative stations, forensics, or remediation.

Next, by referring to and supplementing the vulnerability list of SDNSecurity.org, we define additional SDN security threats as follows:

Switch Table Manipulation: If an SDN switch has a forwarding table, adversaries could try to manipulate this table to redirect traffic to invalid destinations. If a controller has such a table and synchronizes the switches under the controller, this controller can also be a target of switch table manipulation.

Firmware Manipulation: An SDN switch stores its firmware image in memory. Adversaries could try to manipulate this image in order to inject malware functions so as to start the malware at every boot instance. If this firmware is stored in other components, such as a controller or an administrative station, these components could also be targets.

Vulnerability Exploitation of the Switch Program: If the firmware of a switch has a vulnerability that is not yet publicly known, a zero-day attack against the switch is possible. Because we cannot prevent zero-day attacks in general, we have to construct a security incident response team (CSIRT) to monitor the vulnerability information and create an incident response manual.

Vulnerability Exploitation of the Controller Program: This is the same as the vulnerability of the switch program, and a CSIRT must also be organized for its response.

In addition, we assign an ID number to each threat. The "Basic Mechanism" column shows how these threats are launched by adversaries. This information is used for the next risk assessment step.

Table 1. OpenFlow System Security Threats

ID	Vulnerability Check Items	Category	Basic Attack Mechanisms	Threat Vector by Kreutz[]	Vulnerability Genome Project by SDNSecurity.org[]	Our Original
1	Forged or Fake Traffic Flows in Data Plane	D-Plane	Adversaries send forged packets to data plane from the outside of the SDN or from local network.	Threat Vector 1: Forged or Fake Traffic Flows		
2	Firmware Abuse		Adversaries intrude control plane and login to switches.	Threat Vector 2: Vulnerabilities of Forwarding Devices	C-2: Firmware Abuse	
3	Packet_IN Flooding (Switch)		Based on ID 1, adversaries intentionally raise Packet_IN events.		A-1:Packet_IN Flooding	
4	Flow Rule Flooding		Adversaries intrude control plane and issue flow rule configurations.		C-1:Flow Rule Flooding	
5	Control Message Manipulation		Adversaries intrude control plane and send fake control messages.		C-3:Control Message Manipulation	
6	Switch Table Manipulation		Adversaries login switches and manipulate its switch table, or if controllers have switch tables, adversaries login these controllers and manipulate switch table database.			✓
7	Firmware Manipulation		Adversaries login switches and manipulate its firmware images, or manipulate firmware on the management stations.			✓
8	Vulnerability Exploitation of Firmware (Switch)		Adversaries exploit unknown or known vulnerabilities in switch firmware.			✓
9	Packet_IN Flooding (Southbound)	C-Plane	Based on ID 1, adversaries try to waste bandwidth between switches and controller.	Threat Vector 3: Compromise Southbound API	A-1:Packet_IN Flooding	
10	Eavesdrop		Adversaries intrude control plane and eavesdrop messages.		B-1:Eavesdrop	
11	Man-In-The-Middle		Adversaries highjack southbound or northbound to eavesdrop or manipulate messages.		B-2:Man-In-The-Middle	
12	Control Message Manipulation		Adversaries intrude data plane and send forged control messages to controllers.	Threat Vector 4: Compromise Controllers	A-4:Control Message Manipulation	
13	Packet_IN Flooding (Controller)		Based on ID 1, adversaries try to waste computational resources on controllers.		A-1:Packet_IN Flooding	
14	Vulnerability Exploitation of Firmware (Controller)		Adversaries exploit unknown or known vulnerabilities in controller program.			✓
15	Internal Storage Manipulation		Adversaries login the controller and manipulate storage.		A-3:Internal Storage Manipulation	
16	System Variable Manipulation		Adversaries login the controller and change system variables.		A-8:System Variable Manipulation	
17	System Command Execution		Adversaries login the controller and issue system commands.		A-9:System Command Execution	
18	Network Topology Poisoning		Based on ID1 or just login the controller, adversaries manipulate topology database.		A-10:Network Topology Poisoning	
19	Service Chain Interference		Adversaries exploit service chain logic to interfere service chain.	Threat Vector 5: Compromise Northbound API and Applications	A-2:Service Chain Interference	
20	Control Message Abuse		Adversaries abuse northbound API or application to issue invalid control messages.		A-5:Control Message Abuse	
21	Northbound API Abuse		Adversaries at the application layer abuse northbound API to damage controllers and applications.		A-6:Northbound API Abuse	
22	Resource Exhaustion		Based on ID 20-21, adversaries try to waste resources for controllers and applications.		A-7:Resource Exhaustion	
23	Vulnerabilities in Administrative Station	Others	Adversaries exploit the vulnerabilities of administrative stations to launch another attacks.	Threat Vector 6: Vulnerabilities in administrative stations		
24	The Lack of Trusted Resources for Forensics and Remediation		Adversaries intrude control plane and damage to forensics system and data.	Threat Vector 7: The lack of trusted resources for forensics and remediation		
25	The Lack of Trusted Operations for Forensics and Remediation		Adversaries exploit remediation logic and damage to remediation process.			✓

The first column of Table 1 shows the threat category. The second and third columns show the threat vectors from Kreutz et al. [6] and SDNSecurity.org [7], respectively. For the actual network design, we analyze the Kreutz et al. threat vectors to determine finer threats and carry out a risk analysis for each finer threat and determine its countermeasure. The SDNSecurity.org work can also be classified into finer threats. However, we should add some switch and controller related threats: Switch Table Manipulation, (Switch) Firmware Manipulation, Vulnerability of the Switch Firmware in the D-plane, and Controller Vulnerability Exploitation in the C-Plane. We add these security threats in the fourth column.

3.2 OpenFlow Network System Security Assessment Checksheet

Using our proposed table (Table 1), we created a security checksheet for the OpenFlow network system that consists of security risk assessment items and candidate countermeasures. This security checksheet is useful for risk analysis during OpenFlow network system design and operation. This checksheet makes it easy for a network designer or operator to determine risk items and countermeasures for reducing related risks. Table 2 shows the proposed OpenFlow Network System Security Checksheet.

We created the SDN security assessment checksheet (Table 2) based on the threat list shown in Table 1. The contents of each column are explained in detail in the following list.

ID: This column lists the sequence number.

Category: This column includes the D-Plane, C-Plane, or Others categories.

Condition: Using the basic mechanisms shown in Table 1, this column represents the condition under which the threats occur. For example, in order to login to a switch, adversaries must be able to access its management port. If the switch does not have such a management port or there is no path to the port, this threat may not occur. When the network administrator conducts a security risk assessment, this information is useful for selecting the items for risk analysis and countermeasures.

Risk: This column represents the damage against the system when the threat occurs. In order to determine appropriate countermeasures, this information is useful.

Evaluation Points: If the system design meets the attack conditions and the risk is not ignorable, the network administrator conducts an additional evaluation using the points in this column. Based on the evaluation result, the administrator can choose adequate countermeasures.

Countermeasures: This column represents the list of countermeasures, and the network administrator can select solutions from these items. This list should be updated periodically.

Table 2. OpenFlow Network System Security Checksheet

ID	Category	Condition	Risk	Evaluation Points	Countermeasures
1	D-Plane	Adversaries can send packets to data plane of switch.	Waste data plane bandwidth, launch Packet_IN flooding, then service down.	Evaluation with packet generator.	Use of IDS/IPS
2		Adversaries can access management port of switches.	Lead to several risks.	- Check user manual of switches. - Check logging function. - Check intrusion detection function.	- Login Password Management - Logging - Use syslog-based IDS
3		Adversaries can send packets to data plane of switch.	Switch down	- Evaluation with packet generator. - Check monitoring function of abnormal Packet_IN behavior.	Anomaly detection against Packet_IN messages
4		Adversaries can access southbound or controller.	Switch down or flow table disruption.	- Evaluation with flow rule generator - Check monitoring function of abnormal flow rule insertion - Check authentication for flow rule insertion	Anomaly detection against flow rule insertion
5		Adversaries can access southbound or controller.	Switch anomaly	- Evaluation of arbitrary control message insertion - Check message authentication function	Message Authentication
6		Adversaries can login switches or controller.	Flow redirection	- Check switch table integrity check function - Check authentication of switch table manipulation	- Memory Protection - Software Attestation
7		Adversaries can login switches or firmware at the control plane.	Switch untrusted	Check firmware image integrity check function	Secure Boot
8		Adversaries can access from data plane or control plane.	Lead to several risks.	- Check firmware update function - Check ISAC	Firmware Update
9	C-Plane	Adversaries can send packets to data plane of switch.	Waste control plane bandwidth.	- Evaluation with packet generator - Check monitoring function of abnormal Packet_IN behavior	- Anomaly detection against Packet_IN messages at controller - Resource monitor
10		Adversaries can access control plane.	Disclosure of user data.	Check C-Plane confidentiality	C-Plane encryption
11		Adversaries can access control plane.	Disclosure of user data or highjack of controller.	Check authentication between switches and controllers.	Authentication
12		Adversaries can login switches or access southbound.	Highjack of controller.	Check message authentication between switches and controllers.	Message Authentication
13		Adversaries can send packets to data plane of switch.	Waste controller resources.	- Evaluation with packet generator - Check monitoring function of abnormal Packet_IN behavior	- Anomaly detection against Packet_IN messages at controller - Resource monitor
14		Adversaries can access controller.	Lead to several risks.	- Check firmware update function - Check ISAC	Firmware Update
15		Adversaries can login the controller.	Disclosure, manipulation, destruction of data.	Check confidentiality of data store in controllers	- Encryption - Access Control
16		Adversaries can login the controller.	System unstable.	Check integrity check function for system variables	- Memory Protection - Access Control - Logging - Secure Boot
17		Adversaries can login the controller.	Lead to several risks.	- Check system command log - Check anomaly detection function	- Access Control - Logging - Anomaly detection
18		Adversaries can send packets to data plane, or login switches, or login controller.	Hide network anomaly or flow redirection, denial of service.	Check network topology integrity check function	- Topology Database Monitoring - Access Control
19		Adversaries can access data plane, switch, southbound or controller.	Denial of network services.	- Check application behavior logging function - Check application anomaly detection function	- Anomaly Detection of application - Access Control
20		Adversaries can access controller, northbound, application.	Application anomaly or denial of network services.	Check the integrity check function of flow tables and policies.	- Anomaly Detection of application - Access Control
21		Adversaries can access northbound, application.	Block the other application's operation.	- Check Northbound API usage logging function - Check anomaly detection function for Northbound API	Logging and Monitoring Northbound API call
22		Adversaries can access controller, northbound, application.	Application resource exhaustion.	Check application resource monitoring function	Resource monitoring and anomaly detection
23	Others	Adversaries can access administrative station.	Lead to several risks.	Check the behavior logging and monitoring function of administrative stations	- Access Control Logging - Anomaly detection
24		Adversaries can access forensicsremediation system.	Erase attack logs.	Check the confidentiality and integrity of logs	- Encryption - Access Control
25		Adversaries can access forensicsremediation system, and controller.	Drop remediation or backuped firmware and configuration manipulation.	- Check the integrity of config and firmware image - Check the periodic backup functions	- Encryption - Access Control - Periodical Updates

4 Use of OpenFlow Network System Security Assessment

In this section, we evaluate two DoS risk scenarios in an actual OpenFlow network test-bed with typical commercial OpenFlow switches and an open source OpenFlow controller implementation.

4.1 Out SDN/OpenFlow Testbed and Security Assessment

We created an OpenFlow network evaluation environment using the Ryu3.24 OpenFlow controller software and Pica8 P-3297 OpenFlow switch. We evaluated our OpenFlow testbed using our proposed assessment checksheet, which gives qualitative security assessment results. However, quantitative results are desirable for actual network operation. Therefore, we evaluated our OpenFlow network testbed under two DoS scenarios to obtain quantitative results.

4.2 Quantitative Evaluation of DoS Scenario 1 (PACKET_IN Flooding)

In this scenario, we assume that adversaries intentionally send packets that raise vast numbers of PACKET_IN messages to the controller. This results in PACKET_IN messages flooding the controller. The evaluation of this attack can be replaced with an evaluation of the OpenFlow controller, which does not update the flow table.

Experiment Environment For the quantitative evaluation of PACKET_IN flooding, we used our testbed. Figure 2 shows the testbed environment. The OpenFlow switch is connected to hosts A, B, and C. The Ryu OpenFlow controller runs on a VM (virtual machine, consisting of four virtual core CPUs, 4 GB RAM, Ubuntu 14.04 LTS). The VM runs on the physical host machine (Intel Core i7 860 2.8 GHz, 16 GB RAM, Ubuntu14.04 LTS).

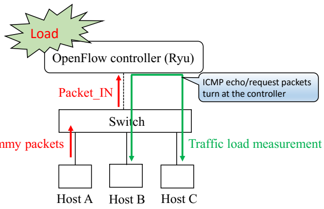


Fig. 2. Environment of PACKET_IN flooding experiment

Experiment We used host A for sending dummy packets, and hosts B and C for measuring the packet's arrival rate. We used packETH for dummy packet generation. Dummy packets were sent at predetermined intervals. When the OpenFlow controller receives a Packet_IN message, it raises a Packet_OUT message; however, it does not update the flow table of the OpenFlow switch. This means that every time the OpenFlow switch receives a packet, a Packet_IN arises from the OpenFlow switch to the OpenFlow controller. However, in order to prevent dummy packets between hosts B and C from affecting the evaluation, the controller discards dummy packets after their Packet_IN is received. Packets between hosts B and C go through the controller. By probing packets' behavior between hosts B and C, we can determine the controller's load during Packet_IN flooding.

We measured the round-trip time (RTT) between hosts B and C when host A sent dummy packets. At the same time, we calculated the packet arrival rate of the number of ICMP echo-reply packets that successfully arrived at host C and compared it to the number of ICMP echo-reply packets that were actually sent from host B.

Results Figure 3 shows the results of the evaluation experiment. The RTT between hosts B and C was measured by the `ping` command. Reachability represents the ICMP packet arrival rate. When the input of dummy packets is less than or equal to 1,000 pps (packets per second), RTT shows no increase and reachability stays at 100%. However, when the input of dummy packets is more than 2,000 pps, RTT starts increasing, and when the input of dummy packets is more than 6,000 pps, reachability falls below 50%. As the pps further increases, RTT increases rapidly and reachability decreases.

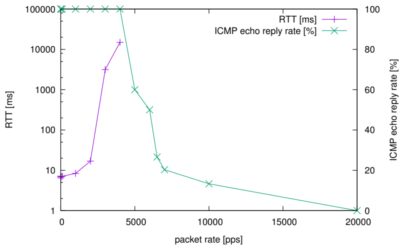


Fig. 3. PACKET_IN flooding experiment using Ryu

Discussion The result of this experiment appears to show that, at most, a rate of Packet_IN messages on the order of thousands of pps causes a serious performance decrease in the OpenFlow controller. This seems to be a result of the limit of the OpenFlow controller's processing ability when running on the host machine. The bandwidth between the switch and the controller may also cause the performance to decrease. If the processing ability of the OpenFlow controller is not sufficiently high, there may be a sudden decrease in performance during DoS attacks.

4.3 Quantitative Evaluation of DoS Scenario 2 (FlowRule Flooding

In this scenario, we assume adversaries intentionally send packets that send vast numbers of Flow_Mod messages to the controller. This results in FlowRule flooding on the controller. The evaluation of this attack can be replaced with an evaluation of an OpenFlow controller that raises a Flow_Mod for every new packet.

Experiment Environment For the quantitative evaluation of FlowRule flooding, we used the testbed. Figure 4 shows the testbed environment. Considering that the OpenFlow system has the ability to run on various kinds of machine, we used a different machine from the one used for the Packet_IN flooding experiment. This machine has a lower performance.

Here, the OpenFlow controller runs on a Raspberry Pi2 and controls an OpenFlow switch. The switch is connected to hosts A, B, and C. We used host A for sending dummy packets at a rate of around 4,000 pps, and hosts B and C for measuring the packet arrival rate. Each dummy packet has a different IP address so that the controller raises a Packet_IN and a flow rule is added to the flow table every time a dummy packet comes to the switch. This causes vast number of flow rules to be added to the flow table of the switch, which could result in flow table overflow. We also set some flow rules in advance on the switch in order to enable communication between hosts B and C. A packet from host B should to be sent toward host C, and a packet from host C should be sent toward host B. That is, all packets between hosts B and C are dealt with within the switch and should never raise a Packet_IN message.

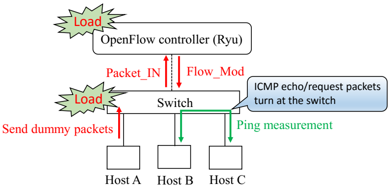


Fig. 4. PACKET_IN flooding evaluation experiment environment

Experiment In this environment, we ran the `ping` command to investigate the effect of flow rule flooding. When an ICMP echo request packet from host B arrives at the switch, the switch looks up its flow table. If a flow rule stating that a packet from host B is supposed to be sent toward host C is already on the flow table, then an ICMP echo request packet is sent toward host C without raising a Packet_IN message. An ICMP echo reply packet from host C is sent to host B in the same way. We sent 50 ICMP echo packets in total. Additionally, we measured the bandwidth, one-way delay time, and packet drop rate between hosts B and C using iperf. We used both TCP and UDP modes for measurement. Note that the maximum UDP bandwidth is limited to 100 Mbps because of the performance limitation of the network interface cards of hosts B and C.

Results Table 3 shows the `ping` evaluation result compared with the values when the controller works as a normal repeater hub. Overall, the RTT of flow rule flooding was higher than that of normal operation. Specifically, the mdev (standard deviation) was higher than normal, which meant there was great variability of the RTT when flow rules flooded the switch. Reachability was lower than the normal value, and 30% of the ICMP packets were dropped when flow rules flooded the switch.

Table 4 shows the iperf evaluation result compared with normal operation values. All values were worse than normal. Specifically, UDP packets dropped by 75%.

Table 3. RTT and reachability of flow rule flooding experiment

		Flow rule flooding	Normal
RTT[ms]	Min	2.882	0.477
	Max	92.476	0.603
	Average	22.409	0.547
	Mdev	21.215	0.04
Reachability[%]		70	100

Table 4. iperf result of the flow rule flooding experiment

		Flow rule flooding	Normal
TCP	Bandwidth [Mbps]	1.04	146
UDP	Bandwidth [Mbps]	24.6	101
	One-way delay [ms]	26.573	0.013
	Reachability [%]	25	99.9

Discussion The results of this experiment appears to show that, at most, a Flow_Mod rate on the order of thousands of pps causes a serious performance decrease in the OpenFlow switch. Vast numbers of Flow_Mod messages may consume the CPU resources of the OpenFlow switch, which may result in an increase of the packet drop rate. We should investigate the data transfer mechanism during Flow_Mod, and, for a secure OpenFlow system, we should design a controller that detects abnormal numbers of Flow_Mod messages.

5 Concluding Remarks

This paper addressed the security threats of OpenFlow network systems and their countermeasures. We classified security threats of the OpenFlow network and clarified its security risks. Furthermore, we discussed a method for the risk assessment and countermeasures for every security risk. We devised a security checksheet for SDN system security. We believe our SDN security checksheet is useful for designing a secure SDN network. In addition, we reported the results of two quantitative evaluation experiments using our OpenFlow testbed.

As future work, we will continuously revise the proposed checksheet to include new threats. In addition, we plan to create decision rules to adopt one or more proper countermeasures for each security threat of the OpenFlow network system.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number 16K00132.

References

1. Y. Jarraya et al.: A Survey and a Layered Taxonomy of Software-Defined Networking, In IEEE Comm. Surveys & Tutorials, Vol. 16, No. 4, pp. 1955–1980 (2014)
2. D.Kreutz et al.: Software-Defined Networking: A Comprehensive Survey. Proc. of the IEEE, Vol. 103, No. 1, pp. 14–76 (2015)
3. N.McKeown et al: OpenFlow: enabling innovation in campus networks, ACM SIGCOMM Computer Communication Review, Vol. 38, Issue 2 (2008)
4. J. Wack, M.Tracy, M. Souppaya: Guideline on Network Security Testing, NIST Special Publication 800-42 (2003)
5. K. Scarfone, M. Souppaya, A. Cody, A. Orebaugh: Technical Guide to Information Security Testing and Assessment, NIST Special Publication 800-115 (2008)
6. D.Kreutz et al: Towards Secure and Dependable Software-Defined Networks, In: ACM SIGCOMM workshop HotSDN'13, pp.55–60 (2013)
7. SDNSecurity.org: An Overview of Misuse / Attack Cases, https://web.archive.org/web/20150423094535/http://sdnsecurity.org/project_SDN-Security-Vulnerbility-attack-list.html(access 2015-12-14)
8. K. Benton, L. J. Camp, C. Small: OpenFlow vulnerability assessment, In: ACM SIGCOMM workshop HotSDN'13, pp.151–152 (2013)
9. S. Shin, G. Gu: Attacking Software-Defined Networks: A First Feasibility Study, In: ACM SIGCOMM workshop HotSDN'13, pp.165–166 (2013)
10. R. Klöti et al.: OpenFlow: A security analysis, In: 21st IEEE Int'l Conf. on Network Protocols (ICNP 2013), pp. 1-6 (2013)
11. S. Scott-Hayward et al.: A Survey of Security in Software Defined Networks, In: IEEE Comm. Surveys & Tutorials, Vol. 18, No. 1, pp. 623-654 (2016)
12. Pica8 switches, Pica8, Inc. http://www.pica8.com/products/pre-loaded-switches(access 2015-12-15)
13. PACKETH. http://packeth.sourceforge.net/packeth/Home.html(access 2015-12-15)
14. Ryu SDN Framework. http://osrg.github.io/ryu/(access 2015-12-15)

Enhanced energy conditioned mean square error algorithm for wireless sensor networks

Duha Binte Asim[1], Nadeem Javaid[2,*]
[1]Institute of Space Technology, Islamabad 44000, Pakistan
[2]COMSATS Institute of Information Technology Islamabad 44000, Pakistan

Abstract Wireless Sensor Networks (WSNs) have found numerous applications in control and monitoring fields. Advancements in the field of electronics have made wireless sensors economical enough to be widely used. WSNs have found wide applications in defence, agriculture, seismic monitoring, health sector, urban area monitoring, etc. The battery life of nodes in such networks is a constraint. Routing algorithms chosen for WSNs should make sure that energy consumption of nodes is minimized. Geographic routing is one of the options. It can be used in large scale networks owing to its low energy consumption properties. It also gives low overhead. Geographic routing comes with an inherent defect of location errors. Location errors impair the performance of geographic routing. In this paper a protocol Enhanced Energy Conditioned Mean Square Error Algorithm (E-ECMSE) is proposed that copes with the location errors of geographic routing and hence shows a fair increase in the packet delivery ratio of the network and a decrease in the energy consumption. The number of hops in the network are controlled which directly reduce the energy consumption.

1 Introduction

Geographic routing is based on known location of the devices [1]. This location can be obtained through sources like Global Positioning System (GPS) or Local Positioning System (LPS), etc. Geographic routing saves sensors the ordeal of maintaining a routing table and hence the overhead(which holds the information such as current neighbours etc). The location based characteristic of georouting protocols make them suitable for dynamic networks where there is a constraint pertaining to

Correspondence: *Corresponding author:
COMSATS Institute of Information Technology, Islamabad 44000, Pakistan.
Emails: nadeemjavaid@comsats.edu.pk, nadeemjavaidqau@gmail.com
Website: www.njavaid.com

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_23

energy, frequency and bandwidth [2].Geographic routing comes with inherent localization errors. All the methods used for gathering location information introduce a certain value of error in the location measurement [3]. These errors can effect the packet delivery ratio as well as the energy efficiency of a network [4]. In the past years a lot of research has been done on formulating a routing protocol that performs well against these errors. In this regard some major contributions were made, all of the contributions mentioned assume that the statistical error associated with each node is known. Most forward within R (MFR) is one of the first contributions made [5], where R is the transmission range. LED prefers creating an energy efficient routing path rather than concentrating on improving other metrics [6].;alongside studying the packet delivery ratio (PDR). LED uses the concept of choosing the optimal energy node as a forwarder. This considerably contributes in improving the energy efficiency. Conditioned Mean Square Error Ratio (CMSER) pays attention to improving the throughput of the network. It chooses the forwarder located nearest to the receiver, thus minimizing the hops. Probability of a successful reception is given preference and paths of lost packets are kept short. Energy Conditioned Mean Square Error (ECMSE) performs even better than CMSER, not just in terms of throughput but also energy efficiency. Along with studying scenarios similar to the ones in CMSER, in ECMSE, cases where acknowledgement is used, are also studied. LED uses the information of selecting the forwarder node on calculations like mean and error variance. These calculations are forwarded to the nodes using anchor nodes [7]. A little difference in calculations though, CMSER uses a similar approach. Both the algorithms use mean and error variance as decision metric, LED also uses an optimal energy position for the selection of best forwarder.

2 Related work

In [8] the authors proposed a scheme that discusses the types of location inaccuracies and their effects on geographic routing. There are four location inaccuracies discussed. Absolute location inaccuracy, relative distance inaccuracy,absolute location inconsistency, relative distance inconsistency. Absolute location inaccuracy and absolute location inconsistency increases packet drop in the network. Relative distance inaccuracy causes the network to choose non-optimal paths for the transmission of packets. Relative distance inconsistency increases the chances of a routing loop to be formed.

In [9] Geographic and Energy Aware Routing (GEAR) is presented. It uses both geographic routing and energy aware routing whichever is suitable. It also chooses the neighbour closest to the destination. In case there is no neighbour for the current sender, it chooses the node which has the minimum cost function while transmitting to the destination node. Nodes also use learned cost functions in GEAR. Once a node notices that it's next hop node does not find neighbours for a good number of rounds, it uses the learned cost function and sends the packet directly to an appropriate node. Concept of recursive geographic routing is also used.

Zeng et al.[10] propose two protocols under the head of Geographic Routing with Environmental Energy Supply (GREES) named GREES-L and GREES-M. A number of factors govern the routing decision in GREES-L and GREES-M, which are, channel condition, residual energy of the sensor,packet advancement to destination and environmental energy supply. The L and M in GREESs stand for linear and multiplication respectively.

In [11] the authors propose Geographic Multi-cast Routing (GMR). GMR is based on geographic routing and selects the neighbours on the basis of a cost function. During multi-cast routing, minimum consumption of resources is ideal. The cost function developed in GMR helps in efficient data delivery as well as minimum bandwidth usage.

In [13] the authors propose two protocols that contribute towards a better lifetime and stability period of the network. In the protocols proposed named, Hybrid Energy Efficient Reactive (HEER) and Multihop Hybrid Energy Efficient Reactive (MHEER), clusters are formed. For cluster head (CH) selection residual energy is taken into account. The difference between the cluster formation in both the protocols is that in MHEER, the number of clusters is predefined and so is the number of cluster heads.

Density controlled Divide-and-Rule (DDR) is proposed in [15]. It is similar to the scheme in [14] in certain aspects. The network is divided into concentric squares and further regions. Each region has a cluster head. The cluster head selection is different from [14].

Popescue et al.[16] present CMSER in which next hop forwarder is chosen on the basis of shortest mean square error (MSER) ratio and minimum error. Optimal energy point is not considered in CMSER. Further, a condition is also applied in CMSER along with MSER that uses the variance of the erroneous distance. The objective function of CMSER is:

$$F_j = argmin(MSER_{ij}) \tag{1}$$

where MSER is given as

$$MSER_{ij} = MSE_{ij}/dij \tag{2}$$

CMSER outperforms LED in terms of throughput under various changes in the network density, transmission range and error variance.

In [7] properties of the two protocols discussed a-priori are combined to achieve even a better throughput and an energy efficient routing protocol. Errors belong to Gaussian distributed making the distances members of Rician distribution. Like LED, it also chooses an energy optimal position. It chooses the forwarder closest to that optimal position. For that it has the distance between the optimal position and the possible forwarders included in it's objective function along with the MSE. Besides the objective function it also uses the condition as in Eq.10.

3 Motivation

Energy consumption is one of the most important parameters of a network, especially in wireless sensor networks where the battery power is limited [17]. According to the nature of a network, the energy consumption is either compromised at the expense of other parameters or other parameters are compromised at the expense of decreasing the energy consumption. Sensor networks often have to be deployed in areas where replacing batteries is a tedious task.Along with scalability, sensor networks should be designed to cater for energy consumption [18]. The network lifetime of sensor networks directly depends on the energy consumed per packet. When it comes to geographic routing, energy consumed can further be increased in case of a wrong choice of forwarder. A geographic routing algorithm needs to be developed that takes into consideration this energy constraint in Wireless Sensor Networks (WSNs).

4 E-ECMSE

The error model developed stands similar to that of ECMSE. In this scheme the location errors are modelled as independent normal random variables. These errors affect the measurement of actual positions of the nodes. This way if x and y are actual coordinates of a node, an errored measurement shall make it:

$$X = x + W_x \tag{3}$$

$$Y = y + W_y \tag{4}$$

Where W_x and W_y are the errors in the x and y coordinates respectively, however in this work the error variance is considered the same on the x and y axis. X and Y are taken as estimated coordinates. The distance between these coordinates is given as:

$$\hat{d_{ij}} = \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2} \tag{5}$$

The estimated distances are Rician random variables. The distance between the actual coordinates can also be measured by using Euclidean distance formula.The expectation of the estimated distances, used in the calculation of the mean square error is given as:

$$E(\hat{d_{ij}}) = \sigma_{ij}\sqrt{\frac{\pi}{2}}L_{1/2}(\frac{-d_{ij}^2}{2\sigma_{ij}^2}) \tag{6}$$

σ_{ij} in the above equation can be calculated as; $\sigma_{ij} = \sqrt{\sigma_i^2 + \sigma_j^2}$ The $L_{1/2}$ in the above equation represents the Laguerre polynomial which is given as:

$$L_{1/2}(x) = e^{x/2}[(1-x)I_0(\frac{-x}{2}) - xI_1(\frac{-x}{2})] \tag{7}$$

I_o and I_1 represent the Bessel function of first class; zeroth and first order respectively[6].I_o can be calculated as:

$$I_0(x) = \frac{1}{\pi}\int_0^\pi e^{x\cos\theta}d\theta \tag{8}$$

ECMSE uses the mean square error associated with each forwarder as a decision metric:

$$MSE = E(\hat{d_{ij}}^2) - 2dijE(\hat{d_{ij}}) + d_{ij}^2 \tag{9}$$

There is a condition further used :

$$(R - \hat{d_{ij}})^2 > Var(\hat{d_{ij}}) \tag{10}$$

$E(d_{ij})^2$ in the MSE can be calculated as:

$$E(\hat{d_{ij}}^2) = 2\sigma_i^2 + \sigma_j^2 + x_i^2 + x_j^2 + y_i^2 + y_j^2 - 2_x ix_j - 2y_iy_j \tag{11}$$

Unlike schemes that choose the forwarder based on the largest distance between the forwarder and the source nodes, ECMSE opts for choosing the forwarder closest to the energy optimal position M. Choosing the forwarder as the farthest node from the source reduces the number of hops [7] and also increases the chances of packet reception. The value of M can be calculated using the energy model discussed in ECMSE as [19]. Where;

$$E_t = e_{tx} + e_{rx} \tag{12}$$

E_t is the total energy spent in the network and e_{rx} is the energy spent on reception and e_{tx} is the energy spent on transmission. The energy spent while communicating includes the energy spent on radio electronics at the transmitter and the receiver and the energy spent on amplification at the transmitter. Energy spent on the radio electronics for both the transmitter and the receiver are assumed as equal. Thus total energy is given by

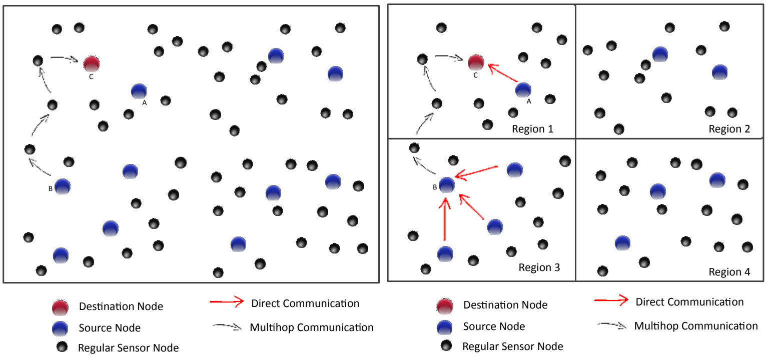
$$E_t = e_{tx-amp} + 2e_{elec} \tag{13}$$

E_t can be further simplified as

$$E_t = \beta(d^\alpha) + c \tag{14}$$

Where $\beta(d^\alpha) = e_{tx-amp}$ and $c = 2e_{elec}$.Value for M can be obtained on solving two equations which are; $(y_i - y_m) = m(x_i - x_m)$ and the equation for distance between source node i and the energy optimal position M, $d_{iM} = \sqrt{(x_i - x_m)^2 + (y_i - y_m)^2}$. Solving these two equations we get; $x_M = x_i \pm \frac{d_{iM}}{\sqrt{1+m^2}}$ and $y_M = y_i \pm \frac{md_{iM}}{\sqrt{1+m^2}}$. The value for d_{iM} can be obtained from the following [6]

$$d_{iM} = \sqrt[\alpha]{\frac{c}{\beta(1-2^{1-\alpha})}} \tag{15}$$



(a) ECMSE (b) E-ECMSE

Fig. 1 Network Models for ECMSE and E-ECMSE

One can use estimated distances instead of original distances in the equations mentioned as in [20]. Enhanced Energy Conditioned Mean Square Error Algorithm (E-ECMSE) makes use of the mean square error and optimal energy position similar to ECMSE. Its objective function is similar to that of ECMSE:

$$F_j = argmin(MSE_{ij} * \hat{d_{jM}}) \tag{16}$$

Alongwith its objective function it also makes use of the condition mentioned above in eq.10. Where F_j stands for the forwarder node. In ECMSE a lot of computational overhead is generated. Reason for that being, most of the packets sent, in the scenarios discussed, are sent through multi-hop communication. Multi-hop communication causes more energy depletion than direct communication. All of the nodes involved in the forwarding, loose their energy each time they are selected. This property puts a serious impact on the energy consumption in cases where successfully received packets have to be responded with an acknowledgement. Acknowledgement goes through the same path as the packet received; thus, even further, depleting the energy of the forwarders. If ECMSE is designed as such that the number of nodes that opt for multi-hop communication is controlled, energy depletion in the network can be reduced.

In E-ECMSE, taking advantage of location properties of geographic routing, we divide the whole field into four regions. All nodes (randomly deployed), including the destination node, now belong to a region. If any source node lies within the same region as the destination node, it sends its packet directly to the receiver. This omits the need to go through rigorous computations involved in multi-hop communication. The other sources, whichever region they belong to, send their packets directly to the source node, in their region, that has the least distance from the destination node. The network has the information of inter-nodal distances a-priori, so calculation of distances does not impact the computations a great deal. This way, no matter how many nodes are nominated as sources, only three source nodes opt for multi-hop communication. This reduces the energy consumption numerous folds. All of the energy consumed in the network, is thus dependant on the number of hops a routing path includes. Number of hops and transmission number affect the energy directly [7], especially in cases where acknowledgement is sent.

As shown in fig.1 node A, which is a source node, lies in the same region as the destination node (marked as node C). It sends its packet to the destination through direct communication. Node B on the other hand lies in region 3 and is the closest to the destination node among all the other sources in region 3. It receives packets from other sources in its region directly. It then transmits these packets to the destination via multi-hop communication. The nodes in the other two regions are going to show similar behaviour.

4.1 Mathematical Formulation

Laguerre polynomial and Bessel's function are used in the calculations of E-ECMSE. Laguerre polynomials have a number of definitions. These are orthogonal polynomials [21]. Laguerre polynomials have a number of applications in daily life. The are used for analysing particles and also help in the study of resonance frequencies and oscillation.

Laguerre polynomial is a solution to the Laguerre equation, which is as under [21]:

$$x\ddot{y} + (1-x)\dot{y} + ny = 0 \tag{17}$$

After solving the above equation we get the Laguerre polynomial which is its solution. Laguerre polynomial is given as under:

$$L_n(x) = e^x((x^n e^{-x})^n) \tag{18}$$

In the Laguerre polynomial discussed here, Bessel's functions are also incorporated. Bessel's functions are used to study various real life problems. It can be used to study the problems of wave propagation, static potentials and also in the study of cylindrical and spherical coordinates. The Bessel's equation can be written as in [22]:

$$x^2\ddot{y} + (1-x)\dot{y} + (x^2 - n^2)y = 0 \tag{19}$$

The first order Bessel's function is as under:

$$J_v(x) = (\frac{x}{2})^v \sum_{k=0}^{\infty} (\frac{(-1)^k}{\Gamma(k+1)\Gamma(v+k+1)})(\frac{x}{2})^{2k} \tag{20}$$

The second order Bessel's function is given by:

$$Y_v(x) = \frac{\cos(v\pi)J_v(x) - J_{-v}(x)}{\sin(v\pi)} \tag{21}$$

Table 1 Simulation Parameters

Simulation Parameters(unit)	Symbol	Value
Path loss exponent	σ	3
Packet size(bits)	p_{size}	1024
Number of packets/source	pkts	1
Energy spent on radio electronics (nJ/bit)	e_{elec}	50
Energy spent on transmission (J/bit)	e_{tx}	2.5e-7
Energy spent on reception (J/bit)	e_{tx}	1.5e-7
Constant (pJ/bit/m^2)	β	100
Network side length (m)	l	50

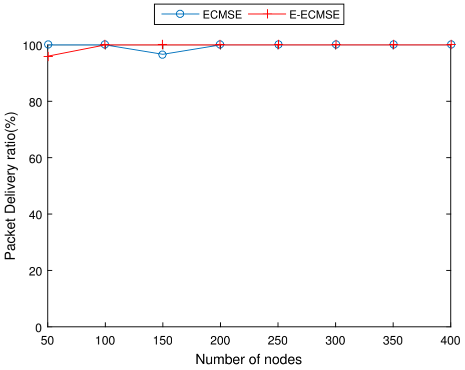


Fig. 2 Routing performance for Scenario 1

5 Simulations and discussions

The performance of the proposed scheme is compared to ECMSE in this section. The simulation parameters are given in the following table:

All of the nodes are randomly distributed and five scenarios in this case are studied. In these scenarios network density, Transmission range and maximum standard deviation of errors is varied. For scenario 4 and 5, acknowledgement is sent for every packet received. ACK is a field for differentiating the cases where acknowledgement is received (Y) and where it is not received (N). Sending acknowledgements adds to the energy consumption of the network. SE is the number of source nodes used. The calculations for energy consumption in the plots are made using the following equations:

$$E_{trans} = TrNo * e_{tx} * pkts * SE * p_{size} \tag{22}$$

$$E_{rcv} = HopNo * e_{rx} * pkts * SE * p_{size} \tag{23}$$

Where E_{trans} is the transmission energy and E_{rcv} is the reception energy. TrNo is the total number of transmissions and HopNo represents the average number of hops. The energy consumed mainly depends on HopNo and TrNo. The greater the value of these two factors,greater is the energy consumed. Simulation scenarios are given in table.2.

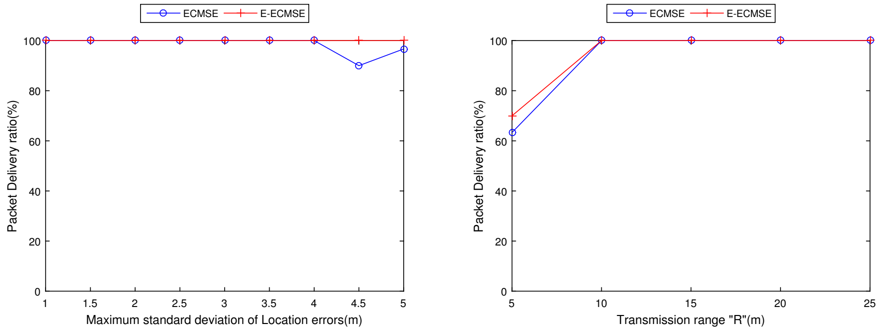
Table 2 Simulation parameters

Scenario	N	R(m)	σ_{max}	SE	ACK
1	50-400	40	8	10	N
2	200	10	1-25	10	N
3	200	5-25	1	10	N
4	100-500	10	1.5	1	Y
5	100-500	10	1.5	25	Y

In fig.2 it can be seen that the PDR stays 100% for most of the readings since the network is dense enough to ensure a 100% PDR. However, for an initial period when the number of nodes is less, E-ECMSE performs better than ECMSE. Since in E-ECMSE source nodes, present in the same region as the destination, directly send the data without having to look for possible forwarders. Possible forwarders if not found, cause packet drop. Direct transmission increases the chance of reception of data. In ECMSE, however, the packets are directly sent to the destination only if they fall under a certain value of transmission range. In case they do not, they switch to multihop communication, multihop communication brings the ordeal of finding the forwarders that fit into the calculations, if not found the packet is dropped. Thus it can be reached that E-ECMSE is better than ECMSE, when it comes to throughput, for sparse and dense networks alike.

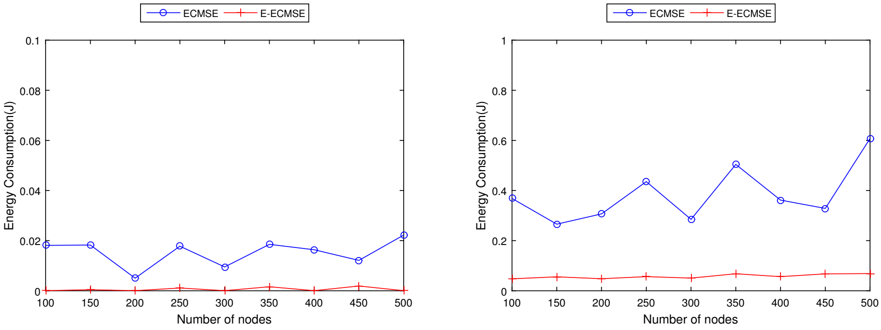
In fig.3a. PDR stays highest for most readings but for ECMSE it decreases as the location errors increase. The location errors cause inaccurate measurements and hence a decrease in the PDR. In scenario 2, although the R is reduced, E-ECMSE still performs better than ECMSE. It can be seen that E-ECMSE is more error resilient than ECMSE and performs better in higher location errors.

In fig.3b It can be seen that there is an obvious increase in PDR for higher values of R. Greater the transmission range, greater the chances of finding the destination within the transmission range. Further, even if the destination does not lie within the transmission range, there is a greater chance of finding forwarders within the range. These factors cause an increase in the throughput as the transmission range increases. E-ECMSE shows considerably better results than ECMSE for smaller R also. When the transmission range is set to 5 meters, ECMSE gives a PDR of around 23%, whereas E-ECMSE gives a PDR of around 70%. Hence it can be noted that the E-ECMSE performs well under small values of transmission range too. For scenario 4 and 5, plots for energy consumption and number of hops is discussed. The plots for similar parameters are plotted together for analytical convenience.



(a) Scenario 2 (b) Scenario 3

Fig. 3 Routing performance for Scenario 2 and Scenario 3

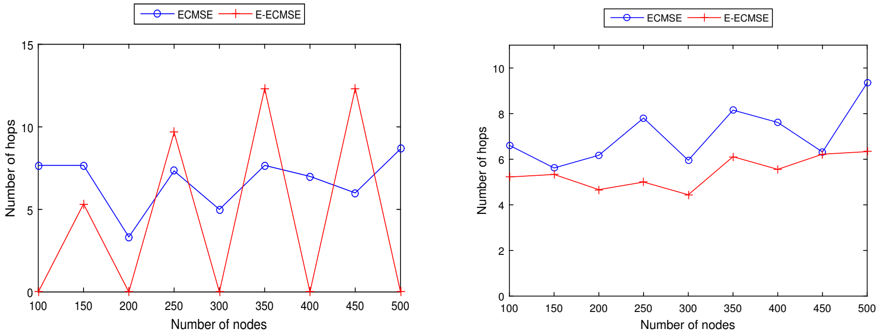


(a) Scenario 4 (b) Scenario 5

Fig. 4 Energy Consumption for Scenario 4 and Scenario 5

In fig.4a There is an increase in the energy consumption across higher node densities. Higher network densities tend to provide a node with greater forwarding options, in both the protocols the least distant neighbors are chosen which in turn increase the number of hops. As the number of hops increases, the energy spent on the transmission of packet also increases. In case the packet is received successfully, there is an acknowledgement sent through the same path, which again increases the energy consumption. As it is evident from the figure, energy consumption in the E-ECMSE is far less than the energy consumption in ECMSE. In fig.4b the values of energy consumption are much higher than in scenario 4. The reason for this is the fact that 25 packets are sent in scenario 5 instead of just 1. The behaviour of the energy consumption is ,however, the same. Energy is depleted more as the number of nodes increases. E-ECMSE still stands far behind ECMSE when it comes to energy consumption.

In fig.5a Number of hops, as discussed earlier, also increase with the increase in the network density. There is a saw tooth graph for number of hops for the E-



(a) Scenario 4

(b) Scenario 5

Fig. 5 Average number of hops for Scenario 4

ECMSE. This is due to the fact that since only one packet is sent in this scenario, if the source sending the packet lies within the same region as the destination, direct communication takes place. Direct communication is a case where number of hops is taken as zero. Network densities that have multi hop mode of communication, show a positive value as the number of hops and wherever direct communication takes place, there is a drop in the value. Collectively the number of hops in E-ECMSE is still less than ECMSE.

In fig.5b the number of hops increases as the node density increases,as mentioned priorly. The E-ECMSE here too, outperforms ECMSE.

6 Conclusion and future work

In this paper, we analysed error resilient geographic routing schemes and presented an enhanced version of ECMSE. E-ECMSE outperforms its contender ECMSE. It performs better than ECMSE when it comes to throughput as well as energy consumption. The analysis is made under variable network densities, transmission ranges and standard error deviations. Cases where acknowledgement is sent for every packet received are also studied. All of the cases under study confirm the superiority of the E-ECMSE.

In future, this study can be taken further considering the effects of the channel, which are not considered here.

References

1. Maghsoudlou, A., St-Hilaire, M., & Kunz, T. (2011). “A survey on geographic routing protocols for mobile ad hoc networks”. Systems and Computer Engineering, Technical Report SCE-11-

03.Carleton University.2011.49 p.
2. Ruhrup, S. (2009). “Theory and practice of geographic routing”. Ad Hoc and Sensor Wireless Networks: Architectures, Algorithms and Protocols, 69.
3. Seada, K., Helmy, A., & Govindan, R. (2004, April). “On the effect of localization errors on geographic face routing in sensor networks”. In Proceedings of the 3rd international symposium on Information processing in sensor networks (pp. 71-80). ACM.
4. Shah, R. C., Wolisz, A., & Rabaey, J. M. (2005, May). “On the performance of geographical routing in the presence of localization errors [ad hoc network applications]”. In IEEE International Conference on Communications, 2005. ICC 2005. 2005 (Vol. 5, pp. 2979-2985). IEEE.
5. Takagi, H., & Kleinrock, L. (1984). “Optimal transmission ranges for randomly distributed packet radio terminals. IEEE Transactions on communications ”, 32(3), 246-257.
6. Peng, B., & Kemp, A. H. (2011). “Energy-efficient geographic routing in the presence of localization errors”. Computer Networks, 55(3), 856-872.
7. Popescu, A. M., Salman, N., & Kemp, A. H. (2014). “Energy efficient geographic routing robust against location errors”. IEEE Sensors Journal, 14(6), 1944-1951.
8. Kim, Y., Lee, J. J., & Helmy, A. (2004). “Modeling and analyzing the impact of location inconsistencies on geographic routing in wireless networks”. ACM SIGMOBILE Mobile Computing and Communications Review, 8(1), 48-60
9. Yu, Y., Govindan, R., & Estrin, D. (2001). “Geographical and energy aware routing: A recursive data dissemination protocol for wireless sensor networks”.
10. Zeng, K., Ren, K., Lou, W., & Moran, P. J. (2009). “Energy aware efficient geographic routing in lossy wireless sensor networks with environmental energy supply”. Wireless Networks, 15(1), 39-51.
11. Sanchez, J. A., Ruiz, P. M., Liu, J., & Stojmenovic, I. (2007). “Bandwidth-efficient geographic multicast routing protocol for wireless sensor networks”. IEEE Sensors Journal, 7(5), 627-636.
12. Zhang, H., & Shen, H. (2010). “Energy-efficient beaconless geographic routing in wireless sensor networks”. IEEE transactions on parallel and distributed systems, 21(6), 881-896.
13. Akbar, M., Javaid, N., Khan, Z. A., Qasim, U., Alghamdi, T. A., Mohammad, S. N., ... & Bouk, S. H. (2015). “Towards network lifetime maximization: sink mobility aware multihop scalable hybrid energy efficient protocols for Terrestrial WSNs”. International Journal of Distributed Sensor Networks, 2015, 10.
14. Latif, K., Javaid, N., Saqib, M. N., Khan, Z. A., Qasim, U., Mahmood, B., & Ilahi, M. (2015). “Energy hole minimization with field division for energy efficient routing in WSNs”. International Journal of Distributed Sensor Networks, 2015, 12.
15. Latif, K., Javaid, N., Saqib, M. N., Khan, Z. A., & Alrajeh, N. (2016). “Energy consumption model for density controlled divide-and-rule scheme for energy efficient routing in wireless sensor networks”. International Journal of Ad Hoc and Ubiquitous Computing, 21(2), 130-139
16. Popescu, A. M., Salman, N., & Kemp, A. H. (2013). “Geographic routing resilient to location errors”. IEEE Wireless Communications Letters, 2(2), 203-206.
17. Kadi, M., & Alkhayat, I. (2015). “The effect of location errors on location based routing protocols in wireless sensor networks”. Egyptian Informatics Journal, 16(1), 113-119.
18. Melodia, T., Pompili, D., & Akyildiz, I. F. (2004, March). “Optimal local topology knowledge for energy efficient geographical routing in sensor networks”. In INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies (Vol. 3, pp. 1705-1716). IEEE.
19. Heinzelman, W. B., Chandrakasan, A. P., & Balakrishnan, H. (2002). “An application-specific protocol architecture for wireless microsensor networks”. IEEE Transactions on wireless communications, 1(4), 660-670.
20. Salman, N., Ghogho, M., & Kemp, A. H. (2014). “Optimized low complexity sensor node positioning in wireless sensor networks”. IEEE Sensors Journal, 14(1), 39-46.
21. Radulescu, V. (2008). “Rodrigues-type formulae for Hermite and Laguerre polynomials”. An. St. Univ. Ovidius Constanta, 16, 109-116.
22. Kreh, M. (2012). “Bessel functions”. Lecture Notes, Penn State-Gttingen Summer School on Number Theory, 82.

In-Vehicle Cloudlet Computing based on Delay Tolerant Network Protocol for Disaster Information System

Masaki Otomo[1], Goshi Sato[1], Yoshitaka Shibata[1]
[1] Iwate Prefectural University,
152-52 Sugo, Takizawa, Iwate, Japan 020-0193

g2311008@s.iwate-pu.ac.jp
sato_g@ipu-office.iwate-pu.ac.jp
shibata@iwate-pu.ac.jp

Abstract. In this paper, we propose an in-vehicle cloudlet computing disaster information system that can flexibly deal with critical network connectivity. In order to achieve this purpose, we develop a dynamic allocation of server resources in accordance with the load change on the system so that it is possible to take full advantage of in-vehicle server and network resources in the disaster areas. Also, by introducing mobile cloudlet computing and DTN protocols, our system can realize rapidly sharing disaster information even if the communication infrastructure is disconnected or challenged in the disaster.

Keywords: Disaster Information System, Cloudlet Computing, System Virtualization, Delay Tolerant Networking

1 Introduction

From the geological conditions of Japan Island, many serious disasters such as earthquake, tsunami and typhoon occur in history. A huge number of people, buildings and communication infrastructure are completely damaged. In fact, the Great East Japan Earthquake on March 11, 2011 [9] and Kumamoto Earthquake on April 16, 2016 brought huge damages [*]. Many Information network infrastructures were destroyed and the network traffics were seriously congested. In order to respond to the anticipated large scale disasters, such as Nankai Trough Quake and Tokai earthquake, GIS based disaster prevention systems which can perform collecting and sharing disaster information, resident safety confirmation, decision making to disaster-response headquarter are developed

On the other hand, in recent, cloud computing is getting popular for various business fields because of its easy and efficient introduction and elastic expandability for computing resource allocation. Using cloud computing services, a series of

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_24

preliminary works including design and maintenance of hardware and software are carried out at a data center. Since the users do not need to newly introduce servers physically for business, the maintenance cost can be largely reduced. Furthermore, by introducing network and server virtualization technologies, user can easily construct and run his own private cloud computing system. Thus, there are many advantages to use the cloud computing to provide Internet and Web services.

On the other hand, so far we have investigated the research of a distributed disaster information sharing system by considering mobile environment on disaster situation [1]. In this system, the network states are monitored at background. If network access to Internet is difficult, then the disaster information is locally stored on the mobile relay station. After moving to the location where the Internet connection can be established, then the stored disaster information can be transmitted to the objective disaster information server on the Internet. However, this system does not consider the case where the server load changes rapidly and the network and server failures occur. When the disaster information server is operated just after the disaster occurred, the system failure and traffic load concentration have to be considered. [10]

In this research, we introduce a mobile cloudlet computing disaster information system for large scale disasters to be able to keep continuously disaster information collection and sharing operations even the network environment is unstable or challenged. The computing resources can be also dynamically provided to different user's groups or organizations such as different local governments and offices as required to maximize the physical resource utilization. Furthermore, this system can not only provide information transmission by introducing DTN protocol on the network, but more quick collection and sharing functions by introducing mobile cloudlet disaster information servers where the communication networks are unstable or even disconnected.

2 Related Works

There are several disaster formation systems so far. The system of previous research [2] was developed to respond to the case where large delay and frequent link disconnection happen as a network environment. In this system, the data can transmit if the disaster server can connect to the network by monitoring the network state. The disaster information can be also smoothly shared with multiple servers of different organization such as different local governments. However, this system cannot consider the case where rapid network and system traffic change and failure. When the servers of the local government are failed by external factors such as tsunami, those systems cannot be served.

In the other previous system, disaster information can be visually shown on the display by combining with GIS system or namely digital map. The user can easily understand what kind of the disaster information is registered in the system by properly using various icons and figures related to disaster properly. Furthermore, by operating 'seek bar' on the window, the registered disaster information can be displayed in temporal order as replay operations on video window.

3 Proposed System Configuration

Figure 1 show a network system of our proposed disaster information system. There are two types of cloud computing including GDC and LDC are introduced. The GDC is a central cloud computing located at somewhere on Internet and integrates all of the disaster information stored in Temporal Servers (TSs) and LDCs in each local area.

The LDC is based on a mobile typed cloudlet computing which is carried on vehicle. The LDC performs as cloudlet computing server. The LDC circulates around the local government office, the evacuation places, the community centers and public places where the TSs are located. The TSs as temporal servers store disaster information after the disaster occurred in each local government area. When the communication network can be available, those TSs can directly share the disaster information with the GDC. When the communication network cannot be available, those LDCs go around the disaster area and collect and store the local disaster information from the TSs until the communication network in this local area recovered.

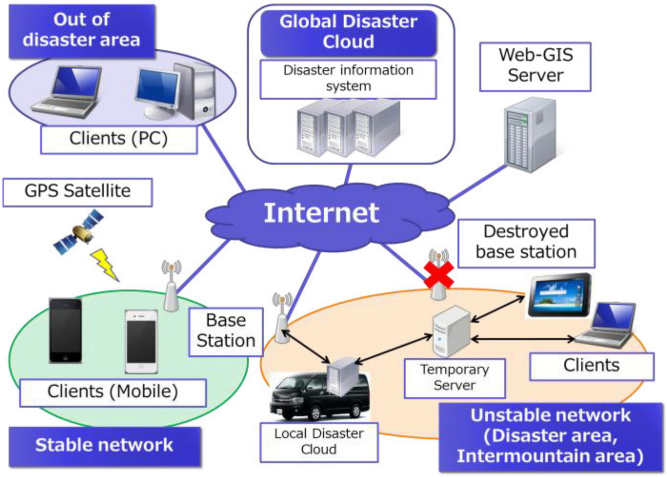


Figure 1 System Configuration

When the communication network is note available, Delay Tolerant Network (DTN) [6][7] is employed. DTN is defined as the communication protocol which can realize reliable data transmission on the challenged network condition with large delay teime and frequent network disconnection in addition to the normal network condition. In our system, DTN protocol is used to exchange the information between the LSs and LDC or LDC and GDC.

Figure 2 shows the system behavior on challenged network condition case where the network connection to Internet is unstable or even though disconnected. Many different wireless network deivces including 3G/LTE, Wi-MAX, Wi-Fi and satellite networks are installed to organize a cognitive wireless network. Just after occurrence of disaster, movile LDC vehicle as cloudlet computing server with the cognitive wireless netowork perambulates around the disaster area and approaches to the LS. Then the LDC automatically recieves those stored data from the LS by DTN protocol. When the LDC approaches to GDC, then the GDC automatically retrieve the stored data from LDC. By collecting those disaster information, all of the data between GDC and TS, GDC and LDC are synchronized to gurrantee the consistency. Thus, data transmission between LSs and GDC can be realized throuth the LDC using DTN protocol even through the Internet cannot be available in the disaster area.

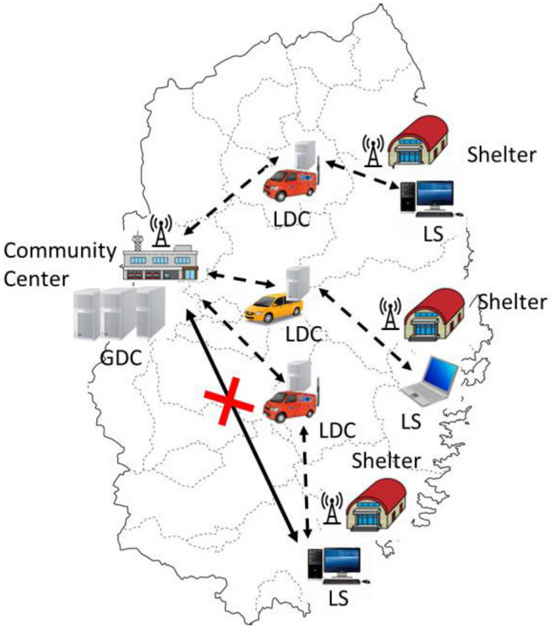


Figure 2 System behavior on challenged network condition

4 System Architecture

Figure 3 shows an architecture of our proposed system. By assigning the required number of Virtual Machines (VM) to each cloud server, the disaster information system can be provided. The proposed disaster information system is consisted of Monitoring Module (MM), Resource Management Module (RMM), VM Control Module (VMCM) and DTN Transport Module (DTM).

The MM monitors resource utilization rates of CPU and memory of the VM and sends them to RMM. The RMM controls VM resource assignment based on the results and sends an operation commands to VMCM such as start, stop, addition and reduction of VMs. The DTM manages sending/receiving/storing data by DTN protocol.

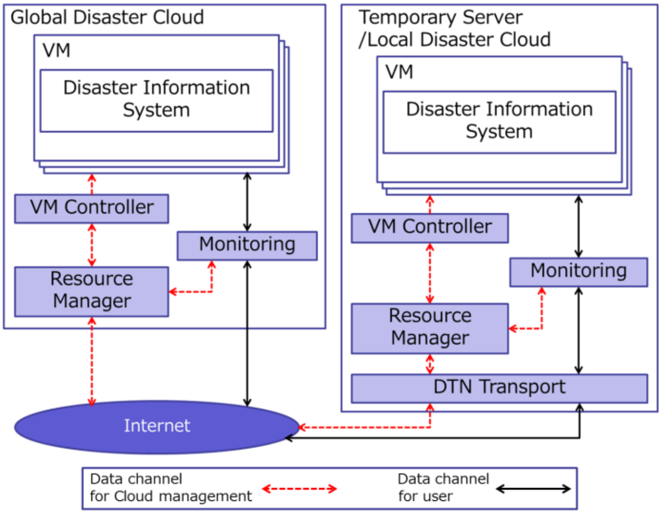


Figure 3 System Architecture

The process for resources addition is based on [3][4][5]. First, the resource utilization rates of VMs are monitored by MM. Second, the number of VM clocks or CPU resources themselves are controlled by referring the utilization rates of CPU and memories. When the CPU utilization rate is more X [%], then the CPU clock frequency is increased until U[GHz] or the number of CPU resources is increased. Also when CPU utilization rate is less than X [%] and memory utilization rate is more than Y [%], the volume of memory is increased is increased until V[Mbytes]. Finally, the process of load valancing for each resource is executed by depending on the change of VM specification.

5 Disaster Information

In our research, two types of disaster information system are supported; one is stricken area information sharing system, and another is safety information sharing system.

Disaster area information sharing system: Just after occurring disaster, the disaster residents register their safety information and location's data, and the disaster aid volunteer send photo images with disaster information, position information and stricken area information using mobile phones to the disaster information server. The collected disaster area information is overlaid on Denshi-Kokudo as icons and displayed by the PC client. All of the disasters are not only categorized and displayed individually, but all of the disaster can be integrated into one category and displayed on the same display. Since the icons correspond to the categories that were selected when the disaster information was registered, the user can easily understand what kind of disaster is dominated in particular disaster area. In addition, when a user reads this information on a PC client, one can use the temporal presenting operations and understand the change of the state of the stricken area through time.

The safety information : The safety information is very important to confirm the lives of the evaluated residents when disaster is just occurred. It is assumed that there are two different registration cases; one is that the residents directly register his safety state with his locations by GPS using their mobile telephone, and another is that the residents evaluate to the pre-specified shelter and the disaster volunteers in the shelter register from the PC server instead of the residents.

6 Prototype System

6.1 Prototype System Configuration

In order to verify of usefulness of our proposed system, a prototype system is constructed and evaluated its functionality and performance. Figure 4 shows system configuration of the prototype of our proposed system. In the prototype system, the cloud system in the prototype, Management Server manages the whole cloud system and Host provides VM resource. As LDC, a note PC is used to easily carry on the mobile vehicle. The hardware and software specification of GDC, TS, LDC is shown in the Table 1.

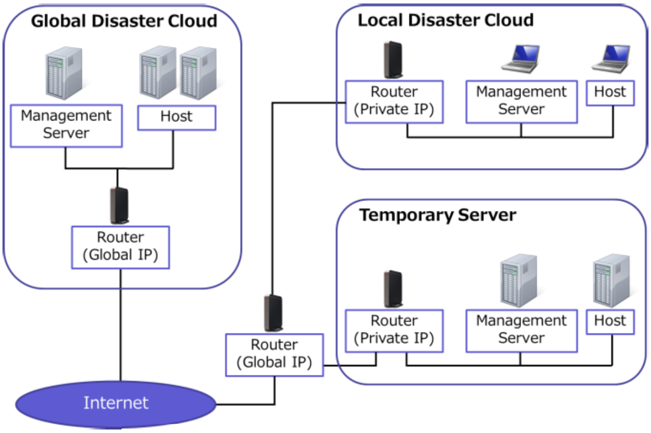


Figure 4 Prototype System

Table 1 Hardware Specification of LDC and GDC Servers

	GDC, Temporary Server	LDC
CPU	Intel(R) Core(TM) i3-3240 Processor (3M Cache, 3.40 GHz)	Intel(R) Core(TM) i7-4712MQ Processor (6M Cache, 2.30 GHz)
RAM	4GB	16GB
HDD	500GB	1TB

6.2 Cloud Stack

As cloudlet computing system environment, we applied CloudStack [8] which is one of open source and can provide an infrastructure as Service (IaaS) to construct both public and private cloud computing system such as Amazon EC2. The CloudStack is used by many organizations because of its excellent GUI and easy operations. Since load balancer and firewall functions as internal architecture are also installed as standard system, more functional expansion can be possible. In our system, dynamic resource control function of VM depending on the resource utilization rate is implemented using CloudStack API on Linux OS environment.

7 Performance Evaluation

7.1 Performance of LDC

In order to understand how many user accesses can be possible in our prototyped system, the performance evaluation of LDC was executed as preliminary experiment. In the experiment, the three cases where the number of VMs assigned to each local government area are one, three and five, respectively are investigated. The disaster information system is installed and operated on each VM. HTTP requests are generated from the client machines and issued to the disaster information system on the VM and the average response time (msec) and throughput (req/sec) are observed by increasing the number of user accesses on the VMs and changing request interval from 1 to 20sec. As experiment, JMeter [11] is used. The experimental parameters and conditions are shown in Table 2.

Table 2 Hardware Experimental parameters

Specification Items	Parameters
CPU	1GHz
Main Memory	1GB
HDD	10GB
No. of Accesses for a VM	100~600
Thread Generation Time	10 sec
Request Interval	1, 2, 5, 10, 20 sec

Figure 5 shows the result of the case where three VMs are assigned for one local government. Regarding with the response time, low values are maintained for less than 260 threads. At the same time, the throughput values are also linearly increased. From this result, 26 requests/sec from users can be processed on one VM.

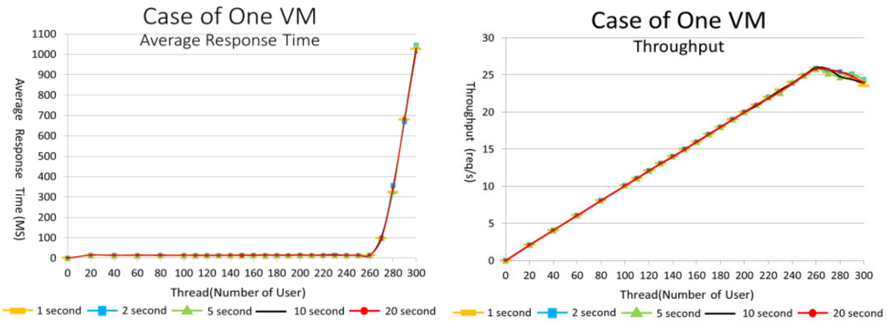


Figure 5 Response Time and Throughput for One VM

One the other hand, Figure 6 shows the result of the case where five VMs are assigned for five local governments. Regarding with the response time, low values are maintained for less than 180 threads. At the same time, the throughput values are also linearly increased. From this result, 18 requests/sec from users can be processed on one VM.

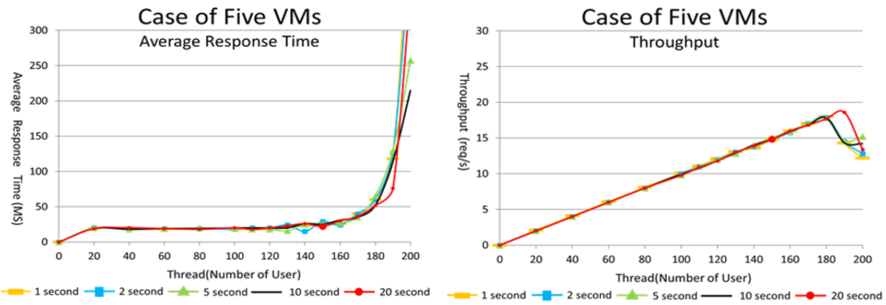


Figure 6 Response Time and Throughput for Five VMs

From those results, the total number of threads can be processed is increased as the number of VMs is increased until some limited points. Namely 260 threads for one VM and 180x5=900 threads for five VMs. Thus, through this experiment, it is clear that although the possible number of accesses is increased as the number of the VMs is increased while keeping the response time almost constant.

7.2 End-to-End Response time between LS and GDC

Next, Performance of end-to-end response time between LS and GDC through the LDC using DTN protocol is evaluated. Figure 7 shows the traverse of LDC to TS and GDC.

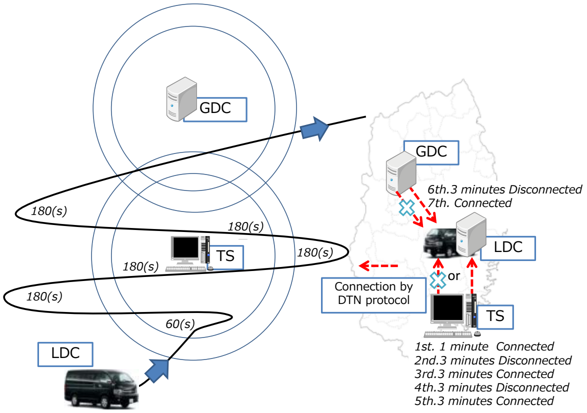


Figure 7 Perambulation of LDC between Each Server

First, the vehicle with LDC traverses in the disaster area and receives the disaster information on TS in the shelter using DTN for 60 sec. ("Connected"). Then, the vehicle releases from the shelter for 180 sec. ("Disconnected"). After that the vehicle comes back to the shelter and again receives the disaster information for 180 sec. ("Connected"). In this procedure, "Connected" or "Disconnected" between the LDC and TS servers are repeated until the 6th step in Figure 8 in this experiment. From the 7th step, the LDC vehicle traverses to Headquarter where the Internet connection to the GDC can be available and transmits all of the disaster information from the LDC server to the GDC server. Thus, in the procedure form the 1st step through to the 7th step, the network communication can be guaranteed even though those servers are in challenged communication environment.

The experimental results are shown in Figure 8 to Figure 10. The data size of the disaster information to be transmitted between TS to the LDC vehicle and the GDC server is 1Mbyte on every 1 second. The buffer size of the DTN bundle layers on each server is large and enough to store the transmitted data. This means that the minimum network transmission speed requires 8 Mbps.

Figure 8 shows the result of the variation in quantity of the transmitted data which are stored in the DTN buffer of TS.

In the scenario of the 1st step, because the network condition between TS and LDC was "Connected", the data on the DTN bundle layer in TS were immediately transmitted to the DTN buffer in LDC, the quantity of the data was 0 for this period. In the network area, the throughput value to transmit data from TS to LDC was 34.91 Mbps. On the 2nd step, because the network condition was “Disconnected”, the data in the TS were accumulated in the DTN buffer. On the 3rd step, although the network condition was “Connected”, the data were not transmitted for 50 seconds. This is due to the time delay of the restarting process transmission on DTN2 in which the source server took time to find the destination server. In addition, another reason except the redundancy of system configuration is the electric field strength. When the vehicle of LDC connected to the network on TS at the 3rd step, the throughput value was 7 to 12 Mbps at -80dBm. This means the network condition was intermittent because the condition needs more than 8 Mbps to send data. After the 50 seconds, those stored data in the DTN buffer were sent to the LDC with the maximum throughput. On the 4th step, the same result was repeated as the 2nd step. Finally, on the 5th step, the same result was repeated as the 3rd step.

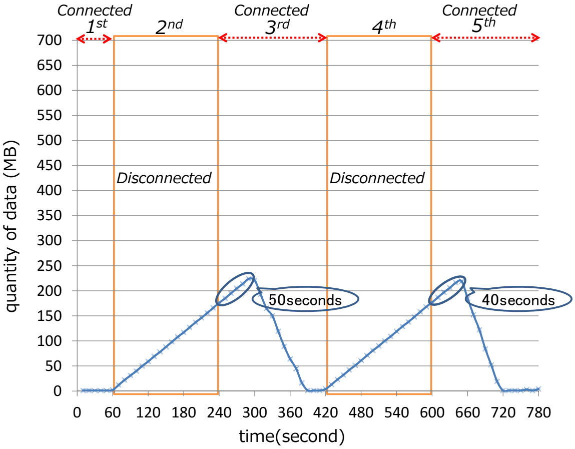


Figure 8 Data Quantity during 1st to 5th Step in TS

Figure 9 also shows the variation in quantity of the received data from TS which are stored in the DTN buffer of LDC.

On the 1st step, the data from TS were received and be accumulated increasingly in the DTN buffer of LDC. On the 2nd step, because the network condition was “Disconnected”, the data transmissions from TS were stopped and the accumulated data of 60 Mbytes were maintained. On the 3rd step, although the network condition was “Connected”, the data from the TS were not transmitted for 50 seconds. This is due to the same reason of TS as shown Figure 8. After 50 seconds, those stored data in DTN buffer were received from TS with the maximum throughput after 290 seconds. On the 4th step, the same result was repeated as the 2nd step and the 5th step, the same result was repeated as the 3rd step. When the iteration in the 5th step finished, the vehicle went out from the TS transmission area and went into the GDC transmission area at 960 seconds. In this transmission area, GDC could receive all of the accumulated data, 744 Mbytes from LDC with the throughput of more 1 Mbyte per second (8 Mbps) as shown in Figure 10. Therefore, these processes were completed at 1710 seconds.

Thus, all of the data generated in TS could be uploaded to the GDC server via LDC even though the network environment is disconnected in the disaster.

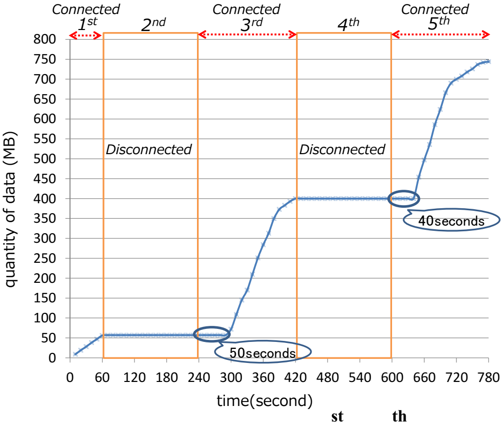


Figure 9 Data Quantity of 1st to 5th Step in LDC

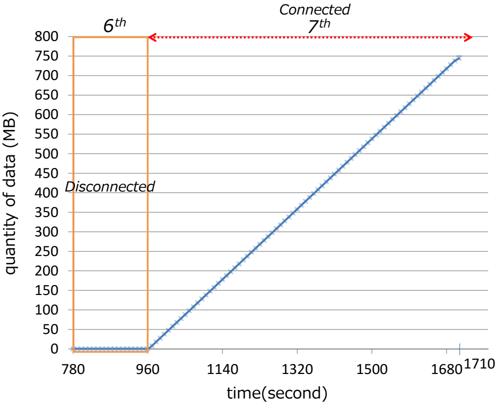


Figure 10 Data Quantity of 6st to 7th Step in GDC

8 Conclusions and future work

In this paper, we proposed mobile cloudlet typed disaster information sharing system based on DTN. Using this system, risk distribution for server failure due to earthquake and tsunami and more flexible server resource control depending on the degree of disaster strength can be realized. Also by introducing mobile typed cloud computing, quick disaster information collection in the disaster areas even just after disaster occurrence can be realized. In addition, using communication means by DTN protocol, data are locally stored when the communication network cannot be available and automatically transmit in the area where the connection to the network can be available, eventually data transmission can be attained on the any network conditions.

As some future researches, we are going to implement and evaluate server resource control functions, data synchronization functions between cloud computing and reliable transmission functions in DTN with the transmission speed improved from TS to LDC.

Acknowledgment. **T**he research was supported by SCOPE (Strategic Information and Communications R & D Promotion Program) Grant Number 142303010 by Ministry of Internal Affairs and Communication in Japan.

References

[1] Y. Sasaki, and Y. Shibata, Construction of Distributed Disaster Information System in consideration of the mobile communication environment, IEICE, "S-65"-"S-66" (2011)

[2] Y. Sasaki, and Y. Shibata, Construction of Disaster information system that enables the display time series uniform, Information Processing Society of Japan (IPSJ), "3-427"-"3-428" (2010)

[3] Trieu C. Chieu, Ajay Mohindra, Alexei A. Karve and Alla Segal, "Dynamic Scaling of Web Applications in a Virtualized Cloud Computing Environment", 2009 IEEE International Conference on e-Business Engineering

[4] Gihun Jung, Kwang Mong Sim, "Agent-based Adaptive Resource Allocation on the Cloud Computing Environment", 2011 International Conference on Parallel Processing Workshops

[5] Zhang Zhang, Jizhong Han, Bo Li, Wei Zhou, Dan Meng, "Lynn: A Multi-Dimensional Dynamic Resource Management System for Distributed Applications in Clouds", 2013 International Conference on Cloud and Service Computing

[6] Delay Tolerant Networking Research Group, <http://www.dtnrg.org/wiki/Home>

[7] Delay Tolerant Networking Research Group "DTN2 Documentation", <http://www.dtnrg.org/wiki/Dtn2Documentation>

[8] CloudStack User Group, <http://cloudstack.jp/>

[9] Ministry of Internal Affairs and Communications "Situation of information and communication in the Great East Japan Earthquake", <http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h23/pdf/n0010000.pdf> 2011

[10] Ministry of Internal Affairs and Communications "Enhance of disaster tolerance in communication", <http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h24/html/nc134210.html> 2012

[11] Apache Software Foundation "Apache JMeter", <http://jmeter.apache.org/>

Towards Heuristic Algorithms: GA, WDO, BPSO, and BFOA for Home Energy Management in Smart Grid

Mudassar Naseem, Samia Abid, Rabia Khalid, Ghulam Hafeez, Sardar Mahboob Hussain, Nadeem Javaid*
COMSATS Institute of Information Technology, Islamabad 44000, Pakistan.

Abstract In this paper, we analyse the scheduling of residential appliances to: 1) reduce cost, and 2) reduce Peak to Average Ratio (PAR) by smoothing load profile. We consider 10 different residential appliances which are categorized into three different groups: shiftable interruptible, shiftable uninterruptible and regular appliances to flexibly control the load. To schedule appliances, Home Energy Management (HEM) systems are designed by using four different heuristic algorithms: Bacterial Forging Optimization Algorithm (BFOA), Genetic Algorithm (GA), Binary Particle Swarm Optimization (BPSO) and Wind Driven Optimization (WDO).

1 Introduction

Smart Grid (SG) is an evolving in electrical grid, which enables the bidirectional communication between electrical utility and consumer while using information and communication technologies. To implement SG Demand Response (DR) and Demand Side Management (DSM) are two important features. DR is a set of rules define by the utility to encourage the consumer to shift their load in response to time varying price signal [1]. DSM prove to be useful in reducing Peak to Average Ratio (PAR) by controlling energy consumption at customer side to improve the reliability of the grid [2]. It also encourage the end user to manage their loads to reduce their electricity bills, by using different scheduling techniques.
In this work our objectives are to 1) reduce cost and 2) reduce Peak to Average Ratio (PAR) by scheduling of residential appliances. The scheduling of different appliances according to given requirements is a complex problem. Alot of substantial research efforts have been put forwarded to investigate the optimal solution of resi-

Corrospondence: *Corresponding author:
COMSATS Institute of Information Technology, Islamabad 44000, Pakistan.
Emails: nadeemjavaid@comsats.edu.pk, nadeemjavaidqau@gmail.com
Website: www.njavaid.com

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_25

dential load management. Different mathematical and heuristic techniques are purposed in literature for scheduling. For example Mixed Integer Liner Programming (MILP) is used in [3] to develop a mathematical model of residential appliances, Photovoltaic (PV) system, storage system, lighting system, heating and air conditioning system. Case studies are carried in Real household, a valuable reduction in cost and PAR is observed in simulations, however system complexity is increased. In [4] scheduling of multiclass appliances is studied, Mixed Integer Non-Linear Programming (MINLP) and modified algorithm are used. Results show that MINLP has high computational complexity than purposed algorithm. MINLP and Genetic Algorithm (GA) are used in [5] to solve the problem of controlling home appliances. Three case studies are conducted in which different residential loads are scheduled. Simulations results show that performance of both techniques are good, however GA requires less computational time. Interruptible load scheduling is studied in [6], Binary Particle Swarm Optimization (BPSO) is used for scheduling. BPSO solve this complex and noncontinuous problem efficiently. Bacterial Foraging Optimization Algorithm (BFOA) is formulated in [7]. Basic working of BFOA, its working procedure is discussed however scheduling problem is not discussed. A comparative study of Particle Swarm Optimization (PSO) and Wind Driven optimization (WDO) is conducted in [8] to solve a Problem of residential load management. Simulation results show that performance of WDO is better than PSO.

A detailed analysis of technical literature clears that both mathematical and heuristic techniques can be used to solve scheduling problems. Although mathematical techniques are quite beneficial but on the other hand they have some drawbacks like high computational complexity. In this regard heuristic algorithms have following characteristics: flexibility for specified constraints, ease of implementation, low computational time and low computational complexity [9]. In our work we analyze four different heuristic based Home Energy Management (HEM) systems, in which each system has same problem but addressed by different algorithm. We discuss two bio-inspired heuristic algorithms (BFOA, GA) and two nature inspired heuristic algorithms (BPSO, WDO) to address the same problem, and then compare their performance parameters. In each HEM system, we consider 10 most commonly used residential appliances, which are categorized into three groups: 1) Shift-able interruptible appliances, 2) Shift-able non-interruptible appliances, 3) Regular appliances. This classification of the appliances is based on the users' behavior. RTP signal is used for pricing. BFOA and GA reduce PAR but percentage reduction in cost is less while BPSO and WDO reduce cost but they create peaks in low peak hours which increases PAR.

2 Related Work

In last few years, energy demand is increased due to which the balance between demand and supply is disturbed. It has created problems like blackout, load sheading etc. To resolve such problems we have to manage the load according to existing generation capacity through scheduling techniques. Many researchers around the globe consider that load management through DSM is the best solution, so in this regard some of papers are discussed. In [10] authors investigate the problem of residential appliances scheduling under dynamic pricing scheme RTP via HEMDAS. This paper purposes an effective home automation system to achieve a favorable trade-off between customers bill and user satisfaction. Authors used their scheme for the operation of both thermal and electrical appliances in a smart home domain. They also considered seasonal price variations and study the effects on cost. Authors use MINLP model for solving optimization problem. The result of purposed model offers a feasible solution for optimal energy management among residential energy users.

Method of smart charging is purposed in [11]. This scheme enhances the efficiency of energy storage systems and scheduling approaches to DSM under real time pricing scheme. An aggregator is introduced which optimally schedule the appliances and battery charging based on DAP scheme to benefit consumers in term of cost reduction and comfort. The results shows that with appropriate scheduling storage devices and carefully designed RTP, customers can apprehend significant saving on their electricity bills. In [12] authors discuss the power demand control scenarios to reduce peak demand. These are default scenario, finite delay request scenario, finite postpone request scenario and finite compressed demand scenario. Recursive formulas are modeled for calculation of peak demand under each scenario. In their modeling they consider finite number of appliances. They associate four different power demand control scenarios with RTP scheme to derive social welfare model in order to reduce cost and peak demand. The results of the purposed model shows an efficient reduction in the peak demand for finite number of devices. Authors in [13] purposes a general and comprehensive optimization base Automated Demand Response (ADR). They schedule the operation of several classes of domestic appliances (Deferrable, Non-deferrable, Thermal, Curtailable and Critical) to minimize energy cost and maximize user comfort. In this paper MINLP is implemented in Advanced Integrated Multidimensional Modeling (AIMM) software to solve the mathematical models.

Theoretical analysis of BFOA is presented in [7]. This paper explains the following: 1) basic theory of BFOA and it parameters, 2) working of important steps of BFOA (chemotaxis, swimming, reproduction and elimination-dispersal), 3) mathematical formulation and 4) flowchart and algorithm of BFOA. Authors in [14] presents a theoretical analysis of BFOA. Theoretical comparative study of BFOA is conducted with other optimization algorithms however, no simulations are conducted to validate the results. Improved version of BFOA (IBFOA) is suggested in [15], in which elimination-dispersal and chemotaxis steps are improved, which increased convergence rate and accuracy.

Implementation method of GA on residential load by using Supervisory Control And Data Acquisition (SCADA) is purposed in [5]. This paper compare the results of GA scheduled load with MINLP scheduled load under three different scenarios while considering power limits. Intelligent Energy Systems Laboratory (LASIE) is used for case study. LASIE consists of a SCADA system, renewable energy sources

(Photovoltaic(PV), wind turbine and fuel cell) and Variable loads for testing. Authors conducted a case study to validated their scheme, they consider three different load scenarios and compare the results. In [16] problem of residential load scheduling is studied under DSM. The main objective of this study is to determine the optimal energy consumption while considering cost function minimized. MATLAB simulations of this study shows that a prominent cost reduction is archived. Energy optimization of small scale smart house having hybrid energy (PV and utility) sources by using GA is presented in [17]. Authors suggest hardware based future work in which controller is build by using GA. Authors in [18] presented a model of optimal integration of Distributed Energy Storage System (DESS) in term of optimal sizing and location. A new method of integration of DESSs and capacitors with SG is purposed to reduce reactive power and cost of power system. Authors conducted a case study in which four different cases of integration of DESS are investigated. Sequential Quadratic Programming (SQP) and GA techniques are used to achieve the required objectives.

Basic concept, variants and applications of PSO in power system is presented in [19]. In this work detail theoretical and mathematical modeling of PSO implementation is presented then different variants of PSO are discussed in detail with their respective mathematical models. Variants that are discussed in this paper are PSO-GA, hybridization of Evolutionary Programming and PSO (EPSO), Multi Objective PSO (MOPSO), Adaptive PSO, Dynamic Neighborhood PSO (DN-PSO), Vector Evaluated PSO (VEPSO), Gaussian PSO (GPSO). In addition, authors discussed the various applications of PSO in power system based optimization problems. Authors in [6] presents the scheduling of interruptible appliances over a 16 hour time period. Objective of this study is to minimize the number of interruptions and minimize the bill payments. BPSO is used for scheduling of 29 interruptible appliances, which are divided into two curtailments i.e. curtailment A and curtailment B. Both of them are scheduled according to required objectives. Problem of managing energy resources by using PSO-Mutation (PSO-MUT)is studied in [20]. Mathematical formulation of objective function is presented in which energy resources like PV system, Wind and Combine Heat and Power (CHP) plant are included. Network constraints of power system are also considered in formulation. Case study is conducted to apply the purposed methodology. A 30 kV distribution network, supplied by substation of 60/30 kV, distributed by 6 feeders, with a 937 buses and 464 MV/LV transformers are used in this study. Results of PSO-MUT shows best average results with slightly high delay than PSO.

A new nature inspired heuristic algorithm is presented in [21] which is called Wind Driven Optimization (WDO). In WDO large number of small air parcels are considered which are randomly moving under the effect of four different forces. Authors model a velocity equation according to these forces which make WDO more controllable and robust. Also a comparative study of WDO, GA, PSO and Differential Evaluation (DE) is conducted on 3 different application of Electromagnetic problems. Results shows an effectiveness of WDO. In [8], comparison of WDO and PSO is studied. Authors investigate that performance of WDO is better then

PSO in term of electric bill reduction and waiting time. furthermore performance of Knapsack-WDO (K-WDO) is also studied.

3 HEM system and Heuristic algorithms

3.1 HEM system

Smart appliances are connected (physical or wireless) to HEM controller to receive instructional signals. Controller is further connected to smart meter which receives DR and RTP signals from Neighbourhood Area Network (NAN). NAN consists of ICT equipments which provides a link between utility and smart homes, it receives usage data and send it to utility similarly DR and RTP signals are received from utility and send it to HEM system. On the basis of this DR and RTP signals HEM system schedule the residential appliances as per according to desired objectives.
In our system, objective are to reduce cost and reduced PAR. To achieve these objectives we schedule 10 different residential appliance which are divided into three categories:
1 Shift-able interruptible appliances: These appliances can be sifted to any time slot and they can be interrupted when required. These appliances include vacuum cleaner, water heater, Water pump, dish washer and hair dryer.
2. Shift-able non-interruptible: These can be sifted to any time slot but when they start their operation, they must complete there operation consecutively i.e. without interruption. These appliances include washing machine and cloth dryer.
3 Regular appliance: Such appliance follows almost same load profile i.e. the scheduled and unscheduled profile is almost same because such appliances are ON in most of the time slots. These appliances include AC, refrigerator and oven.
Appliance along with there power rating and usage hours per day is shown in table 1.
RTP signal is used for electricity bill calculation. Below is the list of formulas that we used for cost and PAR calculations.

$$Cost = \Sigma_{hour=1}^{24}(EP_{Rate}^{hour} * P_{Rate}^{App}) \tag{1}$$

Total un-schedule load formula is:

$$l_{od} = P_{Rate}^{App} * App \tag{2}$$

PAR is calculated by using the formula:

$$PAR = \frac{max(l_{od}^{S})}{Avg(l_{od}^{S})} \tag{3}$$

Table 1 Description of appliances

Group	Appliances	Power Rating (KWh)	Daily usage (Hours)
Shiftable interruptible	Vacuum cleaner	0.7	6
	Water Heater	4	8
	Water Pump	0.8	8
	Dish washer	1.5	10
	Dish washer	1.5	10
	Hair Dryer	1.2	4
Shiftable un-interruptible	Washing Machine	0.7	5
	Cloth Dryer	4	4
Regular Appliances	AC	1.5	15
	Refrigerator	0.18	14
	Oven	2	7

3.2 Heuristic Algorithms

Heuristic algorithms used in our scheme are BFOA, GA, BPSO and WDO. Implementation detail of each algorithm is described as follow, Table 2 shows nomenclature of algorithms.

3.2.1 BFOA

BFOA is well known optimization algorithm inspired by social foraging behavior of real bacteria (Escherichia coli). In which bacteria swims or tumble in search of nutrients and find best nutrients (solutions) to maximize its energy. To solve optimization problems, BFOA has a four principle mechanisms (Chemotaxis, Swimming, Reproduction and Elimination-dispersal) that are observed in real bacterial system. In our BFOA based HEM system, parameters (N_e, N_r, N_c, N_p, N_s, C_i, Θ, Ped, maximum generation) of BFOA are initialized. After initialization of parameters three loops are initialized elimination-dispersion, reproduction and chemotaxis processes respectively. Under chemotaxis loop initial fitness is evaluated and then our system compute new positions of bacteria (solution matrix). Mathematically chemotaxis movement of bacterium can be represented by [7]

$$\theta_i(j,k,l) = \theta_i(j-1,k,l) + C_i \frac{\Delta_i}{\sqrt{\Delta_i^T \Delta_i}} \tag{4}$$

Here $\Delta = (rand(1,D) - ped) \times 2$ and $\theta_i(j,k,l)$ represents the position of i^{th} bacterium at j^{th} chemotactic, k^{th} reproductive and l^{th} elimination-dispersal step. C_1 is the size of step taken in random direction. In next step swimming loop is initialized to find current best solutions in which position of bacteria is updated and evaluated until stoping criteria is met. Fitness function for evaluation in this step is mathemat-

Table 2 Symbols used in BFOA, GA, BPSO and WDO

Symbol	Description
EP^{hour}_{Rate}	electricity price set by utility
$*P^{App}_{Rate}$	Power rating of appliance
θ	Position of bacteria
Δ	Random direction vector
N_e	Number of elimination steps
N_r	Number of reproduction steps
N_c	Number of chemotaxis steps
N_s	Number of swimming steps
N_p	Number of population steps
j_i	Fitness level of bacteria
C_i	Step size
D	Dimension if search space in BFOA
$vmax$	Upper velocity limit in BPSO and WDO
$vmin$	lower velocity limit in BPSO and WDO
P_c	Probability of crossover
P_m	Probability of mutation
v_{new}	Velocity in current iteration
v_{old}	Velocity in previous iteration
x_{old}	Position in previous iteration
w, c_1, c_2	Weighted functions
w_i	Initial weighted function
w_f	Finial weighted function
F_{sig}	Sigmoid function
k	Current iteration
k_{max}	Maximum number of iterations
ΔP	Gradient of pressure
δV	Volume
ρ	Density of air parcels
α	Friction coefficient
ω	Earth rotation
u	Velocity of air parcel
u_{cur}	Velocity of air parcel in current iteration
u_{new}	Velocity of air parcel in new iteration
p_{opt}	Pressure at optimal location
p_{cur}	Pressure at current loacation
x_{opt}	Optimal location of air parcel
x_{cur}	Current location of air parcel
g	gravitational constant
RT	universal gas constand
n	total number of appliances

ically defined as:

$$J_i(j,k,l) = J_i(j,k,l) + J_{cc}(\theta_i(j,k,l), POP[j,k,l]) \tag{5}$$

When swimming step complete its iteration reproduction loop is started in which only healthy bacteria survive for the production of next generation. Least healthy bacteria die and survived bacteria split into two bacteria which keep the swarm size constant. In elimination-dispersion step cells are discarded and new random samples are inserted with a low probability.

3.2.2 GA

GA is heuristic optimization technique inspired by genetic process of living organisms in which new genres are formed which carry the properties of their parents. In GA initial population is randomly generated which consists of N number of chromosomes, where each chromosome hold the solution of given problem. In our GA based HEM system, scheduling of appliances is as follow.
Parameters (Population size, n, Number of iterations, P_c, P_m) of GA are initialized. In next step initial population is generated. Population is a set of solutions that shows the status of each appliance for the given time slot. As these solutions are randomly generated which may or may not satisfied required objectives, so they are evaluated by fitness function (objective function). According to the results of fitness function current best solutions are recorded. On the basis of these current best solutions a new steam of population is generated. Crossover and mutation is applied on this new population. In crossover binary strings are crossovers from two parents and form two new off springs. one point crossover is used in our work as shown in Fig. 1

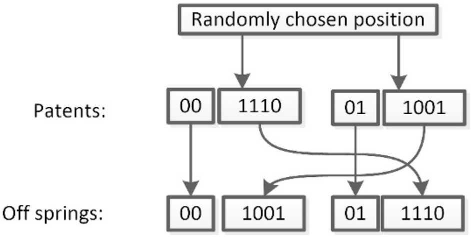


Fig. 1 single point Crossover

A larger the crossover rate mean fast convergence rate so the best crossover is for optimization problem is found to be

$$P_c = 0.9 \tag{6}$$

To create randomness in the results so that repetition of population can be avoided we use mutation process. It changes one or more principles gene in a chromosome from its initial state. Probability of mutation is very low and can be found by the following formula:

$$p_m = \frac{1 - P_c}{10} \tag{7}$$

So for better results from GA value of crossover rate is high while mutation rate is low. once crossover and mutation are done again a population is generated and fitness is evaluated. This whole process keep on repeating until stoping criteria is met.

3.2.3 BPSO

BPSO is binary variant of PSO. It is a nature-inspired optimization technique based on bird flock in search of food. When birds move for food they have some specific positions and velocities by which they move.
In BPSO based HEM system, parameters (Swarm size, n, Number of iterations, c_1, c_2, w_i, w_f, vmax, vmin) are initialized. Initial position matrix is generated randomly, also initial velocity is generated by using following formula:

$$v_i = vmax * 2 * (rand(swarm, n) - 0.5); \tag{8}$$

Here, position matrix is a solution matrix which shows the status of appliances. This initially generated position matrix is evaluated by fitness function (objective function) and best values of position are achieved then velocity of each particle is updated. So in case of BPSO we control position and velocity and define their best values for current iteration called pbest, then collection of pbest values from all iterations are evaluated by fitness function to find global best (gbest). In our work velocity is updated by the following equation [19].

$$\begin{aligned} v_{new} = w \times v_{old} + c_1 \times rand(1) \times (pbest - x_{old})... \\ + c_2 \times rand(1) \times (gbest - x_{old}) \end{aligned} \tag{9}$$

In the above equation the weighted factor w is calculated the following equation

$$w = w_i + \frac{(w_f - w_i) \times k}{k_{max}} \tag{10}$$

Once the velocity is obtained it is converted into binary by using sigmoid (S_g) function which is define as

$$Sig(j,i) = \frac{1}{1 + e^{-v_{new}}} \tag{11}$$

By applying equation 11 position matrix is updated as follow

$$\begin{cases} x_{new} = 1 & ifrand(1) \leq Sig(j,i) \\ x_{new} = 0 & ifrand(1) > Sig(j,i) \end{cases} \tag{12}$$

Then fitness is evaluated on new position, this process will continue until stoping criteria is met. At the end of this process we get ”gbest” values which are optimum solution for scheduling of our appliances.

3.2.4 WDO

Wind Driven Optimization (WDO) is a nature-inspired optimization technique based on atmospheric motion of wind. When wind blows, it equalize the horizontal imbalances. In this algorithm we consider small air parcels that are moving in N dimensional space which experiences different type of forces which effects their velocity and pressure.

In our WDO based HEM system, parameter (Population size, n, Number of iterations, RT, g, α, dimMin, dimMax, vmax, vmin) of WDO are initialized. In next step initial position matrix and initial velocity are generated randomly, velocity is generated by using following formula:

$$v_i = vmax \times 2 \times (rand(populationsize, npar) - 0.5); \tag{13}$$

Here, position matrix is a solution matrix which shows the status of appliances. This initially generated position matrix is evaluated by fitness function (objective function) and best values of position are achieved then velocity of each particle is updated. In WDO the pressure is a fitness function which is used for evaluation of objective function. As we use different iteration to find optimal solution, so instead of using actual values of pressure which make velocity function impractically large, we use ranking based approach. In this approach population of air parcels are ranked on the basis of pressure values in descending order so optimal pressure value has rank 1 which is a global best. In our scheme equation used for updating velocity is as follow [21]:

$$\begin{aligned} u_{new} = (1-\alpha)u_{cur} + gx_{cur} + (\frac{RT}{p_{cur}}|\frac{1}{i} - 1|)\times \\ (x_{opt} - x_{cur}) + (\frac{cu_{cur}^{otherdim}}{i}) \end{aligned} \tag{14}$$

Constants (g,c,RT and) in equation 14 are defined in initialization phase. At each iteration velocity and pressure values must be updated. By using following equation new position of air parcel for each iteration is defined,

$$x_{new} = x_{cur} + u_{new}\Delta t \tag{15}$$

Here Δt is step time which is equal to 1, velocity at each iteration must be bounded by its maximum and minimum values which are define as

$$\begin{cases} u_{new} \ = u_{max} \quad if u_{new} > u_{max} \\ u_{new} = -u_{max} \quad if u_{new} < u_{max} \end{cases} \tag{16}$$

After updating velocity function again new ”position” matrix is generated and evaluated. This process will continue until stoping criteria is met. At the end of this process we get ”gbest” values which are optimum solution for scheduling of our appliances.

Conclusion

Heuristic algorithms: BFOA, GA, BPSO and WDO along with their formulation have been discussed in detail. Applications of all algorithms have been discussed that how they can address the residential appliances’ scheduling problem. From these discussions, it is concluded that every optimization technique has the capability to reduce cost and PAR as compared to unscheduled load. Scheduling of HEM system is a complex task, however, using heuristic techniques this issue can be resolved. It requires proper scheduling of different residential appliances, which efficiently reduces cost and PAR. Cost reduction relieves the consumer while PAR reduction relieves the utility.

References

1. Deng, Ruilong, Zaiyue Yang, Mo-Yuen Chow, and Jiming Chen. “A survey on demand response in smart grids: Mathematical models and approaches.” IEEE Trans. Industrial Informatics 2015, 11, 570-582
2. Nguyen, Hung Khanh, Ju Bin Song, and Zhu Han. “Distributed demand side management with energy storage in smart grid.” IEEE Trans. Parallel and Distributed Systems 2015, 26, 3346-3357.
3. Bozchalui MC, Hashmi SA, Hassen H, Canizares CA, Bhattacharya K. “Optimal operation of residential energy hubs in smart grids.” IEEE Trans. Smart Grid 2012, 3, 1755-1766.
4. Roh, Hee-Tae, and Jang-Won Lee. “Residential demand response scheduling with multiclass appliances in the smart grid.” IEEE Trans. Smart Grid 2016, 7, 94-104.
5. Fernandes, Filipe, Tiago Sousa, Marco Silva, Hugo Morais, Zita Vale, and Pedro Faria. “Genetic algorithm methodology applied to intelligent house control.” Computational Intelligence Applications In Smart Grid (CIASG), 2011 IEEE Symposium on. IEEE, 2011.
6. Pedrasa, Michael Angelo A., Ted D. Spooner, and Iain F. MacGill. “Scheduling of demand side resources using binary particle swarm optimization” IEEE Trans. Power Systems 2009, 24, 1173-1181.
7. Das, Swagatam, Arijit Biswas, Sambarta Dasgupta, and Ajith Abraham “Bacterial foraging optimization algorithm: theoretical foundations, analysis, and applications.” Foundations of Computational Intelligence 2009, 3, 23-55.

8. Rasheed, Muhammad Babar, Nadeem Javaid, Ashfaq Ahmad, Zahoor Ali Khan, Umar Qasim, and Nabil Alrajeh. “An Efficient Power Scheduling Scheme for Residential Load Management in Smart Homes.” Applied Sciences 2015, 5, 1134-1163.
9. Maringer, Dietmar G. “Portfolio management with heuristic optimization.” Springer Science and Business Media 2006, 8
10. Shirazi, Elham, and Shahram Jadid. “Optimal residential appliance scheduling under dynamic pricing scheme via HEMDAS.” Energy and Buildings 2015, 93, 40-49.
11. Chrostopher O. Adika, Lingfeng Wang “Smart charging and appliance scheduling approaches to demand side management” Electrical Power Systems 2014, 57, 232-240.
12. John S. Vardakas, Nizar Zorba, Christos V. Verikoukis “Power demand control scenarios for smart grid applications with finite number of appliance” Applied Energy 2016, 162, 83-98.
13. Sareen Althaher, Pierluigi Mancarella, Joseph Mutale “Automated Demand Response from Home Energy Management System Under Dynamic Pricing and Power and Comfort Constraints” IEEE Trans. Smart Grid 2015, 6, 1874 - 1883
14. Raj, Joshua Samuel, and S. Devi Priya. “Contribution of BFO in grid scheduling.” Computational Intelligence & Computing Research (ICCIC), IEEE International Conference on. IEEE, 2012.
15. Jun Li, Jianwu Dang, Feng Bu, Jiansheng Wang “Analysis and improvement of the bacterial foraging optimization algorithm.” Journal of Computing Science and Engineering 2014, 8, 1-10.
16. Oladeji, Olamide, and O. O. Olakanmi. “A genetic algorithm approach to energy consumption scheduling under demand response.” 6th International Conference on Adaptive Science and Technology (ICAST). IEEE, 2014.
17. Ten, Viktor, Zhandos Yessenbayev, Akmaral Shamshimova, and Albina Khakimova. “Optimized Small-Scaled Hybrid Energy Management of a Smart House Based on Genetic Algorithm” 14th International Conference on Machine Learning and Applications (ICMLA). IEEE, 2015.
18. Carpinelli, Guido, Shahab Khormali, Fabio Mottola, and Daniela Proto. “Optimal integration of distributed energy storage devices in smart grids.” IEEE Trans. smart grid 2013, 4, 985-995.
19. Del Valle Y, Venayagamoorthy GK, Mohagheghi S, Hernandez JC, Harley RG. “Particle swarm optimization: basic concepts, variants and applications in power systems.” IEEE Trans. evolutionary computation 2008, 12, 171-195.
20. Faria, Pedro, Joo Soares, Zita Vale, Hugo Morais, and Tiago Sousa. “Modified particle swarm optimization applied to integrated demand response and DG resources scheduling.” IEEE Trans. Smart Grid 2013, 4, 606-616.
21. Bayraktar, Zikri, Muge Komurcu, Jeremy A. Bossard, and Douglas H. Werner. “The wind driven optimization technique and its application in electromagnetics.” IEEE trans. antennas and propagation 2013, 61, 2745-2757.

Network lifetime maximization via energy hole alleviation in wireless sensor networks

Muhammad Awais Khan, Arshad Sher, Ahmad Raza Hameed, Naeem Jan, Junaid Shabir Abassi, Nadeem Javaid*
COMSATS Institute of Information Technology, Islamabad 44000, Pakistan

Abstract Energy hole creation is one of the most important issues in Wireless Sensor Networks (WSNs). This paper aims to analyze the energy hole boundary for avoiding the creation of energy hole such that network lifetime is prolonged. An analytical model is presented to analyze the network lifetime and location of energy hole from the start of network till the death of last node. Also network area is logically divided to minimize data loss.

1 Introduction

Wireless Sensor Networks (WSNs) have the capability to sense and compute the environmental changes. WSNs are used in many applications like military surveillance, medical diagnosis, pollution monitoring, industrial applications [1] and in wireless communication [2]-[4]. WSNs mainly consist of large number of sensor nodes that are deployed in a given area. These sensor nodes powered with batteries sense the monitoring area and send the collected information to the base station (sink). Direct and multi hop transmission are the two major communications modes in WSNs. Sensor nodes far away from the sink consume high energy in direct transmission mode and die at an earlier stage, while in multi hop transmission mode, sensor nodes near the sink die at an earlier stage due to heavy relaying of farther nodes data. This creates energy hole around the sink. With the presence of energy hole nodes are unable to send data to the sink. Excess amount of energy is wasted which effects the system performance. Results in [5] shows that approximately 90 percent of network energy is left unused when the network is finished in case of uniformly distributed network.

*Corresponding author:
COMSATS Institute of Information Technology, Islamabad 44000, Pakistan.
Emails: nadeemjavaid@comsats.edu.pk, nadeemjavaidqau@gmail.com
Website: www.njavaid.com

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_26

Energy hole has become a major issue in WSNs because it directly impact the network lifetime. Olariu, *et al.*[6] highlight the scenarios of energy hole problems. However, they never mention the factors which leads to energy hole problems. In [7], the authors spotlight the conditions under which the energy hole emerges in the network. Kacimi *et al* lessen the energy hole problem by introducing Load Balancing Technique (LBT) [8]. The technique adjusts the transmission power of sensor nodes which helps in balancing the energy consumption. The scheme is able to improve network lifetime and balances the energy consumption of sensor nodes. However implementing this technique on other scenario is a challenging task. The scheme balances the energy consumption of nodes by adjusting their transmission power. Most protocols try to tackle energy hole in cluster based WSNs. The scheme proposed in [9], [10] present an efficient routing mechanism that helps to overcome the energy hole around the sink as a result increases the network lifetime. However study relates that energy hole does not always emerges close to the sink. Lian et al. [11] propose the nonuniform node distribution strategy that helps to enhance the data capacity in WSNs. However the proposed schemes unable to provide theoretical analysis for their work to locate the energy hole.

In this paper, we alleviate energy hole with the help of sleep schedule mode. The analytical model presented in this paper aim to overcome energy hole problem. Similar to [12] we consider energy consumption for data transmitting, data receiving and energy for idle listening. The assumption in this paper minimizes the energy consumption of the sensor nodes.

Mainly our contribution in this paper are:

1) Analytical model of NEHA (Network lifetime maximization with Energy Hole Alleviation in Wireless Sensor Networks) to determine the traffic load, energy consumption and nodal lifetime of nodes.

2) Energy hole location based on nodal lifetime.

3) Compare the exiting technique with [12].

2 Related Work

Many researchers did work on exploring new techniques to alleviate the problems in WSNs. The work propose by Ozgovde, *et al.* [13] highlight the FNDT (First Node Died Time) and ANDT (All Node Died Time) as X-factor to improve network lifetime. They propose a utility based framework Weighted Cumulative Operational Time (WCOT) to measure the lifetime of the network on the bases of network states history. However the proposed work focuses on network lifetime, not able to highlight energy consumption in proper way. Authors in [14], focus on connection time of sensor nodes rather than working on entire network lifetime and energy consumption. Li, *et al.* in [15] introduce analytical model in data gathering WSNs. The network is divided into different ring sectors to analyze the network lifetime. The propose scheme maximizes network lifetime on the bases of annuli. However the scheme fails to control energy consumption of the network.

In WSNs, the energy hole problem minimizes the lifetime of the the network. In ACH2 [16], Ahmad, *et al.* improve network lifetime and energy consumption by introducing a mechanism of node association with the cluster head. In this scheme, selection of CHs are based upon optimal distance between them. However the first CH is selected on the basis of its threshold value. The proposed scheme improves the network lifetime and energy consumption, however unable to locate the energy hole problem. Similarly, [17] routing protocol is proposed to overcome shortcomings of [8]. It achieves better network lifetime due to non-uniform node distribution, still the proposed scheme are unable to recover from void regions. Data aggregation is one of the most useful paradigm in WSNs. The aim is to overcome unnecessary transmission of data. Energy Balancing Strategy (EBS) is discussed in [18], to improve the network performance. In [19], energy hole problem is alleviated through compressing and aggregating the data. Still, no one is able to provide efficient information about energy hole occurrence. Lin, *et al.* propose an Energy Efficient Ant Colony (EEAC) algorithm to gather data from nodes [20]. In this scheme, each node uses remaining energy to find the probability of remaining next hop node. The scheme improves the network lifetime however, fail to control robustness and scalability issue.

An efficient energy hole alleviating algorithm [21] is suggested to overcome the energy hole problem in the network. The algorithm uses router selection strategy to recover the energy hole to some extent. However, lengthy process increases the end-to-end delay. At 90 m distance, the lifetime reaches up to 22000 rounds. Chang, *et al.* [22] propose a scheme for maximizing network lifetime with the selection of optimal link cost. However, unable to provide a complete solution to recover the energy hole problem. Authors in [23], propose a Balance Energy Efficient Routing protocol with time reliable communication (BERR) for balanced energy consumption throughout the network lifetime. The propose scheme chose minimum depth nodes and include retransmission phenomena to lessen the energy consumption as well as to achieve better reliability. The major flaw in this routing protocol is the selection of next hop node with minimum depth, moreover, there is no such criteria to recover the energy hole.

Ghaffari, *et al.* propose a new improved energy efficient routing protocol that chooses minimum hop neighbours to forward the data [24]. The scheme consider minimum depth nodes that have shortest path to the sink, it also checks the link quality as well as minimum number of hop. It improves the network lifetime in dense regions, yet in sparse regions, it hardly finds the next forwarder node which results in more energy consumption. Energy Efficient tree based Data Collection Protocol (EEDCP-TB) [25] is proposed to maximize the network lifetime. EEDCP-TB control flooding and allocates time schedule for data aggregation to save energy of the nodes. This scheme achieves better network lifetime at the cost of high delay.

Energy consumption is the the major factor which degrades the performance of network. An efficient routing mechanism is the need in this circumstances. Bhattachargee, *et al.* [26] introduce Lifetime Maximizing Dynamic Energy Efficient routing protocol (LMDEE) for multi hop WSNs that takes into account the remaining energy of sensor nodes to forward the data to the sink [27]. LMDEE performs better in term of network lifetime still it fails to cop with the network scalability and data redundancy issues. In [28], the authors propose an Energy Efficient Clustering technique that uses run time recovery mode of sensor nodes due to failure of cluster heads. This scheme helps in prolonging the network lifetime however, at the expense of more errors in dense regions. An Innovative Balance Energy Efficient and real time reliable communication routing protocol (IBEE) is suggested to utilize energy efficiently [29]. IBEE protocol chooses minimum distance nodes as next forwarder nodes and do retransmission to decreases energy consumption. However, efficient energy consumption is achieved with high delay.

Table 1 Used Notation

Notations	Definition
R	Network Radius (m)
r	Transmission range of sensor (m)
E_o	Initial energy of sensor nodes
B	Data transmission rate
ρ	Node density
d_o	Depth threshold
S_m	The m_{th} network stage
A_x	A small region close to the sink

3 NEHA: The Proposed Scheme

In our work, we consider the following aspects:

Network and Energy Consumption Model:

Let sensors nodes be randomly deployed and uniformly distributed in a circular field with centrally positioned sink [31]-[32]. The transmission range of sensor nodes is r and R is the radius of the network. Sensor nodes send data to the sink in a data period (round). They are allowed to forward the received data through direct transmission or via multi hop transmission. Minimum depth neighbours are selected as data forwarder. Radio model in [33], is considered in this paper. Energy consumed in transmitting and receiving k bits data is shown in equation (1) and (2) respectively:

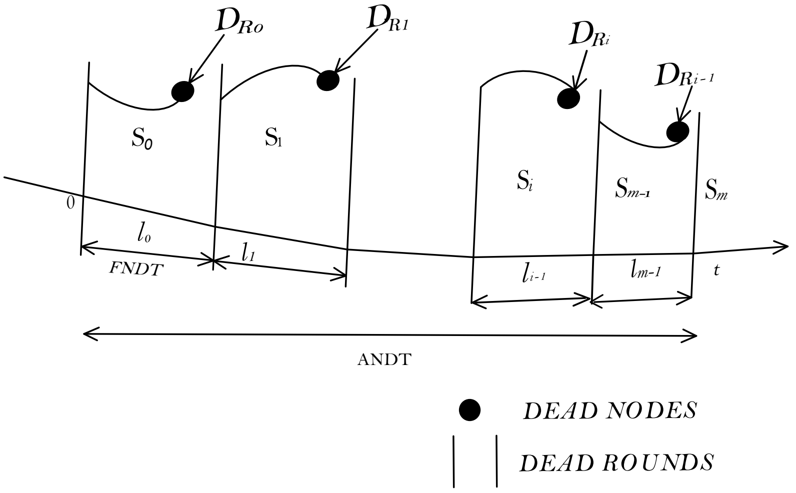


Fig. 1 Entire Process Scenario

$$E_t = \begin{cases} kE_{elec} + k\varepsilon_{fs}d^2 \ if \ d \leq d_o \\ kE_{elec} + k\varepsilon_{amp}d^4 \quad otherwise \end{cases} \tag{1}$$

$$E_r = kE_{elec} \tag{2}$$

Where d_o is the threshold distance and E_{elec} denotes transmitting circuit loss. We adopt free space channel model and multipath fading channel model that are used in [34]. d denote the transmission distance. If d is less than or equal to d_o free spaced channel model is used otherwise, if distance d exceeds depth threshold, we adopt the multipath fading channel model. ε_{fs} denotes the energy for power application in free space channel model [30] while, ε_{amp} denotes the energy for multipath fading channel model.

Sensor nodes send the sense data to sink using greedy geographic routing. Sensor nodes transmit data to their neighboring nodes which are closest to the sink [12]. Nodes transmit and receive data mainly in active mode while no communication occurs between the nodes in sleep mode [12]. Energy is consumed in data transmitting and receiving while energy consumption in sleep mode is negligible. Our initial focus is to find out the location of dead nodes. Figure. 2 shows network forwarding model of the proposed protocol.

Dividing the network into small regions such that to balance the energy consumption in each region [12]. A_x denote a small region close to the sink ε denotes the width of the region and θ be the angle form by the region with the sink. $\{A_{x+r}, A_{x+2r}\}$ denote the upstream regions, r denotes transmission range of sensors. A_{x+r} is supposed to forward data to the region A_x because distance between A_{x+r} lies in transmission range of A_x. Also A_{x+r} relays the data of A_{x+2r} because distance be-

Fig. 2 Data Transmission Model at S_o

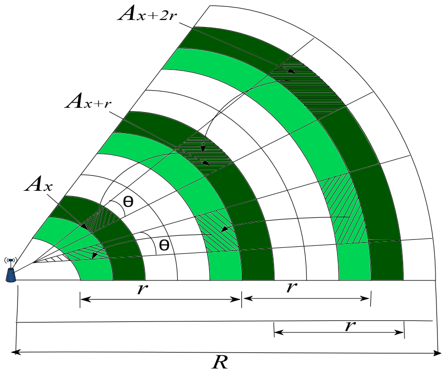
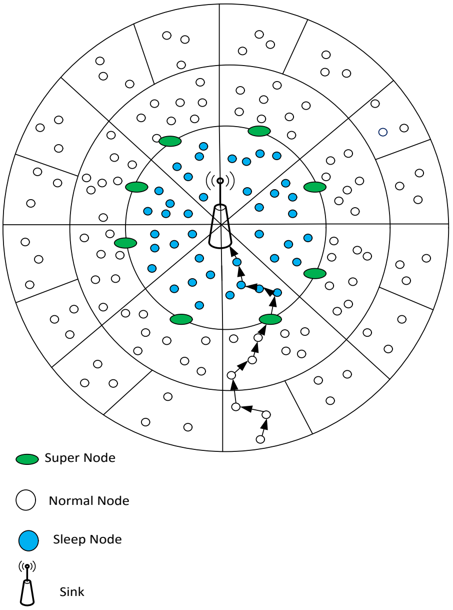


Fig. 3 Data Transmission Model



tween A_{x+r} and A_{x+2r} is equal or close to r. In other words nodes lie in upstream region will transmit data to its downstream region nodes if their distance d satisfies $d < r$ as shown in Fig. 2.

If divided region is too short, there exist few number of nodes or region contain no node. However, when divided region is large, nodes suffer data loss due to large network field which requires high energy. Nodes distance from the sink in A_x region is equal or close to x.

To cop with the network area issue, we have approximated the network area at the start of transmission. As, A_x denote the small region near the sink. $x > \varepsilon$, the volume of network field is approximated as $Z_{A_x} = (x - \varepsilon)\theta\rho$ else $Z_{A_x} = (\varepsilon - x)\theta\rho$.

Fig. 3 shows the complete network scenario of the proposed scheme NEHA. The process consist of two phase:

(i) Initialization phase and **(ii) Data sharing and collecting phase**

In **Initialization phase**, data packets are broadcasted to all the nodes that take part in the data sharing phase. During the start of the network each node transmit the received data to the neighbour nodes if they lie in the transmission range of the node. Nodes, after receiving the data look for a neighbour node to send the received data to the sink through the multihop method. Node chooses minimum depth nodes as a data forwarder. In the next phase i.e. **Data sharing and collecting phase**, the received data are collected by the neighbour node and send this data to the downstream region nodes. When first packet is received by the super node it sends an ACK message to the downstream region nodes to go into sleep mode so that to the received data do not reach to the sink nodes i.e. normal nodes are unable to receive and transmit the data. After that only the super nodes present at the boundary of the sink region are able to receive the data from the upstream nodes but unable to send this data to the normal nodes. When all the data successfully reaches to super nodes, then super nodes again send an ACK message to the normal nodes to come back to their initial position i.e. to wake up from the sleep phase. After this, all the collected data are received by the normal nodes ad send it to the sink to elevate the energy hole problem.

3.1 Data Amount at Stage S_o

S_o denotes the stage at which none of the nodes die in the the network. We find the FNDT from [12] at stage S_o as follows:

$$FNDT = E_o/max({e_x}^0) \tag{3}$$

Where E_o is the initial energy of sensor nodes and ${e_x}^0$ is the total energy of the sensor node including energy for data transmitting, receiving and idle listening at stage S_o.

We calculate the traffic load of sensor nodes on the basis of analytical model [12].

Theorem 1 *Let node i be in a small region A_x with width ε. The distance between A_x and the sink be x and θ be the angle formed by the region A_x with the sink. If i generates one data packet per round then the average data amount sent by i in a single round at stage S_o is:*

$$p_i^0 = \begin{cases} (Z_1+1)+Z_1(Z_1+1)(\varepsilon-r)/2(x-\varepsilon) \; if \; x \leq \varepsilon \\ (Z_2+1)+Z_2(Z_2+1)(\varepsilon-r)/2(\varepsilon-x) \; if \; \varepsilon > x \end{cases} \tag{4}$$

where Z_1 = $(R-x)/r$ and Z_2 = $(R-\varepsilon)/r$ [12] .

Proof: Since A_x is the small region close to the sink and node i is in the region A_x so we are able to calculate traffic load in the region. As ε denotes the width of *the region and θ is the angle formed by the region A_x with the sink and x is the distance of the region A_x from the sink. As nodes in region A_x relay the data of nodes in the upstream region and forward the collected data to the sink so, traffic load is calculated. As A_{x+ir} denotes the upstream region and sensor nodes located in this region are upstream nodes then according to the equation (4) area of upstream region is be approximated as:*

$$N_{A_x} = \begin{cases} ((x-\varepsilon)+ir)\varepsilon\theta\rho \ if \ x \leq \varepsilon \\ ((\varepsilon-x)+ir)\varepsilon\theta\rho \ if \ \varepsilon > x \end{cases} \tag{5}$$

Since, data is generated by sensor nodes and in each round sensor nodes sends one data packet so the total number of data packets must be equal to the number of nodes involve in the process. We write the equation (6) as the sum of data packets from the upstream region is the total data packets on A_x [12].

$$D_{A_x} = N_{A_x} + N_{A_{x+r}} + ... + N_{A_{x+zr}} \tag{6}$$

So, the average traffic load on node i in a region A_x must be equal the ratio of number of data packets sent to the total number of nodes involved i.e. ${p_i}^0 = D_{A_x}/N_{A_x}$ [12]. We find the traffic load at stage S_o of node i, by some doing some arithmetic operation we have ${p_i}^0$ as (4).

3.2 Energy Consumption at Stage S_o

If each node generates λ bits packets then the total amount of data transmitted is ${p_x}^0\lambda$. Energy consumption at stage S_o is calculated according to theorem given below:

Theorem 2 *Let node i lies in the region A_x, x denotes the distance between the region A_x and the sink. If a sensor node transmit data at rate B bits/sec then the average energy consumed by the i in a data round is ${e_i}^0$ = ${e_{i,r}}^0$ + ${e_{i,t}}^0$ + ${e_{i,j}}^0$ [12].*

$$\begin{cases} e_{i,r^0} = ({p_x}^0 - 1)\lambda E_{elec} \\ e_{i,t^0} = {p_x}^0\lambda(\varepsilon_k d^\beta + E_{elec} \\ e_{i,j^0} = E_{idle}(A_t + \lambda/B(1-2{p_x}^0) \end{cases} \tag{7}$$

Where $d = r$ if $x > r$ else $d = x$. For ε_k = ε_{fs} distance must be greater than or equal to depth threshold i.e. $d \geq d_o$ and β = 2 in this case [12], else ε_k = ε_{amp} and β = 4.

.

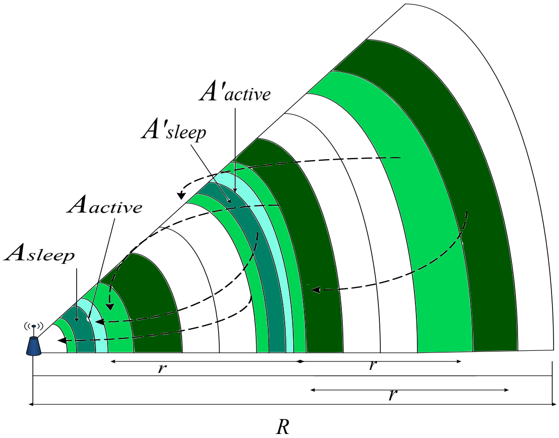
3.3 Avoiding Energy Hole Problem

Energy hole problem is the fundamental aspect in WSNs due to this problem network lifetime tend to decrease with the passage of time. To avoid this, the aim is to cover the loss of dead nodes in order to maximize network lifetime. Nodes in the region close to sink exhaust their energy quickly and die so nodes in upstream region are unable to share data to nodes in downstream region. In LAEHA, this problem is covered by assuming the nodes death region as dead region in which all nodes are dead. The data that are supposed to forward by the dead region are now forwarded by the upstream region that is A_{hot}. We are not allowed to use nodes that are dead during certain rounds as a result we are unable to receive data from that nodes. To cop with this, we set sleep schedule mode by placing nodes at the boundary of the region closed to the sink. We called these nodes as super nodes. They have enough energy to bear the traffic of upstream region nodes. Nodes in the region close to sink are in sleep mode they are unable to send or receive any data from upstream node. Super nodes function is to gather the data from upstream node. The data is transmitted through direct hop or multi hop depends on the scenario whether nodes are in range of super nodes or not. During the start of network nodes send data to their respective nodes. The nodes chooses minimum depth nodes as their neighbor to forward their data. Nodes lie in A_{sleep} send their data to nodes lie in A_{sleep} region, as the nodes in A_{sleep} region are in sleep mode they are unable to receive any data from the upstream nodes so this data is now further gather by the super nodes which further sends this gathered data to the nodes in A_{sleep} region. After all the data is successfully received by these nodes. Super nodes send the message to A_{sleep} nodes to become active. This process decreases the node death probability as a result network lifetime is increased.

4 Operation

Let A_{sleep} denotes the region in which all nodes are in sleep mode, no data is transferred through this region while A_{active} denotes the region in which all nodes are active sending data to its neighbour node through direct transmission or multi-hop transmission. Nodes located in A'_{active} region send the collected data to its downward region A_{active} nodes. While, A'_{sleep} nodes which are in sleep position when nodes in the region become active, they send the data to their downward region nodes i.e. A_{sleep} nodes. As nodes in A'_{sleep} region are supposed to forward the collected data to its downward region A_{sleep} nodes, however, the nodes in A_{sleep} region are in sleep mode they are unable to send or receive the data. Thus, A_{active} nodes collects the data of A'_{sleep} and send the data to the super nodes on the boundary of A_{died}. They transfer this collected data to the sink when nodes in A'_{sleep} region become active. Transmission model is shown in Fig. 5.

Fig. 4 Data Transmission Model of NEHA after S_o



\

Algorithm 1 Calculating the Traffic Load, Energy Consumption and Network lifetime after stage S_o

```
Input: Network range $R$, transmission range $r$ between sensor, $N$ number of nodes, node density $\rho$ etc.
Output: for a node $i \in N$ determine the traffic load ${p_i}^m$ and energy consumption ${e_i}^m$ at stage $S_m$
for each node $j \in N$ and $j = \{1,2,3,.....N\ \}$ at stage $S_o$ calculate the traffic load $[{p_1}^0, {p_2}^0, {p_3}^0, ....., {p_N}^0]$ and energy consumption $[{e_1}^0, {e_2}^0, {e_3}^0, ....., {e_N}^0]$
$m = 0$;
for stage $m+1$
Calculate distance $d$ of node i with super nodes
if $d$ is less than or equal to transmission range of super node, send packet to the super node.
while sink nodes are in sleep mode.
super node continuously gathering the data of nodes in a data period.
if time expires:
Send the collected data to the sink node if lies in transmission range $r$.
while sink receive data do
Calculate lifetime of senor nodes at stage $S_{m-1}$ i.e $l^{m-1}$
Calculate the overall traffic load, energy consumption and lifetime of nodes at stage $S_m$ i.e. $[{p_1}^m, {p_2}^m, {p_3}^m, ....., {p_N}^n]$, $[{e_1}^m, {e_2}^m, {e_3}^m, ....., {e_N}^m]$ as in theorem (1) and (2) for $m$ number of stages.
```

5 Conclusion and Future Work

In this paper, we have proposed NEHA routing protocol for WSNs. An analytical model is used in our technique to calculate the traffic load at each stage of the network field. The shortcomings of LAEHA are addressed in more efficient manner. Moreover, the proposed protocol alleviate the death probability of nodes due to super nodes located at the boundary of the sink region. An energy hole problem is minimized with scheduled sleeping mechanism in the network. The small division of area decreased the number of nodes while reliability of data packet is increased. In future, we have planned to improve network performance with cluster based routing. In order to recover void nodes zone based routing will be considered.

References

1. Yang, Qinghai, Yingji Zhong, Kyung Sup Kwak, and Fenglin Fu. “Outage probability of opportunistic amplify-and-forward relaying in Nakagami-m fading channels”. ETRI journal 30, no. 4 (2008): 609-611.
2. C. Tung, F. Tsang, L. Lam, Y. Tung, S. Li, F. Yeung, T. Ko, H. Lau, and V. R., “A mobility enabled inpatient monitoring system using a zigbee medical sensor network”, Sensors, vol. 14, no. 2, pp. 23972416, 2014.
3. J. Ren, Y. Zhang, and K. Liu, “An energy-efficient cyclic diversionary routing strategy against global eavesdroppers in wireless sensor networks”,Inter. J. Distr. Sensor Netw., vol. 2013, pp. 116, 2013.
4. Y. Chen and Q. Zhao, “On the lifetime of wireless sensor networks”, IEEE Commun. Lett., vol. 9, no. 11, pp. 976978, Nov. 2005.
5. S. Olariu and I. Stojmenovic, “Design guidelines for maximizing lifetime and avoiding energy holes in sensor networks with uniform distribution and uniform reporting”, in Proc. IEEE INFOCOM, 2006, pp. 112.
6. M. Perillo, Z. Cheng, and W. Heinzelman, “On the problem of unbalanced load distribution in wireless sensor networks”, in Proc. IEEE GlobeCom Workshops. 2004, pp. 74 79.
7. R. Kacimi, R. Dhaou, and A. Beylot, “Load balancing techniques for lifetime maximizing in wireless sensor networks” Ad Hoc Netw., vol. 11, no. 8, pp. 21722186, 2013.
8. J. Li and P. Mohapatra, “Analytical Modeling and Mitigation Techniques for the Energy Hole Problems in Sensor Networks”, Pervasive and Mobile Computing, vol. 3, no. 8, pp. 233-254, 2007.
9. S. Olariu and I. Stojmenovic, “Data-Centric Protocols for Wireless Sensor Networks”, Handbook of Sensor Networks: Algorithms and Architectures, I. Stojmenovic, ed., John Wiley and Sons, pp. 417-456, 2005.
10. J. Li and G. AlRegib, “Network lifetime maximization for estimation in multihop wireless sensor networks”, IEEE Trans. Signal Process., vol. 57, no. 7, pp. 24562466, Jul. 2009.
11. J. Lian, K. Naik, and G. Agnew, “Data capacity improvement of wireless sensor networks using non-uniform sensor distribution”, to appear in International Journal of Distributed Sensor Networks.
12. Ren, Ju, Yaoxue Zhang, Kuan Zhang, Anfeng Liu, Jianer Chen, and Xuemin Sherman Shen. “Lifetime and energy hole evolution analysis in data-gathering wireless sensor networks”, IEEE Transactions on Industrial Informatics 12, no. 2 (2016): 788-800.
13. A. Ozgovde and C. Ersoy, “Wcot: A utility based lifetime metric for wireless sensor networks”, Comput. Commun., vol. 32, no. 2, pp. 409 418, 2009.
14. J. Lee, B. K., and C. Kuo, “Aging analysis in large-scale wireless sensor networks”, Ad Hoc Netw., vol. 6, no. 7, pp. 1117 1133, 2008.
15. K. Li, “Optimal number of annuli for maximizing the lifetime of sensor networks”, J. Para. Distri. Comput., vol. 74, no. 1, pp. 17191729, 2014.
16. Ahmad A, Javaid N, Khan ZA, Qasim U, Alghamdi TA. “ACH2: Routing Scheme to Maximize Lifetime and Throughput of Wireless Sensor Networks”, Sensors Journal, IEEE, 2014.
17. A. Liu, X. Jin, G. Cui, and Z. Chen, “Deployment guidelines for achieving maximum lifetime and avoiding energy holes in sensor network”, Inform. Sci., vol. 230, pp. 197226, 2013.
18. M. Haenggi, “Energy-Balancing Strategies for Wireless Sensor Networks”, Proc. Intl Symp. Circuits and Systems (ISCAS 03), pp. 828-831, 2003
19. J. Li and P. Mohapatra, “Analytical Modeling and Mitigation Techniques for the Energy Hole Problem in Sensor Networks”, Pervasive and Mobile Computing, vol. 3, pp. 233-254, 2007.
20. C. Lin, G. Wu, F. Xia, M. Li, L. Yao, Z. Pei, “Energy Efficient Ant Colony alogorithm fordata aggregation in wireless sensor networks”, J Comput Syst Sci 12; 78: pp. 1686-702 ,2012. S. Olariu and I. Stojmenovic, “Design Guidelines for Maximizing Lifetime and Avoiding Energy Holes in Sensor Networks with Uniform Distribution and Uniform Reporting”, Proc. IEEE INFOCOM 06, pp. 1-12, 2006.

21. Xue, Yu, Xiangmao Chang, Shuiming Zhong, and Yi Zhuang. “An efficient energy hole alleviating algorithm for wireless sensor networks”, IEEE Transactions on Consumer Electronics 60, no. 3 (2014): 347-355.
22. J.-H. Chang and L. Tassiulas, “Maximum Lifetime Routing in Wireless Sensor Networks”, IEEE/ACM Trans. Networking, vol. 12, pp. 609-619, 2004.
23. Liu Z-x,Dail-l, Kaim, Guanx-p. “Balance energy-efficient and real-time with reliable communication protocol for wireless sensor network”, J China Univ Posts Telecommun2013; 20 : 3746.
24. Ghaffari A. “An energy efficient routing protocol for wireless sensor networks using a- star algorithm”, J Appl Res Technol 2014 2014; 12: 81522.
25. Jin, Yong-xian, Feng-zhen Chen, Gao-feng Che, and Wei Hu. “Energy-efficient data collection protocol for wireless sensor network based on tree”, In Wearable Computing Systems (APWCS), 2010 Asia-Pacific Conference on, pp. 82-85. IEEE, 2010.
26. S. Bhattacharjee, S. Bandyopadhyay, “Lifetime Maximizing Dynamic Energy Efficient routing protocol for multihop wireless sensor networks”, Simul Modell Pract Theory 2013; 7; pp. 15-29.
27. M. Azharuddin, P. Kuila, PK. Jana. “Energy Efficient fault toulerent clustering and routing algorithm for wirelsss sensor networks”, Computer and electrical engineering, 2014.
28. T. Liu, Q. Li. “An Energy-Balancing clustering approach for gradient based routing in wirelss sensor networks”, Comput Commun 2012a; 35; pp.2150-61.
29. M. Noori and M. Ardakani, “Lifetime analysis of random event-driven clustered wireless sensor networks”, IEEE Trans. Mob. Comput., vol. 10, no. 10, pp. 14481458, 2011.
30. K. Ota, M. Dong, Z. Cheng, J. Wang, X. Li, and X. Shen, “Oracle: Mobility control in wireless sensor and actor networks”, Comput. Commun., vol. 35, no. 9, pp. 1029 1037, 2012.
31. A. Chakraborty, R. Rout, A. Chakrabarti, and S. Ghosh, “On network lifetime expectancy with realistic sensing and traffic generation model in wireless sensor networks”, IEEE Sensors J., vol. 13, no. 7, pp. 2771 2779, 2013.
32. S. Lee and H. Lee, “Analysis of network lifetime in cluster-based sensor networks”, IEEE Commun. Lett., vol. 14, no. 10, pp. 900 902, 2010.
33. G. Anastasi, M. Conti, and M. Di, “Extending the lifetime of wireless sensor networks through adaptive sleep”, IEEE Trans. Industr. Informatics, vol. 5, no. 3, pp. 351365, 2009.

A Smart Card-Based Three-Party Quantum Key Distribution Protocol

Heri Wijayanto[1,2], Hsing-Chung Chen[1,3,*], Wen Yen Lin[4,*]

[1] Dept. of Computer Science and Information Engineering, Asia University
No. 500, Lioufeng Rd., Wufeng, Taichung, Taiwan 41354, R.O.C
[2] Dept. of Information Engineering, Mataram University
No. 62, Majapahit Rd., Mataram, Indonesia
[3] Dept. of Medical Research, China Medical University Hospital,
China Medical University, Taiwan
*Corresponding author‘s emails: shin8409@ms6.hinet.net, cdma2000@asia.edu.tw
[4] Dept. of Digital Multimedia Technology, Vanung University, Taoyuan, Taiwan
*Corresponding author‘s email: qqnice@gmail.com

Abstract. The quantum key distribution (QKD) protocols become the trend research in the computer security field today because those are very strong to prevent the eavesdropping during communication domain. In addition, an existence of the third party as the Trust Center (TC) presents the higher security level and the more effective key distribution such as 3QKDPMA and 3PAQKD-TB. The TC has responsibilities to construct the session key and to authenticate the user on a communication session. Moreover, the TC should have the high-security level to assure all communications done under the secure condition. The 3QKDPMA and 3PAQKD-TB provide the effective method for creating the session key. However, the basic QKD protocol mentioned in this paper do not provide the advanced user authentication. It is because TC only maintains the user's secret key table which is used to authenticate each legal user. To improve the basic QKD protocol, we propose a new scheme for QKD by elaborating the 3PAQKD-TB with the smart card secure user authentication. Finally, it is a lightweight computation approach in order to achieve more efficiently to the basic QKD protocol by using elliptic curve cryptography.

Keywords. User authentication; QKD, smart card; elliptic curve cryptography.

1 Introduction

The classic cryptography algorithms implement diverse mathematics and bit operations to provide a secure communication. On the other hand, the effort of quantum physic science produces an alternative method for secure communication that is called by quantum cryptography. Furthermore, it is claimed that has higher security level than classic cryptography because it solves the problem of sharing long session key for one-time pad encryption and also it resists of eavesdropping without disturbing the channel and detected easily [2,4]. The fundamental theoretical designs of quantum cryptography were proposed by Bennett et al. [2] in 1984 that name is

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_27

BB84 protocol and B92 protocol in 1992 [4]. Since that time, this field of study was attracted many researchers to improve the scheme and it has also been fabricated such as 8505 QKD by MagiQ [7]. However, this technology is still growing up that the researchers are trying to improve the performance and to avoid the problems that are mentioned in [9].

BB84 [2] and B92 [4] protocols are the examples of the two-party protocol and the examples of the three-party QKD are 3QKDPMA [6] and 3PAQKD-TB [1]. The Trust Center (TC) as the third party has responsibilities to construct session key and to authenticate the user on a communication session. Therefore, the more secure and convenient communication session are achieved. Furthermore, the TC should have the high-security level to assure all communications done under the secure condition. The 3QKDPMA [6] and 3PAQKD-TB [1] provide the effective method for creating the session key. However, those protocols do not provide the advanced user authentication. It is because TC only performs user's secret key table to authenticate the legal user that can cause security problems. Three-party authenticated quantum key distribution protocol such as 3QKDPMA [6] and 3PAQKD-TB [1] share a secret key between TC and user. It causes security problems because keeping the password table on the TC, security problems can occur and increase the overhead of verifying legal user [3, 8].

On the other point of view, to improve the security level, many studies about the user authentications without user table kept on the server have been published. Firstly, Hwang et al. [8] proposed a new scheme for authenticating user password without password table. This scheme has been evolving since 1990 and in 2013 Tang et al. [10] proposed the user authentication with smart card based on elliptic curve cryptography and it is improved by Wijayanto and Hwang in 2015 [3].

Based on the weakness of 3QKDPMA [6] and 3PAQKD-TB [1] we proposed a new scheme for the three-party quantum key distribution with strong user authentication based on the smart card. The contributions in this study are as follows. The first, TC does not need to maintain the user password for increasing the security. The second, user does not need to share his/her secret key with the TC. The third, it uses the smart card for user authentication and this scheme is resisted from the smart card lost.

The remaining sections of this paper are organized as follows. Section 2 gives brief reviews of quantum cryptography, 3QKD protocols, and user authentication scheme. In Section 3, we propose the new scheme for 3QKD with strong user authentication. The security analysis of our scheme is given in section 4. And finally, Section 5 concludes the work.

2 Related Works

As the motivation of this study, this section gives an overview of quantum key distribution BB84, the review of three-party quantum key distribution protocols, and the secure user authentication.

2.1 Quantum Key Distribution Protocol BB84

Classical cryptography is based on the mathematics such as DES, AES, RSA, El-Gamal, and Elliptic Curve whereas quantum cryptography is developed by the uncertainty of quantum measurement [2, 6]. It utilizes the polarized photons to transmit digital information and an attacker is impossible to eavesdrop communication channel without disturbing the transmission that will be detected easily [2]. In 1984, Bennet and Brassard proposed the BB84 protocol that uses two measurements basis. Those are rectilinear base and diagonal base [2]. Next, in 1992, Bennett published the other quantum key distribution protocol that is called by the B92 [4]. It performs the other property of the quantum theory; it is polarization-entangled or space-time-entangled two-photon states [4].

The main principle of quantum cryptography is inspired by Vernam cipher [5]. This cryptosystem was firstly invented by Gilbert Vernam for telegraph communication in 1917 and Shanon in 1949 proved that it is an unbreakable cryptosystem [5]. The basic idea of Vernam cipher is to encrypt the plaintext by a key that the length of this key is the same with the length of the message and this key is only used one time. This cryptosystem is also known by One Time Pad (OTP) [5]. However, the main problem in this cryptosystem is the key distribution and the quantum cryptography provides its solution. The quantum cryptography protocol communicates the OTP key by using the quantum channel. Therefore, it is invulnerable from eavesdrops the communication channel to steal the OTP key without disturbing and detected easily as is mentioned in the previous paragraph. For more clear explanation, this article presents the BB84 and B92 protocols in the following subsections.

2.2 The3QKDPMA Protocol

The Three-Party Quantum Key Distribution Protocol with Mutual Authentication (3QKDPMA) Protocol was invented by Hwang T. et al. in 2007 [6]. The user authentication in this scheme is based on the user secret key (K_{TU}) that is one secret key for one user and the secret key is shared in advance by each user to the Trust Center (TC). The 3QKDPMA consists of two phases that those are key setup phase and key distribution phase.

For example, Alice (U_A) and Bob (U_B) are two users that want to create a session key, where K_{TA} is the secret key of Alice, and K_{TB} is the secret key of Bob. In the setup phase, the TC measures the bases utilizing the bit sequence of the KTU. If $(K_{TA})_i$=0 then basis D is chosen else basis R is chosen. After the TC authenticates Alice and Bob, it generates two random number r_{TA}, r_{TB}, and the session key sk. Then the TC computes $R_{TA}=h(K_{TA},r_{TA})\oplus(sk\|U_A\|U_B)$ and $R_{TB}=h(K_{TB},r_{TB})\oplus(sk\|U_B\|U_A)$. Next, the TC constructs the qubits Q_{TA} based on the $(r_{TA}\|R_{TA})_i$ and $(K_{TA})_i$, where i = 1,2,3,...,n, and it follows this rule:

If $(r_{TA}\|R_{TA})_i$=0, and $(K_{TA})_i$=0 then quantum state is a polarization with 45 degrees.
If $(\mathrm{r}_{\mathrm{TA}}\|R_{TA})_i$=0, and $(K_{TA})_i$=1 then quantum state is a polarization with 135 degrees.
If $(r_{TA}\|R_{TA})_i$=1, and $(K_{TA})_i$=0 then quantum state is a polarization with 0 degrees.

If $(r_{TA}||\mathrm{R}_{\mathrm{TA}})_i$=1, and $(K_{TA})_i$=1 then quantum state is a polarization with 90 degrees. By the same way, TC also creates Q_{TB} and sends Q_{TA} to Alice and Q_{TB} to Bob.

On the other hand, Alice receives $(r'_{TA}||R'_{TA})$ and Bob also receives $(r'_{TB}||R'_{TB})$. Next, Alice computes *sk'* by $sk'||U_A'||U_B'=h(K_{TA},r'_{TA}) \oplus R'_{TA}$. And Bob also gets *sk'* by the same way.

The next step is the mutual authentication between Alice and Bob. In the Alice side, she generates a random number r_A and she computes $CS_A=h(sk',r_A)\oplus(U_A||U_B)$, then sends $r_A||CS_A$ to Bob. After receiving $r_A||CS_A$, Bob authenticates Alice by $U_A||U_B=h(sk',r_A)\oplus CS'_A$. On the other hand, Alice also authenticates Bob by the similar process.

2.3 The 3PAQKD-TB Protocol

Chen et al.[1] improved the 3QKDPMA protocol to reduce the computation and this proposed scheme is called by the 3PAQKD-TB or the three-party authenticated quantum key distribution with Time-Bound.

The detail of this protocol is presented in this section by the following example. Let Alice and Bob are two users that want to establish a session key. Firstly, Alice determines a time bound set {*tA1, tA2, tA3, ..., tAn*}, and sends it to *TC*. After receiving this message, *TC* informs Bob and then Bob creates time bound set {*tB1, tB2, tB3, ..., tBm*}and sends it to *TC*. Secondly, *TC* select the intersection of both time bound sets {*tI1,tI2,tI3,...,tIp*} and creates two secret keys *KTA* for Alice, and *KTB* for Bob. And then sends *KTA* and {*tI1,tI2,tI3,...,tIp*} to Alice and *KTB* and {*tI1,tI2,tI3,...,tIp*} to Bob. The next steps are similar to the 3QKDPMA protocol, but Alice and Bob only need the key setup phase once because they have many session keys that the number of session keys depends on the number of intersection time-bound elements. One session key is calculated by *SKx=h(ck',tbx)*.Where *ck'* is the certificate that is got from *TC* and it is same with *sk* (session key) in 3QKDPMA protocol that is explained in the previous subsection.*SKx* is the x^{th} session key of the x^{th} communication session, and *tbx* is the x^{th} element of the intersection time bound set.

2.4 The Secure User Authentication Protocol Based on the Smart Card

The user authentication protocol based on the smart card is divided into five parts [3], those are system setup phase, registration phase, login phase, authentication phase, and password change phase. Besides that, this scheme is based on the elliptic curve cryptography (ECC) that has the shorter keys than RSA or El-Gamal cryptosystem. The details of this scheme are presented as follows that is taken from [3].

2.4.1 System Setup Phase

In this phase, the server selects a secret key x and computes $Q=x.P$ and keeps secret key x. After that, the server publishes the ECC public keys parameters $p, a, b, P, n, h,$ and Q.

2.4.2 Registration Phase

Figure 1 below shows the registration phase. It is done by users once in the first time they log-in to the server that also uses secure communication line. It consists of three steps as follows:

Step 1: The user, U_i select an identity ID_i, password PW_i, and also a high-entropy random number N. Then, users encrypt N by password PW_i as a symmetric key cryptography $C_i = Enc(PW_i, N)$. Next, the user sends ID_i and C_i to the server through a secure channel.

Step 2: After receiving ID_i, and C_i, the server selects a random number k_i that $0<k_i<n$ and also a high-entropy random number M_i. Next, the server computes an EC digital signature by secret key x, and a hash function of a concatenation of ID_i and M_i as $sign(x,k_i,h(ID_i||M_i))$, for a short we call it sign. Then, the server computes $V_i=sign \oplus C_i$, stores V_i into smart card and sends it back to user U_i through a secure channel. Finally, the server maintains an ID table that contains ID_i, status-bit, k_i, and M_i.

Step 3: After receiving a smart card, the user inputs N into the smart card.

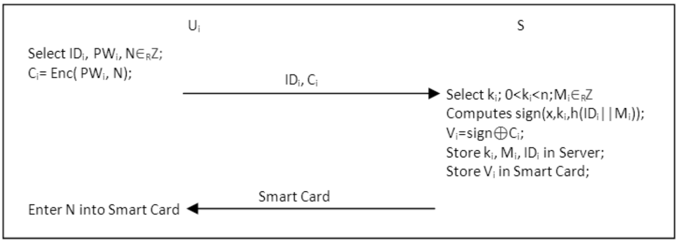

U_i
S
Select ID_i, PW_i, $N \in_R Z$;
C_i= Enc(PW_i, N);
ID_i, C_i
Select k_i; $0<k_i<n$;$M_i \in_R Z$
Computes sign(x,k_i,h(ID_i||M_i));
V_i=sign$\oplus C_i$;
Store k_i, M_i, ID_i in Server;
Store V_i in Smart Card;
Smart Card
Enter N into Smart Card

Fig. 1. Registration Phase

2.4.3 Login Phase

In the login phase, the interaction between users and server are utilized common channel. Firstly, the user inputs his or her identity ID_i and password PW_i into the smart card. Then smart card computes $s=V_i \oplus C_i$ that equals to sign because of $V_i=sign \oplus C_i$. Secondly, smart card chooses a random nonce $r_1 \in_R Z^*_n$, computes $R_1=r_1.P$, and $R_2=r_1.Q$. Thirdly, smart card encrypt $C_1=ENC(R2,ID_i||R_1||R_2||s||Tc)$ where Tc is the timestamp of U_1 and then sends R_1 and C_1 to servers. This phase is shown in Fig. 2 below.

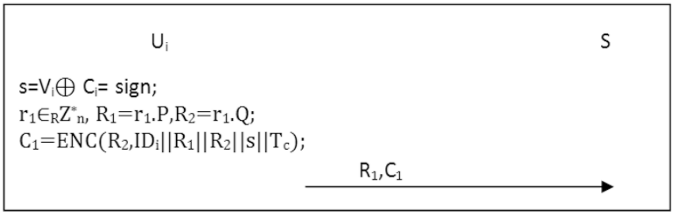

U_i
S
$s=V_i \oplus C_i=$ sign;
$r_1 \in_R Z^*_n$, $R_1=r_1.P, R_2=r_1.Q$;
$C_1=ENC(R_2, ID_i||R_1||R_2||s||T_c)$;
R_1, C_1

Fig. 2. Login Phase

2.4.4 Authentication Phase

When a log-in request that are R_1 and C_1 arrive at the Server *S*, *S* will do four steps that are described as bellow.

Step 1: Server *S* computes the session key R_2' by secret key *x* as $R_2'=x.R_1$. Then, Server decrypts C_1 by R_2', this result is $ID_i||R_1||R_2||s||Tc$. If this decryption is failed for producing those parameters, this login phase is rejected, and informs the sender.

Step 2: *S* checks the ID_i in the database. If this *ID* is not available in the database, *S* will reject this request and informs U_i in encrypted text by password R_2'.

Step 3: *S* checks status-bit. If status-bit is equal to one, the server rejects this request and informs U_i about it in encrypted text by password R_2', otherwise, the server sets it to be one.

Step 4: *S* checks *Tc*. If $(Ts\text{-}Tc)<=0$ or $(Ts\text{-}Tc)>\Delta T$ server rejects this request and informs U_i in encrypted text by password R_2'.

Step 5: Server computes its signature as $s'=sign(x,k_i,h(ID_i||M_i))$ and compares it with *s*. If those are not equal, *S* rejects this request and informs U_i about it. Otherwise, U_i has passed this authentication phase in the server side. And then, *S* encrypts $S||Ts$ by R_2' and sends back C_2 to user U_i in encrypted text by password R_2'.

The next step is done on the user side. U_i decrypts C_2 by R_2 and check *S* and *Tc* by the same way as the server did. If those parameters are not satisfied the requirement criteria, U_i will reject this session.

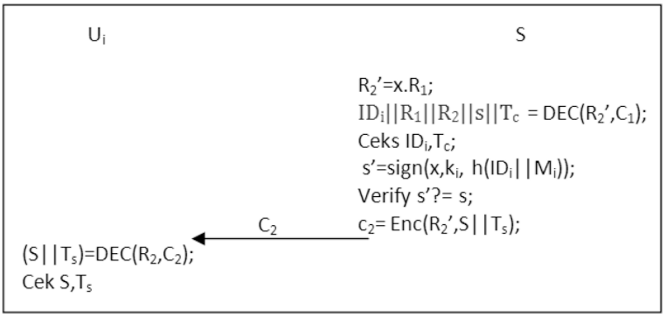

U_i
S
$R_2'=x.R_1$;
$ID_i||R_1||R_2||s||T_c = DEC(R_2', C_1)$;
Ceks ID_i, T_c;
$s'=sign(x,k_i, h(ID_i||M_i))$;
Verify $s'?= s$;
C_2
$c_2= Enc(R_2', S||T_s)$;
$(S||T_s)=DEC(R_2, C_2)$;
Cek S, T_s

Fig. 3. Authentication Phase

2.4.5 Password Change Phase

The password change phase is shown in Fig. 4 below. When the Ui wants to change his or her password for some reasons, U_i should keys his or her identity ID_i and password PW_i to smart card first before changing the password. After that, Smartcard will perform login protocol and if the login process is passed, U_i can input the new password $PW_{i,new}$. After that, smart card generates new random number N_{new} and computes $V_i=V_i \oplus Enc(PW_i,N) \oplus Enc(PW_{i,new},N_{new})$. Next, smartcard replaces V_i and N by $V_{i,new}$ and N_{new}. Finally, smartcard informs U_i that changing password is successful.

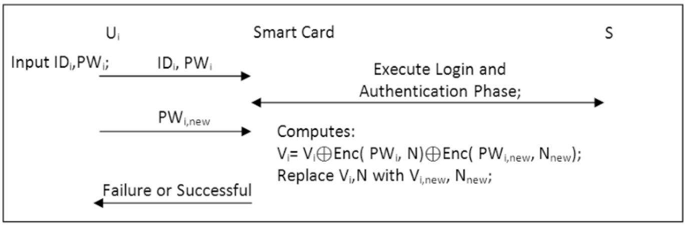


Fig. 4. Password Change Phase

3 Method

In this section is presented a new scheme for three-party quantum key distribution that adopted the Wijayanto's smart card user authentication scheme [3] and the 3PAQKD-TB [1]. Generally, it is divided into five phases that are system setup phase, registration phase, login phase, authentication and key generation phase, and password change phase. However, the first two phases and the password change phase are the same with Wijayanto's protocol but all communications are done by BB84 protocol [2]. Therefore, in this section, we only present the two remaining phases. Furthermore, it is similar to the Wijayanto's scheme that performs the classic public key cryptosystem ECC [5].

3.1 Login Phase

In the login phase, the interaction between a user U_1 and server S uses the quantum channel and *sign* or s as the measuring basis where $s_{Ui}=V_i \oplus C_i$ and the V_i is stored in the user's smart card and C_i is entered by the user. First, U_1 enters his or hers identity ID_1, the identity of user U_2 as a couple of the communication session, and password PW_1 into the U_1*'s* machine. Second, the U_1*'s* machine generates a random $r_1 \in_R Z^*_n$, calculates $R_{1-U1}=r_1.P,R_{2-U1}=r_1.Q$, and determines a time bound set $TB_{U1}=\{t_{U1-1},t_{U1-2},t_{U1-3},...,t_{U1-n}\}$. Third, the U_1's computer encrypt $c_1=ENC(R_2,ID_1||ID_2||R_{1-U1}||R_{2-U1}||s_{U1}||Tc||TB_{U1})$ where Tc is the U_1 time-stamp. And next, U_1 sends R_{1-U1} and c_1 to server S. This phase is presented in Fig. 2 below.

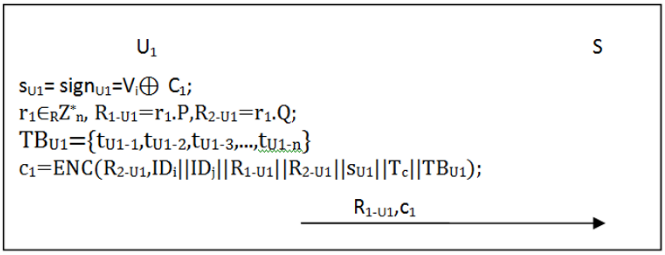

U_1
S
$s_{U1}= sign_{U1}=V_i\oplus C_1;$
$r_1\in_R Z^*_n, R_{1-U1}=r_1.P, R_{2-U1}=r_1.Q;$
$TB_{U1}=\{t_{U1-1},t_{U1-2},t_{U1-3},\ldots,t_{U1-n}\}$
$c_1=ENC(R_{2-U1},ID_i||ID_j||R_{1-U1}||R_{2-U1}||s_{U1}||T_c||TB_{U1});$
R_{1-U1},c_1

Fig. 5. Login Phase

3.2 Authentication and Key Generation Phase

After receiving R_{1-U1} and c_1, the server S calculate R'_{2-U1} and decrypt c_1 by $R'_{2-U1}=x.R_{1-U1}$ and $ID_i||ID_2||R_{1-U1}||R_{2-U1}||s||Tc||TB_{U1}=ENC(R'_2,c_1)$. The next step is the server verifies U_1 as follows by the same process in Wijayanto's scheme. If it passes the verification step, the server S signs the identity of U_2 by $s_{U2}=sign(x,k_2,h(ID_2||M_2))$. Then the server S sends a message to U_2 through the quantum channel and uses s_{U2} bits as the measurement bases. The U_2 verifies the server by the signature of S, it is the same within login phase of U_1 where $s_{U2}=V_2\oplus C_2$. Then U_2 does the same way with U_1 login phase with $TB_{U2}=\{t_{U2-1},t_{U2-2},t_{U2-3},\ldots,t_{U2-n}\}$.

The next step is done by the server by determining the intersection set TI of the two time-bound sets and sends it to the both U_1 and U_2. And the following steps are the same with 3PAQKD-TB protocol that U_1 and U_2 authenticate each other (mutual authentication) by their time bound set and keep the session key for the further communication session. Besides that, after keeping the session key sk, the next communication between U_1 and U_2 uses the session key bits as the measurement bases. Fig. 6 below presents the Authentication and the Key Generation Phase

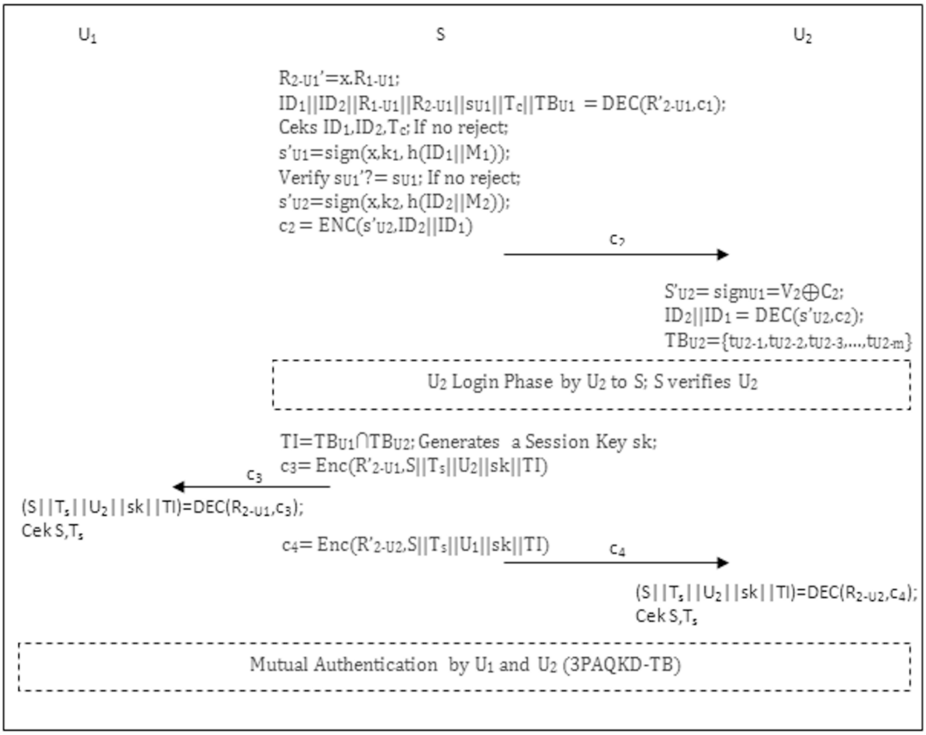

U_1
S
U_2
$R_{2\text{-}U1}'=x.R_{1\text{-}U1};$
$ID_1||ID_2||R_{1\text{-}U1}||R_{2\text{-}U1}||s_{U1}||T_c||TB_{U1} = DEC(R'_{2\text{-}U1},c_1);$
Ceks ID_1,ID_2,T_c; If no reject;
$s'_{U1}=sign(x,k_1, h(ID_1||M_1));$
Verify $s_{U1}'?= s_{U1}$; If no reject;
$s'_{U2}=sign(x,k_2, h(ID_2||M_2));$
$c_2 = ENC(s'_{U2},ID_2||ID_1)$
$c_?$
$S'_{U2}= sign_{U1}=V_2\oplus C_2;$
$ID_2||ID_1 = DEC(s'_{U2},c_2);$
$TB_{U2}=\{t_{U2\text{-}1},t_{U2\text{-}2},t_{U2\text{-}3},\ldots,t_{U2\text{-}m}\}$
U_2 Login Phase by U_2 to S; S verifies U_2
$TI=TB_{U1}\cap TB_{U2}$; Generates a Session Key sk;
$c_3= Enc(R'_{2\text{-}U1},S||T_s||U_2||sk||TI)$
c_3
$(S||T_s||U_2||sk||TI)=DEC(R_{2\text{-}U1},c_3);$
Cek S,T_s
$c_4= Enc(R'_{2\text{-}U2},S||T_s||U_1||sk||TI)$
c_4
$(S||T_s||U_2||sk||TI)=DEC(R_{2\text{-}U2},c_4);$
Cek S,T_s
Mutual Authentication by U_1 and U_2 (3PAQKD-TB)

Fig.6. Verification and Key Generation Phase

4 Security Analysis

First of all, this scheme is derived from the 3PAQSK-TB [1] scheme that resists of eavesdropping. As mentioned in [2], in the quantum key distribution protocol, an attacker cannot eavesdrop the channel without disturbing the communication that can be detected easily. It is because the attacker cannot measure the basis accurately and retransmit again the correct qubits.

The other benefit of 3PAQSK-TB [1] is that has strong mutual authentication among two users in the communication session. However, the authentication of the users and TC is only done by static pre-shared secret key.It reduces the security principle of QKD that based on one-time pad encryption (OTP) [5]. The OTP utilizes a long secret key that has the same length with the plaintext length and this secret key is only used one time. In our scheme, it uses R2 to request the session key to the TC that is based on the secure random number of an integer. On the other words, a user uses the different temporal secret key or it is called by the session key. Therefore, it increases the security level of the 3PAQSK-TB protocol.

Besides that, the scheme proposed in this article does not require the TC to save the user's secret key on a table that is not immune to the insider privileged attacker stolen-verifier attack [10]. This security is only based on the secret key of ECC signature (x) and the ECC is known as the strong public key cryptosystem [5]. An attacker also cannot guess the random N since it is a high-entropy random number and the user's password is chosen properly by the user that should be strong enough from off-line guessing attacks such as brute force and dictionary attack.

The *TC* should be secure enough from the DoS attack that tries to make *TC* overload and shuts down the service. This proposed scheme is invulnerable from the DoS because it performs status bit and each user that want to generate a session key should be authenticated first. The status bit is set by "1" for a user while the user request of a session key. Therefore, one user cannot request more than one at one time. Besides that, every user uses unique signature bits to connect to the *TC* and only the *TC* can make it so that an attacker cannot perform dummy users to do the DoS attack.

The man in the middle attack or the impersonation attack can not be performed in this scheme because an attacker cannot pretend the legal user. To impersonate the legal user, an attacker should have the legal user password and the signature of the *TC*. Assume that an attacker is successful to occupy a legal user's computer. Then he or he got the N but he or she cannot get the *TC* signature because he or she needs the legal user password to unwrap the *TC*'s signature from V_i.

This proposed scheme is also provided by time-stamp *Tc* in every encrypted message. It is very important to avoid the reply attack. Assume that an attacker has stolen the reply message from the *TC* after one legal user passes the verification phase. Then he or she uses this reply message to request a new session key. The problem for doing this is because the *Tc* of the stolen reply message is out of date.

5 Conclusions

The main propose of the scheme proposed in this article is to provide the 3PAQKD-TB [1] with the TC's user authentication. This scheme still maintains the advantages of the previous scheme that those are mutual user authentication and eavesdropping resistance. Moreover, this scheme also immunes of the insider privileged attack or the stolen-verifier attack, the DoS attack, the man in the middle attack or the impersonation attack and the reply attack.

References

1. Hsing-Chung Chen, Syuan-Zong Lin, and Tzu-Liang Kung, "Three-Party Authenticated Quantum Key Distribution Protocol with Time Constraint,"Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, 2012.
2. Charles H.Bennett, Gilles Brassard, "Quantum Cryptography: Public Key Distribution and Coin Tossing," Theoretical Computer Science, Vol. 560, Part 1, pp. 7–11, 2014.
3. Heri Wijayanto, Min-Shiang Hwang, "Improvement on Timestamp-based User Authentication Scheme with Smart Card Lost Attack Resistance," International Journal of Network Security, Vol 17, No.2, pp. 160-164, 2015.
4. Charles. H. Bennett, "Quantum Cryptography Using Any Two Nonorthogonal States," Physical Review Letters, Vol. 68, No. 2, pp.3121-3124, 1992.

5. Alfred J. Menezes, Paul C. van Oorschot, Scott A. Vanstone, “Handbook of Applied Cryptography,” CRC Press, Boca Raton, London, New York, Washington D.C., 1996.
6. Tzonelih Hwang, Kuo-Chang Lee, and Chuan-Ming Li, “Provably Secure Three-Party Authenticated Quantum Key Distribution Protocols,” IEEE Transactions on Dependable and Secure Computing, Vol. 4, No. 1, January-March, 2007.
7. MAGIQ QPN 8505 Uncompromising VPN Security TM, http://www.magiqtech.com/Products_files/8505_Data_Sheet.pdf, accessed in Aug. 24th, 2016.J
8. Tzonelih Hwang, Yihwa Chen, Chi-Sung Laih, “Non-Interactive Password Authentications without Password Tables,” In Proceedings of IEEE Region 10-th Conference on Computer and Communication Systems, Hong Kong, Sep., 1990.
9. Valerio Scarani, Christian Kurtsiefer, “The black paper of quantum cryptography: Real implementation problems,” Theoretical Computer Science, Vol. 560, pp. 27–32.
10. Tang Hong-Bin, Liu Xin-Song, Jiang Lei, “A Robust and Efficient Timestamp-based Remote User Authentication Scheme with Smart Card Lost Attack Resistance”, International Journal of Network Security, Vol. 15, No.6, PP.446-454, Nov., 2013.

TCP with network coding meets loss burstiness estimation for lossy networks

Nguyen Viet Ha, Kazumi Kumazoe and Masato Tsuru

Abstract Although Transmission Control Protocol (TCP) is still dominant in both wired and wireless networks, TCP in lossy networks suffers from goodput reduction because it considers any packet loss as a network congestion signal and decreases its sending rate even if the loss is not due to congestion. TCP with network coding (TCP/NC) had been proposed to overcome this problem. It is expected to recover the lost packets without retransmission at the sink by proactively sending the redundant combination packets encoded at the source. However, the original TCP/NC is ineffective in burst loss channels because it does not provide any means to determine appropriate values of the network coding parameters, e.g., redundancy factor and coding window size, to adapt burst loss channels. We propose the new scheme called TCP/NC with loss rate and loss burstiness estimation (TCP/NCwLRLBE), which periodically estimates not only the loss rate but also the loss burstiness based on observation of the past communications and adaptively determines appropriate network coding parameters based on computation of the necessary recovery capability. We implemented and validated our proposal in Network Simulator 3. The results show that the TCP/NCwLRLBE can improve the TCP goodput in the burst loss channels.

1 Introduction

The conventional transmission control protocol (TCP) recognizes all packet losses to be a sign of network congestion and cuts down the sending rate, even if they are

Nguyen Viet Ha
Kyushu Institute of Technology, Japan, e-mail: nvha@infonet.cse.kyutech.ac.jp

Kazumi Kumazoe
Kyushu Institute of Technology, Japan, e-mail: kuma@ndrc.kyutech.ac.jp

Masato Tsuru
Kyushu Institute of Technology, Japan, e-mail: tsuru@cse.kyutech.ac.jp

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_28

caused by a lossy network. Reducing mistakenly the sending rate in the cases of non-congestion packet loss makes the goodput performance of TCP considerably degrade. TCP with network coding (TCP/NC) was presented [1, 2] to address this problem. The term network coding (NC), while is being used in a broader sense, is mostly referred as a technique in which multiple original packets are combined into multiple coded packets to traverse a network. Those packets are decoded to the original packets after traversing the network, in order to improve the throughput, delay, and/or resilience. In TCP/NC, the source sends the data as the random linear network coding combination packets (referred to as combination packets or combinations) to the sink across a lossy network. The sink is expected to recover all data using the remaining combinations without retransmission if some of combinations are lost. It avoids the unnecessary reduction of the sending rate due to TCP congestion windows (CWND) control. A new network coding layer is added into the protocol stack between TCP and IP layers to provide the recovery ability shown in Fig.1. This layer operates transparently with upper and lower layer; thus, it can take the functionality of the original TCP protocol such as congestion control and retransmission mechanism; and the benefit of the network coding in recovering the lost packets quickly.

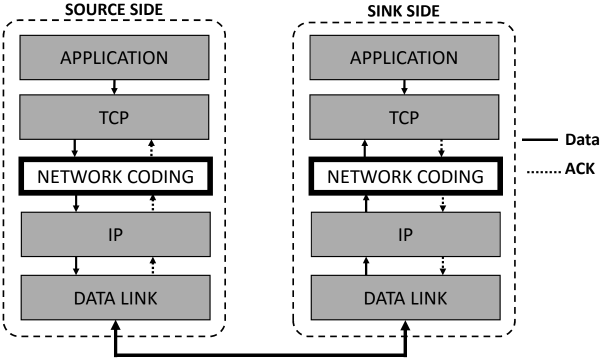


Fig. 1 The network coding layer in the TCP/IP model

Since the original TCP/NC cannot change the network coding-related parameters (NC parameters), it cannot work well in the time-varying packet loss conditions. Some studies improved the original TCP/NC by adding the ability of adjusting automatically the NC parameters to adapt to a change of channel such as Self-adaptive network coding with TCP (SANC-TCP) [3] and Adaptive network coding with TCP (ANC-TCP) [4]. In these studies, they retained the basic algorithm to calculate the NC parameters which is proposed in [1]. However, the NC parameters that are calculated based on the basic algorithm cannot achieve the best performance ([2]) because this algorithm only limits the minimum value of the NC parameters to recover all packet losses in the ideal condition of channel. The value of NC parameters can be chosen higher than that of the calculated value but it involves the increase of the cost link resource which is limited. In our previous study, we proposed a new scheme called TCP/NC with loss rate estimation (TCP/NCwLRE) [5] to choose the NC parameters based on the success probability of transmission. The transfer performance is significantly increased when comparing with the basic algorithm. However, all the previously methods only considered the random losses in which packet losses happen independently and separately. The fact is that the burstiness loss condition is more severely affected to the data transfer performance than the random loss condition even though the average link loss rate of the both conditions is the same [6]. In general, burst packet losses hinder the potential of TCP/NC due to the following problems. First, there is no optimization of the NC parameters to adapt the burst loss channels. Secondly, the "size of the burst loss" sometimes increases suddenly to a value higher than the average value, and exceeds the recovery capacity of the network coding system. The performance of system is decreased in a short time after retransmission because the standard retransmission of TCP/NC is ineffective in the burstiness loss condition.

In this paper, to cope with the above mentioned first problem, we consider the adaptation of the NC parameters in the burst loss condition in addition to the random loss condition. We already proposed TCP/NCwLRE to estimate and adapt to the link loss rate on the random loss channels [5]. Based on this, we propose TCP/NCwLRLBE to estimate and adapt to the link loss rate and the loss burstiness, i.e., the "average size of the burst loss" on the burst loss channels.

Note that, to cope with the above mentioned second problem, we also combine an efficient retransmission mechanism of unrecoverable lost packets (TCP/NCwER) in TCP/NC, which was proposed in our previous work [6]. TCP/NCwER helps the NC layer retransmit the lost packets in an efficient way and all the retransmitted packets are also encoded (combined). TCP/NCwER with its functions can determine the number of packet losses as well as the exactly packets needed to retransmit; thus, it can retransmit these multiple packets in one round trip time.

The remainder of this paper is organized as follows. In Section 2, we describe the fundamental concept of TCP/NC. The proposed protocol is presented in Section 3. Simulations and results are described in Section 4 and the conclusions is discussed in Section 5.

2 Network coding fundamental

TCP/NC protocol was presented in 2008 [2] which successfully implemented the network coding into protocol stack with a minor change by adding a network coding layer between TCP and IP layer, as shown in Fig.1. TCP/NC allows the source to send m combination packets (C) created from n original packets (p) with $m \geq n$ using Eq. (1) where α is the coefficient on a certain Galois Field. If the number of lost combinations is less than $k = m - n$, the sink is expected to recover all the original packets using the remaining combinations without retransmission. Therefore, TCP layer is unaware of light loss events occurring and maintains the CWND appropriately to improve the goodput performance. The processes of creating m

combinations and regenerating n original packets are called encoding and decoding, respectively. In theory, the encoding/decoding process handles in each coding window (CW) which is a group of n original packets; thus, n is actually the coding window size in packet.

$$C[i] = \sum_{j=1}^{n} \alpha_{ij} p_j ; \quad i = 1,2,3,...,m \tag{1}$$

There are two NC parameters to decide the performance of the transfer system. Redundancy factor R is the number of linear combination packets sent to IP layer on average; hence R equals the quotient of m and n. And the other is the mentioned parameter k, the recovery capacity in one CW (the CW recovery capacity), which is understood as the maximum number of packet losses in each CW can be recovered without retransmission. R can be chosen based on the link loss rate. If the link loss rate is high, an appropriate R can improve the goodput performance but a large R incurs the unnecessary redundancy and reduces the goodput. And k is chosen based on the types of channels (e.g., random or burtiness loss condition), the time taken by the sink to wait for decoding (decoding delay) and the hardware limitations (e.g., processor, memory).

Besides executing the encoding/decoding process, network coding layer allows a new interpretation of ACKs by using the degree of freedom concept and the seen/unseen definition [1]. The ACK number in ACK packet is be changed to the sequence number of the oldest "unseen" packet, which will be decoded when the sink receives the additional combinations.

Definition 1. (seeing a packet). A node is said to have seen a packet p if it has enough information to compute a linear combination of the form $(p+q)$, where q is a linear combination itself involving only packets that arrived after p at the sender.

Fig. 2 is an example of the encoding, decoding and ACK packet returning processes. The packets p_1, p_2, p_3 and p_4 are encoded to the combination $C[1]$, $C[2]$, $C[3]$, $C[4]$, $C[5]$ and $C[6]$. When a new packet comes to NC layer, the combinations will be created and transported immediately. Due to the two lost combinations, the NC layer cannot decode any combinations until receiving the combination $C[6]$. For each received combinations, NC layer returns an ACK packet whose ACK number corresponds to the smallest unseen packet. During the process, the TCP layer totally unawares with any loss events; thus, the TCP congestion window (CWND) keeps increasing and the performance is stable.

TCP/NC only uses the original retransmission and congestion control mechanisms of TCP layer. In other words, NC layer is completely transparent in these mechanisms. If the number of lost combinations exceeds the recovery capability, one or some packets will be "unseen" in all received combinations. Then TCP layer will receive duplicate ACK numbers from NC layer and retransmit the "unseen" packets to NC layer; NC layer simply forwards them to the lower layer.

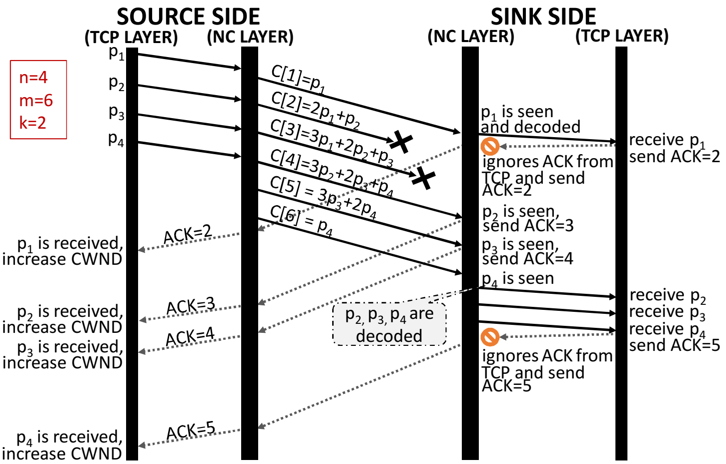

SOURCE SIDE
(TCP LAYER)
(NC LAYER)
SINK SIDE
(NC LAYER)
(TCP LAYER)
n=4
m=6
k=2
p_1
p_2
p_3
p_4
C[1]=p_1
C[2]=$2p_1+p_2$
C[3]=$3p_1+2p_2+p_3$
C[4]=$3p_2+2p_3+p_4$
C[5] = $3p_3+2p_4$
C[6] = p_4
p_1 is seen and decoded
ignores ACK from TCP and send ACK=2
p_2 is seen, send ACK=3
p_3 is seen, send ACK=4
p_4 is seen
receive p_1 send ACK=2
receive p_2
receive p_3
receive p_4
send ACK=5
ignores ACK from TCP and send ACK=5
p_2, p_3, p_4 are decoded
ACK=2
ACK=3
ACK=4
ACK=5
p_1 is received, increase CWND
p_2 is received, increase CWND
p_3 is received, increase CWND
p_4 is received, increase CWND

Fig. 2 Network coding process

3 The proposed method

In this section, we describe the idea of TCP/NC with loss rate and loss burstiness estimation (TCP/NCwLRLBE) which can dynamically adjust R to work well in the random and burst loss channel. And we present the new method for R estimation to improve the goodput performance of the TCP/NC system.

3.1 Estimating the link loss rate and the burstiness loss rate

To adjust R, the system has to estimate the link loss rate r. In other words, NC layer must determine the number of packet losses. The basic work was done in [6]. We proposed to use some additional information called NC-ACK header besides the normal TCP header in ACK packet. We added the new Packet-id (*Pid*) field to NC header as well as the new Packet-id echo-reply field (*Pid-reply*) and Redundancy flag (*R-flag*) in the NC-ACK header, as shown in Fig.3. Hence, each combination has a unique *Pid* number which can be understood as the identification number. After receiving the combination, the sink will return an ACK packet which has the *Pid-reply* number equal to the *Pid* number in the received combination. Based on the *Pid-reply* in the ACK packet, the source can determine the number of losses to estimate r by dividing the number of packet losses to the number of sending combinations.

The redundant combinations which should be ignored by original TCP/NC have to be counted in this situation. The sink has to return an ACK packet for all combinations including the redundant combinations. To avoid the mistaken forwarding, the sink must inform the source of which ACK packet is used only for counting and does not forward to TCP layer. To aid this goal, the sink uses *R-flag* in the returned

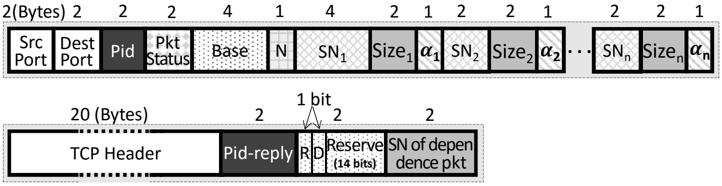


Fig. 3 The network coding header (above) and the network coding ACK header

Table 1 The NC header fields

Field name	**Description**
SrcPort	The source port number
DestPort	The destination port number
Pid	The packet identity
Pkt status	The packet status. Using for the returning ACK process
Base	The sequence number (SN) of the oldest packet in the NC buffer of the source. Using for buffer management at the sink
N	The number of original packet in the combination
SN_1	The SN of the first original packet
SN_n	Equal to the SN of the *n-th* packet subtract to SN_1
$Size_n$	The payload size of n-*th* packet
α_n	The n-*th* NC coefficient
Pid-reply	The packet identity echo reply
R	The redundancy flag
D	The dependence flag
Reserve	Reserved for the future use
SN of the dependence pkt	The SN of the dependence packet at the sink. Using to notify the source to retransmit this packet

ACK packet to indicate two types of ACK packets which are used to return for the normal and the redundancy combinations. If the source receives the ACK packet of the normal combination, it will forward to TCP layer. Otherwise, the source will drop the ACK packet after getting the control information.

Note that, we do not consider all fields of the header in this paper. The remaining fields which are useful for the effective retransmission mechanism of TCP/NCwER are minutely described in [6]. However, the short explanations of all fields in the NC header and NC-ACK header are shown in Table 1.

Other value considered to estimate is the average length of continuous losses (L). The *Pid* and *Pid-reply* fields are still used for this purpose. Based on the received *Pid-reply*, the source can know exactly which packets are lost; thus, it can determine of the length of continuous losses in each loss event (prefer as the burst loss size). L is calculated periodically by using Eq. (2), where l is the maximum of the burst loss size, N_j is the number of loss events of the burst size at j and N is the sum of N_j or the total loss events.

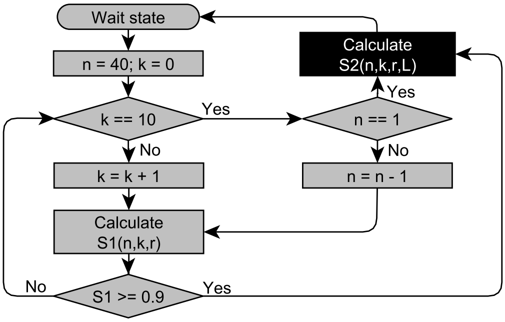


Fig. 4 Estimating n and k based on $S_1(n,k,r)$

$$L = \sum_{j=0}^{l} \frac{N_j \times j}{N} \tag{2}$$

3.2 Estimating the network coding parameters

After estimating r and L, two processes will be performed sequentially to estimate the good n and m. These processes are used to determine the minimum of R to help the system work well in the channel that has the link loss rate r. And it also can work well in the burstiness loss condition that has the average length L of continuous losses. At first, the success probability of transmission in one CW with the link loss rate r, $S_1(n,k,r)$, can be calculated by Eq. (3) where n and k are variables. The limiting of $S_1(n,k,r)$ depends on the specific conditions. In the burstiness loss condition, we find that $S(n,k) \geq 0.9$ can achieve the best goodput performance. The process of choosing n and k based on $S_1(n,k,r)$ is shown in Fig. 4. In our simulation, n is limited up to 40 and k is limited up to 10.

$$S_1(n,k,r) = \sum_{j=0}^{k} \binom{n+k}{j} r^j (1-r)^{n+k-j} \tag{3}$$

The above estimation of n and k based on $S_1(n,k,r)$, the probability that the number of lost packets in $(n+k)$ sending packets is no more than k in case that the packet loss happens independently (i.e., random loss) with the loss rate of r, can be seen as baseline values of them. They are expected to result in a good performance in case of the random loss condition, but may not sufficient in case of the burst loss condition. Therefore, we propose the second process to find better n and k based on $S_2(n,k,r,L)$, the probability that the number of lost packets in $(n+k)$ sending packets is no more than k in case that the packet loss happens dependently (i.e., burst loss) with the loss rate of r and the average length L of the continuous packet losses. Note that the exact form of S_2 depends on the burst loss channel model.

In this paper, we focus on a burst loss channel model which is available in the simulation software (Network simulator 3 - NS3 [7]), although our approach can be applied to other burst loss channel models. In NS3 burst loss model, a loss event happens with probability of e, and for each loss event, j-continuous packets are lost with equal conditional probability of $\frac{1}{l}$ where $j{=}1,2,\ldots,l$. Note that "the loss event rate" e can be expressed by $\frac{r}{L}$ and "the maximum number" l of continuously lost packets in one loss event can be expressed by $2L{-}1$, where r is the loss rate and L is the average length of continuous losses.

To investigate the statistical property of the number of lost packets in one CW, we consider the states (F,S) and their state transition probabilities, where F and S is the number of failure sending (i.e., lost packets) and successful sending in one CW, respectively. Therefore, in one CW, the number of the states is equal to $\frac{(m+1)(m+2)}{2}$; where $m{=}n{+}k$. The probabilities of state transition from state (f,s) to state (f,s), to state $(f,s{+}x)$, and to state $(f{+}x,s)$ are defined in Eqs. (4), (5), and (6), as follows. Let $e{=}\frac{r}{L}$ and $l{=}2L{-}1$, for $x = 1,2,\ldots$,

$$\Pr[(f,s) \rightarrow (f,s)] = \begin{cases} 1 & \text{if}\, f{+}s = m \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

$$\Pr[(f,s) \rightarrow (f,s+x)] = \begin{cases} 1-e & \text{if}\, x = 1,\, f{+}s{+}1 \le m \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

and

$$\Pr[(f,s) \rightarrow (f{+}x,s)] = \begin{cases} \dfrac{e}{l} & \text{if}\, 1 \le x \le l,\, f{+}s{+}l \le m \\ \dfrac{e}{d} & \text{if}\, 1 \le x \le d,\, f{+}s < m < f{+}s{+}l, \\ & \text{where}\, d = m{-}(f{+}s) \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

Let Q be the transition probability matrix with the dimension of $\frac{(m+1)(m+2)}{2}$, which is built using Eqs. (4), (5), and (6). The simple example of evolution of state transition is shown in Fig. 5.

Let $V_{ini}[(f,s)]$ be the existence probability of state (f,s) at the beginning of CW. Since each CW starts with $(f,s){=}(0,0)$ definitely, $V_{ini}[(0,0)]{=}1$ and $V_{ini}[(f,s)]{=}0$ for any $(f,s){\neq}(0,0)$, that is, vector $V_{ini}{=}(1,0,0,\ldots,0)$. At the end of CW, all $m{=}n{+}k$ packets have been sent and some of them have been lost; the all possible final states are $\{(f,s)|f{+}s{=}m\}$. Let $V_{fin}[(f,s)]$ be the existence probability of state (f,s) at the end of CW. The final state existence probability vector V_{fin} can be computed by:

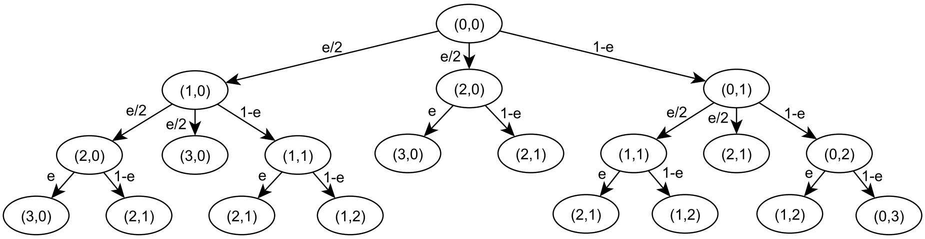


Fig. 5 The moving of the transition state with m=3 and l=2

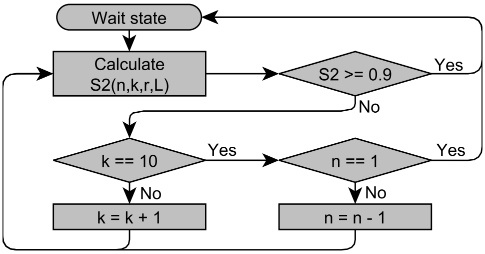


Fig. 6 Estimating n and k based on $S_2(n,k,r,L)$

$$V_{fin} = V_{ini} \times Q^m \qquad \text{where } V_{ini} = (1,0,0,\ldots,0) \tag{7}$$

Hence the probability that the number of lost packets in $(n+k)$ sending packets is no more than k is:

$$S_2(n,k,r,L) = \sum_{i=0}^{k} V_{fin}[(n+k-i,i)]. \tag{8}$$

The determination process of n and k based on $S_2(n,k,r,L)$ is shown in Fig. 6.

3.3 Adjusting the network coding parameter

The estimated $R = \frac{n+k}{n}$ will be updated periodically in each predetermined time period. To have enough information for estimating, besides the interval time, the number of sending combinations is used. On other words, the process of estimation R will be performed if two following conditions satisfy. The interval time and the number of ACKed packets and lost packets have to be greater than or equal the predefined value. Moreover, to increase the accuracy of the estimation, the moving average method can be used such as the simple moving average (SMA) in this paper.

In other studies [3, 4], they change the NC parameter at the starting of a new network coding process or after sending m combinations. However, it may take long time to wait due to the congestion control of TCP layer (decreasing the sending rate).

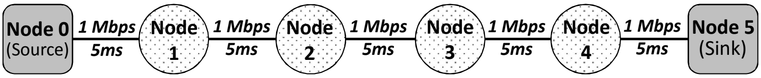


Fig. 7 Simulation topology

Consequently, the estimated parameters may be outdated. The method to adjust the network coding at any time without any affections to the current coding process should be considered. In our previous study in [5], we proposed a mechanism which can adjust n and k as soon as possible without affecting the recovery ability of the current *CW*. It can determine the state to safely finish the current encoding process or to keep using the current process and just change a new coefficient matrix. In here, "safely finish the current encoding process" means the encoding process will send k instead only one combination. After that, the encoding process can change to a new process with a new coefficient matrix. This proposed method is used in this paper as the method to adjust the NC parameter.

4 Simulation result

The implementation of TCP/NCwLRLBE was accomplished using Network Simulator 3. The topologies of the simulation consist of four tandemly arranged routers. There are the sources and the sinks on either side of these routers. To evaluate the basic performance in lossy networks, one source and one sink are used, as shown in Fig.7. All links have a bandwidth of 1 Mbps and a propagation delay of 5ms. The TCP type used in the simulation is NewReno, and the payload size is 1000 bytes. The transferred data size is 100 Mbytes. In all simulations, we run the system in totally 10 times to obtain the average value.

There are two types of the channel used for evaluating. The first is the random loss channel that simulates a slow fading channel. This channel causes the losses independently and separately. Another one is the burst loss channel that simulates a fast fading channel e.g., mobile wireless channels. All those packet losses happen on the links connected between the sources and the router.

In the calculation of $S_1(n,k,r)$ and $S_2(n,k,r,L)$, n and k are limited up to 40 and 10, respectively. The n and k are chosen based on the smallest $S_1(n,k,r)$ and $S_2(n,k,r,L)$ but they have to be greater than or equal to 0.9. Both TCP/NCwLRE and TCP/NCwLRLBE estimate the loss rate and compute a new R every 5 seconds. The SMA is used to avoid a sudden change of R by getting the mean of the 3 nearest estimated values of the loss rate. Noted that, we add TCP/NCwER, which has the efficient retransmission mechanism, to TCP/NCwLRE as the network coding scheme for the fairness comparison in this paper. Fig. 8 and Fig. 9 show the goodput performance of TCP NewReno, TCP/NCwLRE and TCP/ NCwLRLBE.

In the random channel (noted with l=1). TCP/NCwLRLBE has the best performance due to its efficient in adjusting n and k. As mentioned, TCP/NCwLRE is only

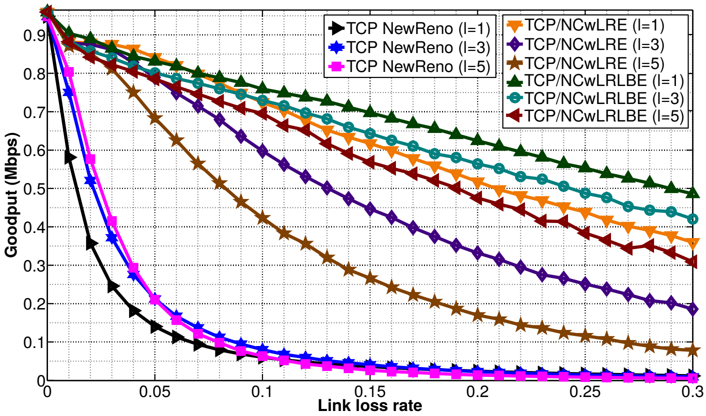


Fig. 8 The goodput comparison in the random loss channel (l=1) and the burst loss channel with l=3,5

designed for the random loss channel; thus, it keeps k as a constant (equal to 3) and just adjusts n. Although the channel in this simulation is randomly changed with the mostly separated loss, the burstiness still happens at a large link loss rate. Therefore, the TCP/NCwLRLBE, that estimates and adapts with the burstiness condition, is better than TCP/NCwLRE from the link loss rate equal to 0.07.

The results of the burst loss channel are noted with $l>1$. As mentioned before, in NS3 burst loss model and under a given link loss rate, the number of the continuous packet losses in one loss event is randomly chosen from 1 to a changeable number l. In this simulation, l is set at 3, 5, 7 and 9. In this second simulation case, the TCP/NCwLRLBE totally get the best performance compare with the other protocols. Besides, we can see the affection of the burst loss size l to the goodput performance. When the l is increased, the affection of the burstiness is increased and hinders the estimation process; thus, the goodput performance is increased following the increase of l. However, TCP/NCwLRLBE can keep the goodput be stable to sending the data.

5 Conclusions

In this paper, we have proposed a new scheme of network coding called TCP/NC with loss rate and loss burstiness estimation (TCP/NCwLRLBE) that can enable TCP/NC to automatically adjust the NC-layer behaviors to adapt to time-varying loss rates. We have implemented the mechanism to estimate the link loss rate and the loss burstiness from the continuous observation of the packet transmission between the source and the sink. We have also proposed new algorithm to calculate the redundancy factor (R) based on the probability distribution of the number of lost packets and the loss burstiness in one CW. This method is more appropriate than

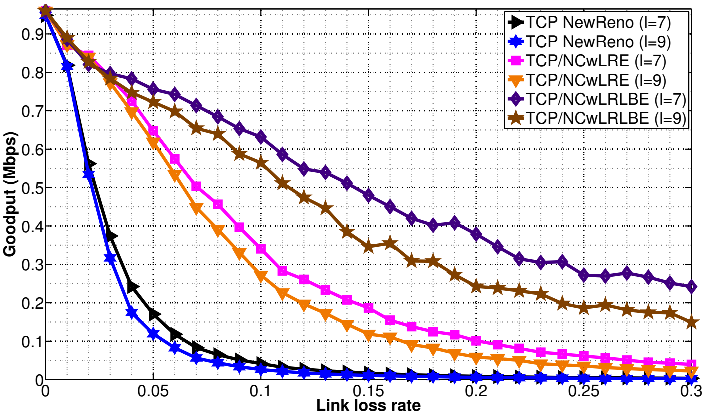


Fig. 9 The goodput comparison in the burst loss channel l=7, 9

the basic method that bases on the expectation (average) of the number of lost packets. The simulation results on Network simulator 3 (NS-3) show that the proposed TCP/NCwLRLBE can achieve the best goodput performance compared with other protocols e.g., TCP NewReno and TCP/NCwLRE.

Acknowledgements This work is partly supported by JSPS KAKENHI (16K00130) and KDDI Foundation.

References

1. J. K. Sundararajan, D. Shah and M. Medard (2008) ARQ for network coding. International symposium on information theory, 1651–1655
2. J. K. Sundararajan, D. Shah, M. Medard, S. Jakubczak, M. Mitzenmacher and J. Barros (2011) Network coding meets TCP: theory and implementation. In: Proc. of the IEEE, doi: 10.1109/JPROC.2010.2093850.
3. S. Song, H. Li, K. Pan, J. Liu and S Y R Li (2011) Self-adaptive TCP Protocol Combined with Network Coding Scheme. International Conference on Systems and Networks Communications, 20–25. Barcelona, Spain
4. C. Y. Cheng and H. Y. Yi (2013) Adaptive Network Coding Scheme for TCP over Wireless Sensor Networks. International Journal of Computers, Communications and Control, doi: http://dx.doi.org/10.15837/ijccc.2013.6.26
5. N.V. Ha, M. Tsuru, K. Kumazoe (2016) Making TCP/NC adjustable to time varying loss rates. 8-th International Conference on Intelligent Networking and Collaborative Systems (INCoS), 6 pages. Ostrava, Czech Republic
6. N.V. Ha, M. Tsuru, K. Kumazoe (2017) TCP Network Coding with Enhanced Retransmission for heavy and bursty loss. Journal of IEICE Transactions on Communications, doi: 10.1587/transcom.2016EBP3101
7. Network simulator (ns-3). https://www.nsnam.org/. Accessed 30 March 2016

Reduction of Network Traffic by Using the Peer Cache Mechanism in Co-located Collaborative Web Search on Smartphones

Tsuyoshi Donen, Shingo Otsubo, Ryo Nishide, and Hideyuki Takada

Abstract As mobile terminals become popular, more people are getting involved with collaboratively searching Web sites in order to achieve a common purpose with others. When performing such a task, we want to save the usage of the Internet traffic as much as possible because the transfer speed of mobile terminals may be limited when the data transfer usage exceeds a certain amount of gigabytes. To reduce the Internet traffic, we build a proxy system with the peer cache mechanism to share the Web contents stored on participating mobile terminals, focusing on the existence of the Web contents which are accessed multiple times from different terminals. Our experimental results reveal that about 20% of the Internet traffic has been reduced when four people engaged in a collaborative Web search task to find good restaurants for a year-end party.

1 Introduction

As Web search on mobile and tablet terminals becomes a part of our daily activity, collaborative Web search is often performed with multiple users who want to achieve a common goal[1]. For example, a group of people may search for a restaurant and sightseeing spots, or a route to destination.

On the other hand, the Internet traffic is increasing on account of the increasing usage of mobile terminals. It is reported that 55% of the Internet traffic of mobile communication is occupied by video streaming and Web browsing[2]. In order to avoid high network load, the telecommunication carriers have deployed bandwidth

Tsuyoshi Donen · Shingo Otsubo
Graduate School of Information Science and Engineering, Ritsumeikan University
e-mail: {t_donen, s_otsubo}@cm.is.ritsumei.ac.jp

Ryo Nishide · Hideyuki Takada
College of Information Science and Engineering, Ritsumeikan University
e-mail: r_nishide@cm.is.ritsumei.ac.jp, htakada@cs.ritsumei.ac.jp

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_29

control for mobile terminals. A mobile terminal with bandwidth control is limited with the speed of mobile data transmission. If the Internet traffic of mobile terminals can be reduced, the terminals are less likely to receive the bandwidth control.

In this research, we focus on the existence of Web contents accessed multiple times from different terminals during collaborative Web search, and propose a method to reduce the network traffic. In detail, Web contents are stored in the local cache space in mobile terminals, and the cache is shared among terminals using the Bluetooth ad hoc network. We expect that the method has a possibility to reduce the network traffic for mobile terminals. The evaluation experiments are conducted to examine how much the Internet traffic can be reduced in a real situation of collaborative Web search.

2 Related Work

Many kinds of cache mechanisms for Web browsers are available to accelerate the response time and reduce bandwidth. The Web browser caches the Web contents once they are accessed. When the same Web contents are accessed again, they are retrieved from the Web cache.

To leverage the use of cache mechanism, the Web cache can be shared with multiple terminals. Proxy servers are used to share Web cache. Squid[3] is one of the general cache mechanisms which work as a proxy server for caching Web contents accessed from the LAN environment. In contrast to Squid which works on the server/client system, a method on reducing the Internet traffic on P2P network is also proposed[4]. This method enables multiple terminals to share their local Web cache on the P2P network, applicable to small scale LAN environment.

One of the problems in these systems is that the client terminal must be connected to the system running a proxy server, or must join the P2P network in LAN environment. In our research, we aim to construct the ad hoc network among client terminals using Bluetooth to share the Web cache.

3 Reduction of Network Traffic in Collaborative Web Search

3.1 Use Case

Our method is intended to be applied to co-located collaborative Web search conducted by multiple people using mobile data communication. Each terminal is connected either directly to the Internet using the mobile data communication or through a mobile router. Co-located terminals communicate each other using Bluetooth. The number of terminals connected each other is limited to at most eight due to the capacity limitation of simultaneous Bluetooth connections.

3.2 Peer Cache Mechanism

3.2.1 Overview

An overview of the proposed system is shown in Figure 1. We build a proxy application which has the peer cache mechanism. An only thing that users have to do is setting the proxy IP address to "localhost". The proxy runs on every terminal, and sends and receives data toward the Web browser. The proxy manages the Web cache which is implemented with the local file system. In order to share the Web cache among terminals, the proxy communicates with the proxy of other terminals by using Bluetooth. The procedure of cache access is shown in Figure 2. The proxy works as follows.

1. The proxy gets an HTTP request for the Web content from the Web browser.
2. If the Web content is cached in its own Web cache, the proxy returns that Web content to the Web browser.
3. If the Web content is not cached in its own Web cache, the proxy sends a search message to all connected terminals with Bluetooth.
4. If the terminal responds with a message notifying that it has the Web content, the proxy sends a Web content request message to that terminal.
5. The proxy gets the Web content from that terminal and returns the Web content to the Web browser.
6. Because some of the terminals may not respond in some reason, timeout is set to every message. Non-responding terminals are regarded as those that do not have the Web content.
7. If the proxy gets Web content that is not cached in its own Web cache, the proxy caches the Web content in the local Web cache.

If the Web content request from the Web browser is using HTTPS protocol, the proxy only relays the request between the Web browser to the Internet.

We implemented the system on Android OS.

3.2.2 Referencing the Web Cache of Other Terminals

The proxy attempts to reference the Web caches of other terminals if the target Web content is not cached in its own terminal. This section describes a way to reference the Web cache of other terminals.

The proxy each other exchanges a message which implies to a unit of search and request of the Web content. A message consists of the method section, the ID section and the data section. The proxy decides how to operate according to a type of methods shown in Table 1. A URL hashed with MD5 is applied to the ID section of a message.

The proxy references the Web cache on other terminals by the following procedure.

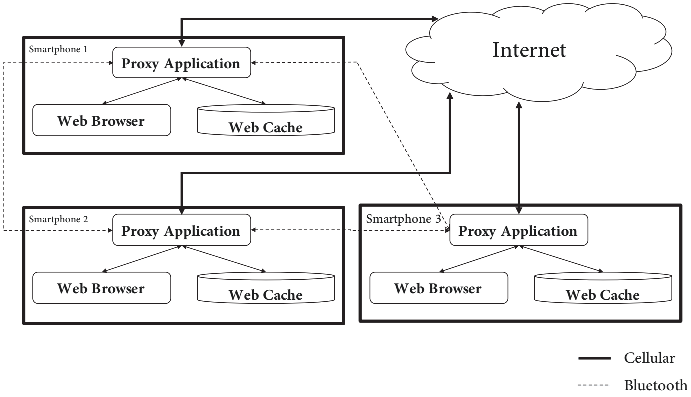


Fig. 1 System overview

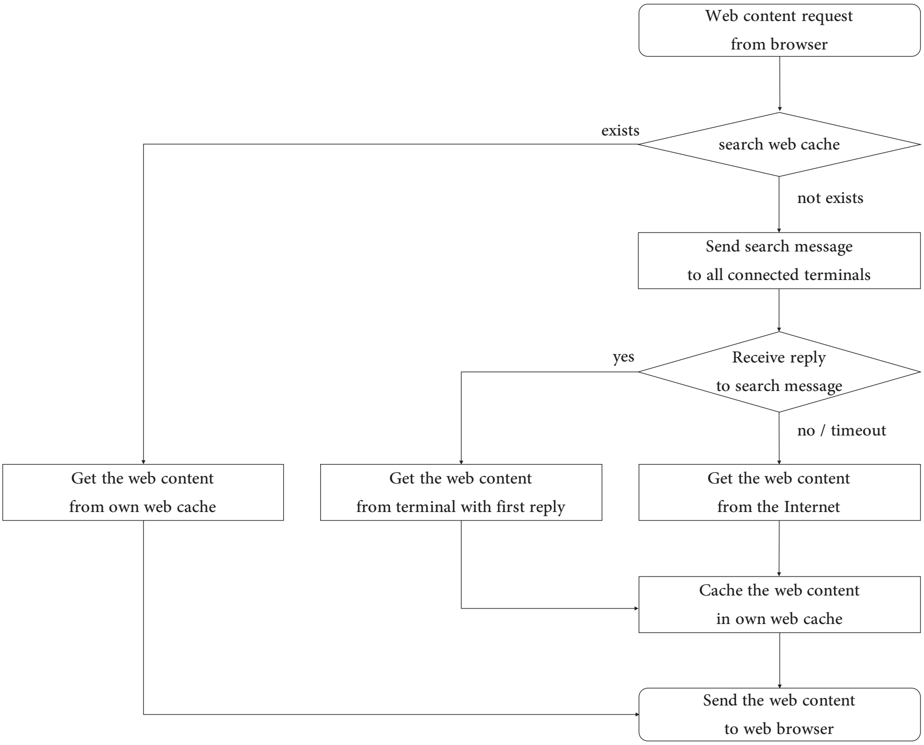

Web content request from browser
search web cache
exists
not exists
Send search message to all connected terminals
Receive reply to search message
yes
no / timeout
Get the web content from own web cache
Get the web content from terminal with first reply
Get the web content from the Internet
Cache the web content in own web cache
Send the web content to web browser

Fig. 2 Procedure of referencing to the Web cache

1. The proxy generates a *find* message with a message ID set to a hashed URL of the target Web content.
2. The proxy sends the *find* message to all connected terminals.

3. The proxy on each of the terminals receives the *find* message and searches for the target Web content in its local Web cache. Then the proxy replies with a *res*1 message including the search result in the data section. The data section is set to "yes" if the target Web content is found in its Web cache, otherwise "no".
4. If the proxy requesting the Web content receives the *res*1 message with the search result "yes", it sends a *get*1 message to that proxy which has the target Web content. At the same time, it sends a *quit* message to the proxy on other terminals. If none of the proxy replies with the search result "yes", this procedure is terminated.
5. The proxy which has the target Web content replies with a *res*2 message including the Web content in the data section.

In order to handle an unexpected failure of terminals, the requesting proxy waits for a response for a timeout period. When timeout happens, the procedure is terminated and the proxy tries to get the Web content from the Internet.

Table 1 Methods for Web cache reference

Method Type	Data Section	Role
find		search for Web content
*res*1	yes / no	reply to *search* message
*get*1		request Web content
*res*2	Web content	reply to *get*1 message
quit		quit sending and receiving message with an ID that this massage has

4 Evaluation

We have conducted experiments to examine how much Internet traffic is reduced.

4.1 Overview of the Experiment

In this experiment, two groups of four students performed a collaborative Web search task for 15 minutes. Topics for the task was to "find five restaurants around JR Okayama Station for a year-end party" or "find five restaurants around JR Kanazawa Station for a year-end party".

In the experiment, three ZenFone2 (Android 5.0) and one Nexus5 (Android 6.0) were used. The Chrome browser was used for browsing the Web. Before starting a task, the Web cache of each terminal was set to empty.

4.2 Result

In the experiment, the reduced Internet traffic was measured by keeping the size of the Web contents in a log file. Searched keywords were examined by using the history of Web searches.

Figure 3 and 4 show the search keywords used by the first and second group, respectively. In addition, the search engine is used four to six times. In the first group, all participants searched with “Kanazawa Station, Year-end Party (金沢駅 忘年会)” and accessed the first entry of search results. In the second group, three of four participants searched with “Okayama Station, Year-end Party (岡山駅 忘年会)”, and likewise accessed the first entry of search results. Then, one of the participants proposed an interesting restaurant, and other participants also searched with the name of its restaurant and accessed the restaurant Websites or restaurants rating sites. The keywords of “Kanazawa Station, Tsujiya Restaurant (金沢駅 つじや商店)” and “Kanazawa Station, Kuroya Restaurant (金沢駅 くろや)” correspond to those in the first group.

Table 2 and 3 show the Internet traffic of the first group and second group, respectively. The Internet traffic without using our method was 178.53MB in the first group, and 153.20MB in the second group. When only the local Web cache was used, the reduced Internet traffic was 18.31MB in the first group and 20.38MB in the second group. When the peer cache mechanism was used, the reduced Internet traffic was 38.23MB(21%) in the first group, and 37.16MB(24%) in the second group.

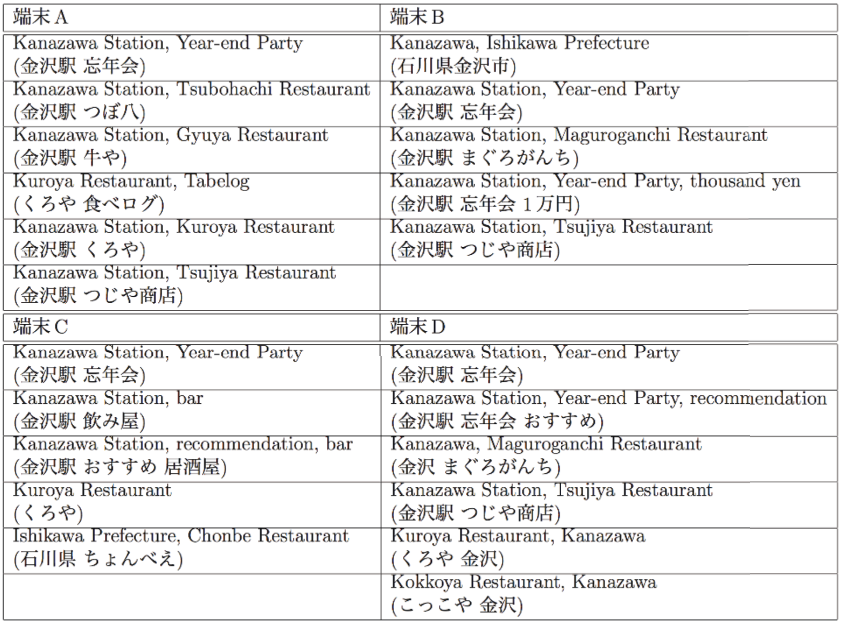
The results of the analysis on search keywords and the reduced Internet traffic reveal that the peer cache mechanism works effectively for a collaborative Web search task which shares a common goal among participants.

Table 2 Reduced Traffic in the First Group

	No cache	Only Own Web Cache	Our Approach
A	80.03MB	7.45MB	10.61MB
B	37.70MB	4.73MB	9.99MB
C	37.03MB	3.82MB	10.71MB
D	23.77MB	2.31MB	6.92MB
Total	178.53MB	18.31MB	38.23MB

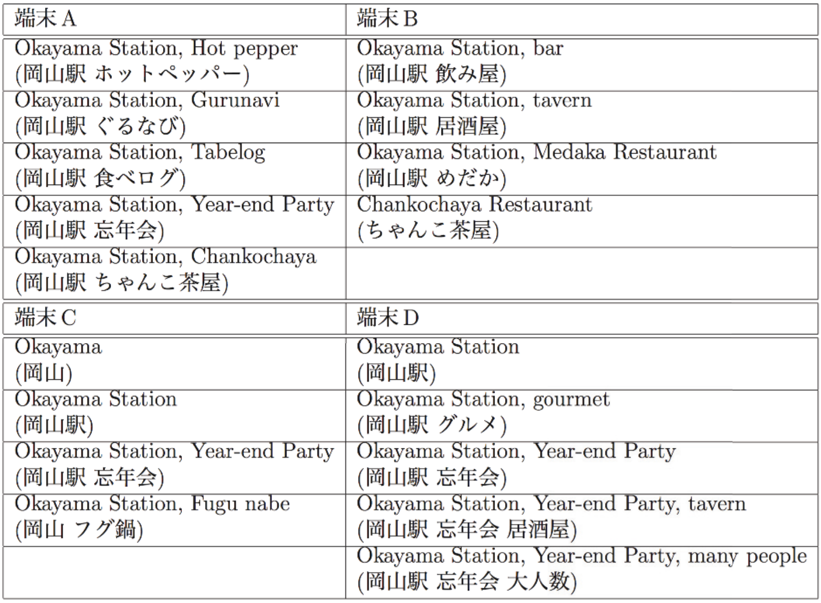
Table 3 Reduced Traffic in the Second Group

	No cache	Only Own Web Cache	Our Approach
A	42.44MB	4.94MB	9.11MB
B	47.02MB	6.94MB	10.91MB
C	31.11MB	4.17MB	9.10MB
D	32.63MB	4.33MB	8.04MB
Total	153.20MB	20.38MB	37.16MB



端末 A	端末 B
Kanazawa Station, Year-end Party (金沢駅 忘年会)	Kanazawa, Ishikawa Prefecture (石川県金沢市)
Kanazawa Station, Tsubohachi Restaurant (金沢駅 つぼ八)	Kanazawa Station, Year-end Party (金沢駅 忘年会)
Kanazawa Station, Gyuya Restaurant (金沢駅 牛や)	Kanazawa Station, Maguroganchi Restaurant (金沢駅 まぐろがんち)
Kuroya Restaurant, Tabelog (くろや 食べログ)	Kanazawa Station, Year-end Party, thousand yen (金沢駅 忘年会 1万円)
Kanazawa Station, Kuroya Restaurant (金沢駅 くろや)	Kanazawa Station, Tsujiya Restaurant (金沢駅 つじや商店)
Kanazawa Station, Tsujiya Restaurant (金沢駅 つじや商店)	
端末 C	端末 D
Kanazawa Station, Year-end Party (金沢駅 忘年会)	Kanazawa Station, Year-end Party (金沢駅 忘年会)
Kanazawa Station, bar (金沢駅 飲み屋)	Kanazawa Station, Year-end Party, recommendation (金沢駅 忘年会 おすすめ)
Kanazawa Station, recommendation, bar (金沢駅 おすすめ 居酒屋)	Kanazawa, Maguroganchi Restaurant (金沢 まぐろがんち)
Kuroya Restaurant (くろや)	Kanazawa Station, Tsujiya Restaurant (金沢駅 つじや商店)
Ishikawa Prefecture, Chonbe Restaurant (石川県 ちょんべえ)	Kuroya Restaurant, Kanazawa (くろや 金沢)
	Kokkoya Restaurant, Kanazawa (こっこや 金沢)

Fig. 3 Keywords searched in the first group



端末 A	端末 B
Okayama Station, Hot pepper (岡山駅 ホットペッパー)	Okayama Station, bar (岡山駅 飲み屋)
Okayama Station, Gurunavi (岡山駅 ぐるなび)	Okayama Station, tavern (岡山駅 居酒屋)
Okayama Station, Tabelog (岡山駅 食べログ)	Okayama Station, Medaka Restaurant (岡山駅 めだか)
Okayama Station, Year-end Party (岡山駅 忘年会)	Chankochaya Restaurant (ちゃんこ茶屋)
Okayama Station, Chankochaya (岡山駅 ちゃんこ茶屋)	
端末 C	端末 D
Okayama (岡山)	Okayama Station (岡山駅)
Okayama Station (岡山駅)	Okayama Station, gourmet (岡山駅 グルメ)
Okayama Station, Year-end Party (岡山駅 忘年会)	Okayama Station, Year-end Party (岡山駅 忘年会)
Okayama Station, Fugu nabe (岡山 フグ鍋)	Okayama Station, Year-end Party, tavern (岡山駅 忘年会 居酒屋)
	Okayama Station, Year-end Party, many people (岡山駅 忘年会 大人数)

Fig. 4 Keywords searched in the second group

5 Conclusion

We have proposed a method to reduce the Internet traffic by sharing the Web cache in the co-located collaborative Web search. We also have conducted an evaluation experiment to examine the efficiency of the method, and verified that the Internet traffic was actually reduced when performing a collaborative Web search task.

Performing collaborative Web search with exceeding the number of simultaneous Bluetooth connection is not currently considered. The security issues when accessing to Web cache on other terminals should also be considered. It is necessary to increase the applicability for use in real environment by solving these problems.

References

1. M. R. Morris and E. Horvitz, "SearchTogether: an interface for collaborative web search," in *Proceedings of the 20th annual ACM symposium on User interface software and technology*, 2007, pp. 3–12.
2. Ericsson, "Ericsson Mobility Report," http://www.ericsson.com/res/docs/2014/ericsson-mobility-report-november-2014.pdf.
3. "Squid," http://www.squid-cache.org.
4. Y. Matsumoto, E. Kawai, T. Okuda, and Y. Kadobayashi, "Web Cache System Using Peer-to-Peer Network," in *DPS Workshop*, 2002.

Building a Content Delivery Network among Tens of Nearby Devices Using Multihop Wireless Communication

Yuki Takeda, Syuhei Yamamoto, Ryo Nishide and Hideyuki Takada

Abstract As mobile device technology advances, content delivery among nearby devices becomes increasingly important at such places as where people gather face-to-face for seminars or meetings. Conventional ways such as cloud-based services, ad-hoc contents dissemination methods and file sharing peripheral devices are not enough to support this style of content delivery. In this paper, we propose a two-layered content delivery using multihop wireless communication built on an ad-hoc network. This network supports content delivery among tens of nearby devices even though the underlying ad-hoc network has limitation of the number of devices. We also show experiment results to reveal how the contents delivery proceeds in the network and how the network topology affects the delivery time. Important findings in the experiments are that 1MB data can be delivered to 16 devices in approximately 5.25 seconds using Wi-Fi based ad-hoc connection on iOS devices, and that reducing the number of devices simultaneously sending the data leads to the less required time for delivery.

1 Introduction

Mobile devices such as smartphone and tablet have been widely spread in recent years. Owing to this advancement, paper materials have changed to digitalized information, and opportunity to deliver the contents among mobile devices has increased. For example, we need to be able to deliver the contents to multiple participants in seminars or meetings.

Yuki Takeda · Syuhei Yamamoto
Graduate School of Information Science and Engineering, Ritsumeikan University
e-mail: {y_takeda, s_yamamoto}@cm.is.ritsumei.ac.jp

Ryo Nishide · Hideyuki Takada
College of Information Science and Engineering, Ritsumeikan University
e-mail: r_nishide@cm.is.ritsumei.ac.jp, htakada@cs.ritsumei.ac.jp

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering
and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_30

Several methods can be used for contents delivery. Examples of the methods are using the cloud-based services, the ad-hoc communication among multiple devices, and the storage device. However, these methods are not suitable for performing content delivery among multiple nearby devices. The communication speed may be reduced due to access congestion to a communication infrastructure. Although, ad-hoc communication is considered to be a method for communication among a group of nearby devices, the restricted number of simultaneous connections can be a problem.

In this paper, we propose a content delivery network suitable for co-located tens of devices. Our contribution can be summarized as follows.

- We build a two-layered contents delivery network among tens of nearby devices using multihop wireless communication even though the underlying ad-hoc network only supports at most 8 devices.
- We give experimental results of the contents delivery time in terms of how the contents delivery proceeds in the proposed network and how the network topology affects the delivery time.

The target of this work is iOS devices, where the Multipeer Connectivity framework is used for connecting the devices.

2 Contents Delivery among Group of Devices

This section discusses the existing contents delivery methods, and presents a scheme of our content delivery framework.

2.1 Content Delivery among Multiple Nearby Devices

Our work deals with contents delivery among multiple nearby devices. We assume that all devices involved are co-located in a place where ad-hoc communication can be performed mutually with other devices.

There are several methods for contents delivery, such as cloud-based services, ad-hoc contents dissemination methods and file sharing peripheral devices. However, there are some conditions needed in these methods. In the cloud-based services, all of the devices must have Internet connection in order to deliver or receive the contents. Therefore, this method is dependent on communication infrastructure, and the transmission speed may be decreased due to the access congestion.

Communication infrastructure is not necessary for ad-hoc network or peripheral device, but they are not always a good choice for the content delivery among tens of devices. In ad-hoc network, the transmission speed may decrease due to the increase of simultaneously connected devices. Otherwise, the number of devices may be restricted, which may prevent contents delivery with the large number of devices. For example, the number of simultaneous connections on Bluetooth communication is limited to at most 8 devices.

We employ multihop communication in order to deliver contents to multiple co-located devices, considering a method to overcome the limitation for ad-hoc communication.

2.2 Related Work

2.2.1 Wireless Communication between Devices

One of the methods to perform contents delivery among multiple nearby devices is to propagate the contents widely by forwarding the contents to nearby devices using wireless communication. A content-centric network proposed by Jung-Jae Kim uses multihop communication with Wi-Fi Direct[1]. FireChat[2] delivers messages to multiple distant devices using multihop communication via devices running the same application.

These systems are aimed to disseminate contents in a wide range, without assigning the destination of contents delivery. Moreover, the necessary hop count to reach the entire network cannot be estimated. It is not guaranteed that the contents reach the target devices.

The scheme of our study is assumed to mutually communicate with all devices co-located in the same place, and has a possibility to estimate the necessary hop count.

2.2.2 Utilization of Peripheral Devices

Several works have been conducted to support contents delivery using peripheral devices. Local Cloud Storage[3] performs contents delivery automatically by connecting the device embedded with storage space and wireless communication module. Portable Cloud[4] provides a movable high-performance wireless LAN access point embedded with PC cluster, battery and external devices. It provides communication infrastructure to support contents delivery with multiple devices.

These works can provide an environment to contents delivery among multiple nearby devices without being restricted to places and situations. However, these devices can not be always prepared anywhere. In this paper, we explore a method which does not require such devices in order to deliver the contents to multiple nearby devices.

2.3 Multipeer Connectivity Framework

Our content delivery network uses the Multipeer Connectivity framework on iOS for communication among devices. The Multipeer Connectivity framework supports ad-hoc communication with nearby devices by Wi-Fi communication through access points or ad-hoc connection among devices using Wi-Fi or Bluetooth communication. The communication method depends upon the device settings. There are two operations to communicate among devices: “Browse” to search and invite other devices and “Advertise” to notify the presence of device. The devices can be connected when the “Browsing” devices discover the “Advertising” devices. A device can perform “Browsing” and “Advertise” at the same time.

The participating devices are managed in a unit called “Session”. The number of devices which can be connected in a session is limited to at most 8.

A single device can join multiple sessions simultaneously. All devices are mutually connected in a session, and the contents are sent by specifying one or more destination devices.

3 Content Delivery Network

This section proposes the method of content delivery among tens of nearby devices using wireless communication network.

3.1 Network Overview

Network must overcome the limitation of the Multipeer Connectivity framework to deliver the contents among multiple nearby devices. A proposed network overview is shown in Figure 1. The network has a hierarchical structure which consists of two layers, super-layer and regular-layer. The necessary hop count can be estimated while determining a path to send contents to a specified destination. Each of super-layer and regular-layer has sessions where the devices are capable of communicating mutually with other devices in the same session. We refer to a device in super-layer session as super-peer, and a device in regular-layer session as regular-peer. Super-peer is assigned in each regular-layer session to communicate with different sessions through super-layer. All devices can communicate with each other in a different session by passing through super-layer. The maximum number of devices in each session can be configured from 2 to 8 for each layer. The maximum number of devices in the entire network is calculated by multiplying the maximum number of devices in super-layer and regular-layer.

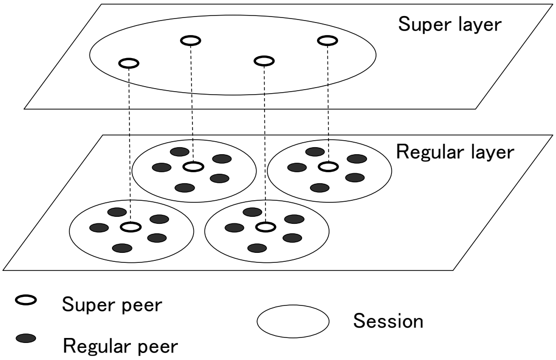


Fig. 1 Network Overview

3.2 Composing Network among Multiple Nearby Devices

Each device performs the following steps to construct the network.

1. Start advertising to and browsing other devices when the application is launched.
2. Start connection when other devices are found or when the invitation is received. If the number of super-peers is below the limit of the number of devices in the super-session, set itself as a super-peer. If invited by a super-peer, set itself as a regular-peer and stop browsing and advertising.
3. Super-peers stop browsing and advertising to the super-session if the number of super-peers reaches the maximum number of devices in the super-layer session. Then, the super-peers make a regular-session and start browsing other devices in the session.
4. A super-peer stops browsing if the number of regular-peers reaches the maximum number of devices in its regular-layer session.

In order for each device to be able to specify any device even in a different session, it must have a list of names of all devices and names of devices relaying the contents to each of the devices. When a new device participates in the network, all devices in the network share its name and relaying information about the new participating device through the super-peers.

3.3 Delivering Contents

This section describes a flow of delivering the contents in the proposed network. A device which initiates the delivery and a device which receives the contents perform the following to deliver contents.

A device which initiates the delivery: When a user specifies destination devices and the contents to send, the list of destination and the contents are sent to the destination. If the destination does not exist in the same session, they are sent to a super-peer in the same session.

A device which receives the contents: A device which has received the contents sends the list of destination and the contents to a super-peer of the destination in the same session which has not received the contents yet. If the list of destinations contains itself, then the contents is saved.

4 Evaluation Experiment

This section describes the verification if the proposed content delivery network works correctly in a real environment, and provides experiment results to validate the effect of transmission speed depending on the network topology.

4.1 Required Time for Contents Delivery

4.1.1 Details of the Experiment

In this experiment, we measure the time of the contents delivery for each of the devices in the network.

We prepared the environment where iOS devices were placed on a desk in a meeting room within an ad-hoc communication area. The number of participating devices was 16, and the maximum number of devices in a session in both super-layer and regular-layer was limited to 4. Wi-Fi ad-hoc communication was used, and the size of data to send was 1MB. In order to measure the time required to deliver the contents to all devices in the network, one of the regular-peer initiates the contents delivery by specifying all devices as the destination.

We measured the required time in each devices from the time when the delivery was initiated to the time when all devices finished to receive the contents. This trial has been conducted 5 times.

4.1.2 Experimental Result

Figure 2 shows the time course of the number of devices receiving the contents as time goes on. The contents have been delivered to all devices in 4.5 - 6 seconds.

The time course of the number of receiving devices is divided into three steps. The following steps show which devices are performing the content delivery in the network.

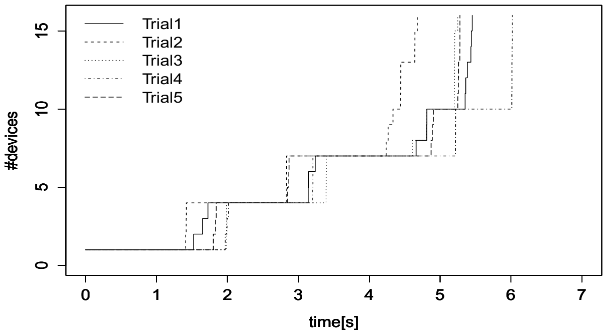


Fig. 2 The time course of the number of receiving devices

Step 1 In the regular-layer session, the contents are delivered from an initiating device to other devices in the same session.

Step 2 One of the devices receiving the contents in Step 1 also belongs to the super-layer session. This device then delivers the contents to all other devices in the super-layer session.

Step 3 Devices in the super-layer session receiving the contents in Step 2 also belong to one of the regular-sessions. Each of the devices then delivers the contents to all other devices in the same regular-layer session.

In the current implementation, content delivery to the next peer is started after the entire contents is received. Delivery time could be reduced if the peer simultaneously receives and sends the contents as file streaming.

4.2 Performance Evaluation for Variations of the Network Topology

4.2.1 Content of Experiment

Even when the contents are delivered in the same number of devices, the delivery time may change by changing the topology. In this experiment, the optimal topology of the network will be examined, by measuring the delivery time due to changes of the network topology.

The number of participating devices is 14. Two conditions were prepared in this experiment. In condition 1, the maximum number of devices in the super-layer session is seven and the maximum number of devices in regular-layer sessions is two. In condition 2, the maximum number of devices in the super-layer session is two and the the maximum number of devices in regular-layer sessions is seven.

Experimental environment and the data size are the same as those for the previous experiments. The trial has been conducted 5 times.

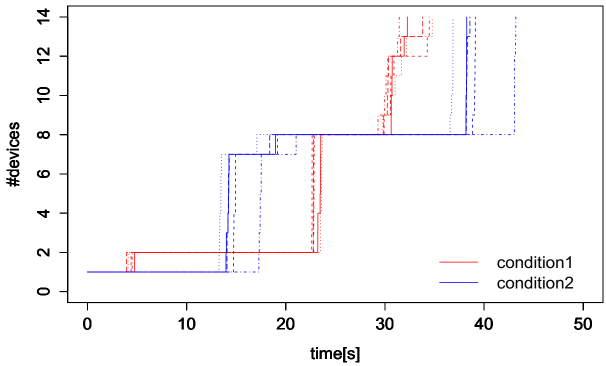


Fig. 3 The time course of the number of the receiving device(compare topology)

4.2.2 Experimental result

Figure 3 shows the time course of the number of receiving devices in each trial for each of the conditions. The result shows that the delivery time gets longer if the number of receiving devices increases in each of the sessions.

5 Conclusion

We have proposed a content delivery network using multihop wireless network communication among tens of nearby devices over the restriction of ad-hoc communication. Evaluation experiment was conducted to examine the characteristics of the network, and discussion has been made focusing on the time required for content delivery and the time course of the number of receiving devices. Modifying the algorithm to correspond with the leaving devices is left for one of the future works.

References

1. W.-S. Jung, H. Ahn, and Y.-B. Ko, “Designing content-centric multi-hop networking over wi-fi direct on smartphones,” in *Wireless Communications and Networking Conference (WCNC), 2014 IEEE*, 2014, pp. 2934–2939.
2. “Firechat,” https://itunes.apple.com/jp/app/firechat/id719829352.
3. Y. Arakawa, Y. Tanaka, S. Tagashira, and A. Fukuda, “Local cloud storage: Temporal local file sharing with previously paired wireless memories and cross-layer simultaneous data transmission mechanism,” in *Computer and Information Technology (CIT), 2012 IEEE 12th International Conference on*, 2012, pp. 640–646.
4. T. Yamanoue, S. Tetaka, K. Oda, and K. Shimozono, “Portable cloud computing system: A system which makes everywhere an ict enhanced classroom,” in *Proceedings of the 2014 ACM SIGUCCS Annual Conference on User Services Conference*, 2014, pp. 85–88.

A Workbook Design for Fill-in-Blank Problems in Java Programming Learning Assistant System

Nobuo Funabiki, Minako Dake, Khin Khin Zaw, and Wen-Chung Kao

Abstract To advance Java programming educations, we have developed a *Java Programming Learning Assistant System (JPLAS)* as a Web application system. JPLAS provides *fill-in-blank problems* for novice students to study the grammar and basic programming skills through *code reading*. To select the blank elements with grammatically correct and unique answers from a given code, we have proposed the graph-based *blank element selection algorithm*. Then, we generated and assigned fill-in-blank problems to students in Java programming course for two years. Unfortunately, the teacher selected original Java codes rather arbitrarily, which may degrade educational effects. Besides, this algorithm has been continuously extended to enhance the variations of generated problems. In this paper, we present a *workbook design for fill-in-blank problems in JPLAS* by collecting suitable Java codes from textbooks and Web sites, and applying the extended algorithm, so that they can be instantly and properly assigned to students. This workbook design consists of 15 categories with a considerable number of problems that follow the conventional learning order of Java programming. For the preliminary evaluation, we assign some problems to novice students. In the coming semester, we will use this workbook in the course to verify the adequacy of our proposal for novice students.

1 Introduction

As a reliable and portable object-oriented programming language, *Java* has been extensively used in a variety of practical systems, including Web application systems, mission critical systems at large enterprises, and small-sized embedded systems for real time controls. Thus, the cultivation of Java programming engineers has been

Nobuo Funabiki, Minako Dake, Khin Khin Zaw
Okayama University, Okayama, Japan, e-mail: funabiki@okayama-u.ac.jp

Wen-Chung Kao
National Taiwan Normal University, Taipei, Taiwan e-mail: jungkao68@gmail.com

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_31

highly demanded amongst industries. As well, a number of universities and professional schools have designed Java programming courses to deal with these demands.

To advance Java programming educations, we have developed a Web-based *Java Programming Learning Assistant System (JPLAS)* [1]-[6]. As a function, JPLAS provides the *fill-in-blank problem* to support self-studies of students who are novices at Java programming. The goal of the fill-in-blank problem is to improve self-studies of students for learning the grammar and basic programming skills through *code reading*.

In a fill-in-blank problem, a Java code with several blank elements is shown for each student, where it is requested to fill in the blanks. The Java code is designed to be of high-quality, most worth for code reading. An *element* is defined as the least unit of a code, such as a reserved word, an identifier, an operator in a conditional expression, or a control symbol. To be more precisely, a *reserved word* is a fixed sequence of characters that has been defined in the grammar to represent a specified function, and students must master it as a priority. An *identifier* is a sequence of characters defined in the code by the author to represent a variable, a class, or a method. An *operator* in a conditional expression, such as “”, “&&”, and “++”, often determines the algorithm or the processing in the code. A *control symbol* indicates other grammar elements such as “ . ” (dot), “ : ” (colon), “ ; ” (semicolon) , “ (“, “) ” (bracket), “ { “ , “ } ” (curly bracket).

To assist a teacher to prepare fill-in-blank problems in JPLAS, we have proposed the *blank element selection algorithm* to generate a fill-in-blank problem from a given code such that all blanks will have the grammatically correct and unique answers [2][3]. First, in this algorithm, a *compatibility graph* is generated by selecting a candidate element for a blank in the code as a *vertex*, and connecting any pair of vertices that can be blanked together by an *edge*. For this purpose, we define the conditions that a pair of elements cannot be blanked simultaneously. Then, we extract a *maximal clique* [4] of the compatibility graph, which becomes a maximal set of correct blank elements. Empirically, it is observed that a fill-in-blank problem will become more difficult when a larger number of elements are blanked [6]. Therefore, by blanking a subset of selected elements by the algorithm, a variety of fill-in-blank problems with different levels may be produced.

To evaluate the effectiveness of fill-in-blank problems in JPLAS, we assigned them to students in the Java programming course in our department for two years [5][6]. Unfortunately, the teacher selected Java codes for them rather arbitrarily, which may degrade the educational effects. Besides, the blank element selection algorithm has been continuously extended to enhance the variations of generated problems to deal with students at various levels [3].

In this paper, we present a *workbook design of fill-in-blank problems for JPLAS* by collecting a set of suitable Java codes from textbooks and Web sites for Java programming introduction and applying the extended algorithm so that they can be assigned to students by the teacher instantly and properly. This workbook design consists of 15 categories that are arranged in the conventional learning order of Java programming, where each category contains a considerable number of problems.

For the preliminary evaluation, we assign eight problems in the workbook to four novice students and analyze the results. In the coming semester, we will use this workbook in our Java programming course to estimate the adequacy of it for novice students.

This paper is organized as follows: Section 2 shows related works. Section 3 introduces the extended blank element selection algorithm. Section 4 presents the workbook design of fill-in-blank problems for JPLAS. Lastly, Section 5 concludes this paper with future studies.

2 Related Works

In this section, we show related works to the fill-in-blank problem in JPLAS for Java novice students.

In [7], Piech et al. presented a *Hidden Markov Model (HMM)* to create a graphical model of how students in a fundamental programming course progress through a homework assignment. They found that several sink states or milestones exist where students clearly had serious functional problems, and once a student transitioned to a state, he/she remained there with a high probability through several code updates.

In [8], Hosseini not only presented a fine-grained indexing tool for the automatic indexing of Java problems, but explored the application during exam preparations. It was discovered that smaller grain size of knowledge units was critical to finding the sequence of problems to fill the gaps in student knowledge.

In [9], Delev et al. reported the data generated by usage of the system *Code* that supports compilations, executions, and testing of the test cases of programming problems for exercises and examinations in programming courses for C, C++, and Java. The results comprised three categories, namely, recursion, matrix, and files. Through the questionnaire, it was presumed that students prefer example problems with solutions, helps in locating and fixing errors in solutions, and relevant materials and similar problems.

In [10], Brown et al reported a study to determine if programming educators form a consensus about which Java programming mistakes are the most common and found that educators formed a weak consensus about which mistakes are most frequent. Experimentally, students most often made mistakes in "mismatched parentheses " , " calling method with wrong types " , and " missing return statement " . These three frequent mistakes were also reported in [11].

In [12], Busjahn et al. studied the visual attention distribution when 15 programmers with various expertise read short source codes. Furthermore, it was found that most attention is oriented towards understanding of identifiers, operators, keywords, and literals, and relatively little reading time for separators. This result supports the importance of blank element selections in our fill-in-blank problems.

Table 1 Vertex information in constraint graph.

item	content
symbol	symbol of element
line	row index of element
column	column index of element
count	number of element appearances
order	appearing order of element in the code
group	statement group index partitioned by “ { ” and “ } ”
depth	number of “ { ” from top

3 Overview of Extended Blank Element Selection Algorithm

In this section, we overview the extended *blank element selection algorithm* [2][3] using the *constraint graph* that is generated to describe the constraints in the blank element selection.

3.1 Vertex Generation for Constraint graph

In the constraint graph, each vertex signifies a candidate element for being blank. The candidate elements or vertices are extracted from the Java code using open source software *JFlex* [13] and *jay* [14]. Each vertex contains the associated information in Table 1 that is necessary for the following edge generation.

JFlex is a lexical analyzer generator for a Java code, which is also coded by Java. It transforms a code into a sequence of lexical units that represent the least meaningful elements to compose the code. It can classify each element in the code into a reserved word, an identifier, a symbol, or an immediate data. For example, a statement “`int value = 123 + 456;`” is divided into “`int`”, “`value`”, “`=`”, “`123`”, “`+`”, “`456`”, and “`;`”. Unfortunately, *JFlex* cannot identify an identifier among a class, a method, or a variable. Thus, *jay* is applied as well. Since *jay* is a syntactic parsing program based on the LALR method, it can identify an identifier.

3.2 Edge Generation for Constraint graph

An edge is generated between any pair of two vertices or elements that should not be blanked at the same time. There are three categories to represent the constraints in selecting blank elements with unique answers.

3.2.1 Group Selection Category

In the *group selection category*, all the elements related to each other in the code are grouped together. To generate edges between them of the constraint graph, first, the vertex that has the largest number of incident edges in the constraint graph is selected for each group. Then, edges are generated between this vertex and the other vertices to confirm that at least this selected element is not selected for blank. It is noted that if the element with many contending elements in the constraint graph is not blanked, more elements can be blanked as a result. Six conditions are included in this category.

(1) Identifier appearing two or more times in the code

The multiple elements representing the same identifier of a variable, a class, and a method by using the same name, are grouped together. If all of such elements are blanked, a student cannot answer the original identifier.

(2) Pairing reserved words which are composed of three or more elements

The three or more elements representing the reserved words in pairs are grouped together. If all of them are blanked, the unique answers may become too difficult as the following two cases:

- switch-case-default
- try-catch-finally

(3) Data type for variables in equation

The elements representing the data types for variables in one equation are grouped together. For example, in " `sum = a + b` ", the data types of the three variables, " `sum` ", " `a` ", and " `b` ", must be the same.

(4) Data type for method and its returning variable

The elements representing the data type of a method and its returning variable are grouped together.

(5) Data type for arguments in method

The elements representing the data type of an argument in a method and its substituting variable are grouped together.

After every group is found, the groups from (3)-(5) that contain an overlapped element are merged together into one group because they must be the same data type.

(6) Operators in conditional expression

The elements representing the operators in the same conditional expression are grouped together. For example, in the conditional expression: " `for (int i = 0; i < j && i != k; i ++)` ", the four elements: " `<` ", " `&&` ", " `!=` ", and " `++` ", are grouped together.

3.2.2 Pair Selection Category

In the *pair selection category*, the elements appearing in the code in pairs are grouped together. For each pair, an edge is simply generated between the two corresponding vertices to assure that at least one element be not selected for blank.
(1) Elements appearing continuously in a statement

The non-blanked elements in a problem code become excellent hints to solve the fill-in-blank problem. Here, a *problem code* represents a Java code with blanked elements that should be filled by students. As more non-blanked elements exist between blanked elements, the fill-in-blank problem becomes easier. To control the number of non-blanked elements between two blanks, the *blank gap number BG* has been introduced. Then, for each element, any element in the same statement in the code that exists within its BG neighbors is paired. For $BG > 0$, the two elements connected with a dot (" . ") are also paired here.
(2) Variables in equation

The elements representing any pair of the variables in an equation are paired. If both are blanked, it will become impossible to access the unique answers. For example, for " `sum = a + b` ", " `sum = b + a` " is also feasible.
(3) Pairing reserved words

The two elements representing the paring reserved words are paired. If both are blanked, the unique correct answers may not be guaranteed. The following five paring reserved words are considered:

- if-else
- do-while
- class-extends
- interface-extends
- interface-implements

(4) Pairing control symbols

The two elements representing a pair of control symbols, namely " (,) " (bracket) and " {, } " (curly bracket), are paired. The novice students are expected to thoroughly examine them in their codes to decrease the amount of mistakes.

3.2.3 Prohibition Category

In the *prohibition category*, an element is prohibited from the blank selection because it does not satisfy the uniqueness with the high probability. There are three conditions for this category. However, an element in a fixed sequence of elements indicating a specific meaning in a Java code, such as "`public static void main`" and " `public void paint(Graphics g)` ", is excluded from this category, because they should be mastered by students.

(1) Identifier appearing only once in code

The selected element representing the identifier in this category appears only once in the code. If it is blanked, a student cannot answer the original identifier.

(2) Access modifier

The element representing an access modifier for an identifier is selected for this category. If it is blanked, either “`public`”, “`protected`”, or “`private`” can often be grammatically correct.

(3) Constant

The element representing a constant is selected for this category. If it is blanked, a student cannot answer the original constant.

3.3 Compatibility Graph Generation

By taking the complement of the constraint graph, the *compatibility graph* is generated to symbolize the pairs of elements that can be blanked simultaneously.

3.4 Maximal Clique Extraction of Compatibility Graph

Finally, a maximal clique of the compatibility graph is extracted by a simple greedy algorithm to find the maximal number of blank elements with unique answers from the given Java code. A clique of a graph represents its subgraph where any pair of two vertices is connected by an edge. It is emphasized that as more blanked elements continue in a problem code, it becomes harder. To control the number of continuously blanked elements, the *continuous blank number CB* has been introduced, where the maximum of *CB* elements can be blanked continuously in a statement. We note that when *CB* is 2 or larger, *BG* must be 0.

The procedure for our algorithm is described as follows:

1) Calculate the degree (= number of incident edges) for each vertex in the compatibility graph.
2) Select one vertex among the vertices whose degree is the maximum. If two or more vertices have the same maximum degree, select one randomly.
3) If the selected vertex is a *control symbol* and the number of selected control symbols exceeds 1/3 of the total number of selected vertices, remove this vertex from the compatibility graph and go to 5).
4) Add the selected vertex for blank if the number of continuously blanked elements does not exceed *CB*, and remove this vertex from the compatibility graph. If it is blanked, its non-adjacent vertices are also removed.
5) If the compatibility graph becomes null, terminate the procedure.

3.5 Fill-in-blank Problem Generation

In the maximal clique procedure, 3) is used to sustain the total number of blank control symbols, because a code is generally composed of plenty of control symbols. Here, we examined the average number of blanks for control symbols and other symbols by the algorithm. Then, we empirically selected $1/3$ as an appropriate ratio to generate the feasible *fill-in-blank problems* for novice students.

4 Workbook Design of Fill-in-blank Problems

In this section, we present a workbook design of fill-in-blank problems in JPLAS.

4.1 Contents of Workbook

In the workbook of fill-in-blank problems in JPLAS, we collected source codes from textbooks for Java programming [15]-[17], and related Web sites [18]-[28]. By referring to the contents of the textbooks that have been used in the introductory Java programming course in our department, we select 15 categories to classify the Java codes as in Table 2. The first 12 categories (ID: 1-12) are related to Java grammar, and the remaining three categories (ID: 13-15) are for applications. The data structure, sorting algorithms, and graph algorithms are selected since they have been educated in our department in the corresponding courses. In future works, we will increase the number of categories that are relevant to significant applications of Java programming for novice students. Subsequently, to generate fill-in-blank problems using the blank element selection algorithm, we adopt three combinations of the two parameters (*BG*, *CB*), as well, to examine the numbers of selected blanks by changing them.

Table 2 shows the code topic, the number of codes, the average number of statements or lines (LOC) for one code, and the average number of blanks with (3, 1), (1, 1) and (0, 3) for (*BG*, *CB*) in each category. The number of selected blanks is smaller for the problems related to Java grammar (ID: 1-12) than those for applications (ID: 13-15). It could be caused by the fact that codes for grammar usually have shorter statements than codes for applications. The former codes were rather artificially made for grammar studies, whereas the latter codes were made for practical use. The results prove that as the number of blanks increases, the problem becomes more difficult [6]. Thus, depending on performances of students in the course, the teacher would be suggested to carefully select problems as assignments.

Table 2 Workbook design of fill-in-blank problems.

category ID	code topic	# of codes	ave. # of lines (LOC)	ave. # of blanks (*BG*, *CB*) (3, 1)	(1, 1)	(0, 3)
1	variable	5	9.6	8.6	9.4	13.2
2	operator	7	9.43	8.86	9.0	14.14
3	conditional statement	6	21.17	15.33	15.83	27.33
4	loop, break, continue	15	13.33	10.67	11.4	18.0
5	array	11	16.91	16.36	19.91	29.54
6	class: field, method, member	4	16.5	10.25	13.25	20.75
7	class: overload, constructor, this	4	21.0	14.75	19.5	26.25
8	class: library, string, class method	6	17.17	16.0	17.83	26.83
9	class: inheritance, superclass, override	6	20.5	13.67	15.5	23.33
10	interface	5	24.0	16.0	18.2	26.8
11	package, file	3	27.0	16.0	20.0	31.0
12	exception	8	21.0	16.75	17.5	25.0
13	data structure	2	30.5	19.5	28.0	39.5
14	sorting algorithms	4	26.75	22.0	39.0	51.0
15	graph algorithms	7	51.0	42.71	76.14	102.0

Table 3 Trial application results for four students.

problem ID	category ID	LOC	*BG*	*CB*	ave. # of blanks	ave # of corrects	ave. correct rate (%)
Q1	6	32	1	1	22	20.75	94.31
Q2	9	28	3	1	21	18.75	89.29
Q3	4	26	1	1	17	16	94.12
Q4	3	19	0	3	24	23.75	98.96
Q5	5	18	1	1	19	18	94.74
Q6	7	18	1	1	23	22	95.65
Q7	3	12	0	3	20	19.5	97.5
Q8	4	11	0	3	18	16	88.89

4.2 Trial Application Results to Novice Students

Next, we selected eight problems related to Java grammar in this workbook design and asked four novice students from Indonesia to solve them. These students, with complete knowledge in C programming, have studied Java programming for merely 10 days. Table 3 reflects the category ID, the number of statements (LOC), the adopted values of (*BG*, *CB*) for the problem generation, the average number of selected blanks, and the average correct answer rate for each problem. Here, we note that the original source codes for these problems come from [15] or [16].

Table 3 indicates that the two problems Q2 and Q8 had lower correct answer rates than the others. The following **Problem Q2** and **Problem Q8** illustrate their problem codes respectively. As shown there, the problem code for Q2 includes the *object array* at lines 19-22, and the code for Q8 includes *double loops* at lines 3 and 5. It can be considered that they are difficult for the novice students. On the other hand, LOC and the values of (*BG*, *CB*) are not sensitive in solving performances of students, because these codes for Java grammar have a considerable sum of simple short statements. It is necessary to investigate their performance progress when students solve fill-in-blank problems using application codes such as sorting algorithms and graph algorithms, which will be in our future studies.

Problem Q2

```
class  _1_ {
   protected int num;
   protected double gas;
   public Car() {
      _2_  = 0;
      _3_  = 0.0;
      System.out.println("generate a car");
   }
}
_4_  RacingCar extends Car {
   private int course;
   public  _5_ () {
      _6_  = 0;
      System.out.println("generate a racing car");
   }
}
_7_  CodeQ2 {
   public  _8_  void main( _9_ [] args) {
      _10_ [] cars;
      cars =  _11_  Car[2];
      _12_ [0] =  _13_  Car();
      _14_ [1] =  _15_  RacingCar();
      _16_ (int i=0; i<cars.length; i++){
         Class clsName =  _17_ [i] _18_ getClass();
         _19_ .out. _20_ (class of (i+1) +
         "th object is" +  _21_  );
      }
   }
}
```

Problem Q8

```
public  _1_  CodeQ8 {
   public  _2_   _3_  main( _4_ [] args) {
      _5_  ( _6_  i = 0;  _7_  _8_  10;  _9_ ++) {
         _10_ .out.print _11_ i + ":");
         _12_  (int j = 0; j  _13_  _14_ ; _15_ ++) {
            System.out. _16_ ("*");
         }
         _17_ .out. _18_ ("");
      }
   }
}
```

For reference, the problem code for Q4 is revealed as follows. With a simple structure of “ `if` ” “ `else` ”, the average correct rate for Q4 is the highest among the eight problems.

Problem Q4

```
_1_  java.io.*;
_2_  CodeQ4{
   public  _3_   _4_  main( _5_ [] args)  _6_
   IOException{
      System. _7_ .println("Please enter an integer");
      BufferedReader br = _8_   _9_ ( _10_
      InputStreamReader( _11_ .in));
      _12_  str =  _13_ .readLine( _14_ ;
      int res = Integer.parseInt( _15_ );
      _16_ ( _17_  == 1) {
         System.out. _18_ ("the input is 1");
       }
       else _19_ ( _20_  == 2) {
          _21_ .out.println("the input is 2");
       }
       _22_ {
           _23_ .out. _24_ ("Please enter 1 or 2");
       }
    }
 }
```

5 Conclusion

In this paper, we presented a workbook design of fill-in-blank problems in JPLAS by collecting a set of Java codes from textbooks and Web sites, and applying the extended blank element selection algorithm. This workbook design consists of 15 categories that are arranged in the conventional learning order of Java programming. We showed trail application results of eight problems with four novice students. In the coming semester, we will assign problems in the workbook to students in the Java programming course, evaluate the effects of LOC and the two parameter values in their solving performances, and verify the adequacy of this workbook in Java programming educations for novice students.

References

1. Funabiki, N., Matsushima, Y., Nakanishi, T., Amano, N.: A Java programming learning assistant system using test-driven development method. Int. J. Comput. Sci. 40(1), 38-46 (2013)
2. Tana, Funabiki, N., Ishihara, N.: A proposal of graph-based blank element selection algorithm for Java programming learning with fill-in-blank problem. In: IMECS2015, pp. 448-453 (2015)

3. Zaw, K. K., Funabiki, N., Kuribayashi, M.: A proposal of three extensions in blank element selection algorithm for Java programming learning assistant system. to appear In: GCCE2016 (2016)
4. Garey, M. R., Johnson, D. S.: Computers and intractability: A guide to the theory of NP-completeness. Freeman, New York (1979)
5. Tana, Funabiki, N., Ishihara, N., Kao, W.-C.: Correlation analysis of fill-in-blank problem solutions to final programming results in Java programming course. In: GCCE2015, pp. 348-349 (2015)
6. Funabiki, N., Tana, Ishihara, N., Kao, W.-C.: Analysis of fill-in-blank problem solution results in Java programming course. to appear In: GCCE2016 (2016)
7. Piech, C., Sahami, M., Koller, D., Cooper, S., Blikstein, P.: Modeling how students learn to program. In: SIGCSE '12, pp. 153-160 (2012)
8. Hosseini R., Brusilovsky, P.: JavaParser: a fine-grained concept indexing tool for Java problems. In: AIED 2013 (2013)
9. Delev T., Gjorgjevikj, D.: A study on implementation and usage of web based programming assessment system: Code. In: ICT Innovations 2014, pp. 76-85 (2014)
10. Brown, N. C. C., Altadmri, A.: Investigating novice programming mistakes: educator beliefs vs student data. In: ICER ' 14 (2014)
11. Altadmri, A. Brown, N. C. C.: 37 million compilations: investigating novice programming mistakes in large-scale student data. In: SIGCSE ' 15 (2015)
12. Busjahn, T., Bednariky, R. Schulte, C.: What influences dwell time during source code reading? Analysis of element type and frequency as factors. In: ETRA 2014, pp. 335-338 (2014)
13. JFlex. http://jflex.de/
14. jay. http://www.cs.rit.edu/~ats/projects/lp/doc/jay/package-summary.html
15. Yuki, H.: Java programming lesson. Softbank Creative, Tokyo (2012), http://www.hyuki.com/jb/#download
16. Takahashi, M.: Easy Java. Softbank Creative, Tokyo (2013), http://homepage3.nifty.com/~mana/yasaj.html
17. Kondo, Y.: Algorithm and data structure for Java programmers. Softbank Creative, Tokyo (2011)
18. ITSenka. http://www.itsenka.com/
19. tutorialspoint, http://www.tutorialspoint.com/java/index.htm
20. Java program samples. http://www7a.biglobe.ne.jp/~java-master/samples/
21. Shellsort. http://www.thelearningpoint.net/computer-science/arrays-and-sorting-shell-sort-with-c-program-source-code
22. Sinapova, L.: Lecture Notes, http://faculty.simpson.edu/lydia.sinapova/www/cmsc250/LN250_Weiss/Contents.htm
23. Chang, S. K.: Data structures and algorithms. World Scientific Pub., New Jersey (2003)
24. Dijkstra Algorithm. http://www.ifp.illinois.edu/~angelia/ge330fall09_dijkstra_l18.pdf
25. Prim Java. http://cs.fit.edu/~ryan/java/programs/graph/Prim-java.html
26. Graph Java. http://www.sanfoundry.com/java-program
27. Depth First Search. https://en.wikipedia.org/wiki/Breadth-first_search
28. Breadth First Search. https://en.wikipedia.org/wiki/Depth-first_search

Design and implementation of software consistency detection system based on Netty framework

Jun Yang[1], Haipeng Zhang[1,3], Lifang Han[2], Baojiang Cui[1,3], Guowei Dong[4]

[1] School of Computer Science, Beijing University of Posts and Telecommunications, China
junyang@bupt.edu.cn; cuibj@bupt.edu.cn;

[2] China Electric Power Research Institute, Beijing, China
hanlifang@epri.sgcc.com.cn;

[3] National Engineering Laboratory for Mobile Network Security, China
harperzhang1989@163.com;

[4] China Information Technology Security Evaluation Center, Beijing, China
donggw@itsec.gov.cn

Abstract. For studying the consistent detection problem of software code deployed on the server, analysing the existing domestic and foreign consistency detection technology, based on the Netty framework and consistent hash comparison, achieved a software consistency detecting system for remote server. The system can effectively detect software's consistency information which deployed on the server, and realize communication between server and client by Netty, including comparing task management, comparing information interaction, and through with traditional IO, asynchronous NIO of comparative tests proved the effectiveness and efficiency of the system.

1 Introduction

Computer as an important tool to improve productivity, occupies a very important position in the development of modern production, with the development of computer science, the scope of application of computer software more and more, software structure is becoming more and more complex, once the structure or content is changed, to those of widely used software, the problems brought by the loss is huge, software consistency detection technology arises at the historic moment.

Software consistency detection is mainly used to detect whether the software being used is consistent with the initial installed software, to detect whether the software being used is malicious tampering or not. This technology can be applied in many fields such as digital library, intellectual property protection and software version consistency checking.

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_32

2 Related work

The current domestic and international mainstream software consistency detection technology are the following.

2.1 Software digital watermarking technology

Software digital watermarking [1] is one of the many branches of information hiding, and its mainly protection object is computer code, including source code and machine code, to avoid or reduce the risk of copying and tampering. Because computer code can't tolerate any mistakes, some traditional methods of using a computer program that can be used to modify the computer program in the allowable error range will not be applied to computer software. For example, the modified LSB (Lease-Significant-Bits) bit method for digital images. The software protection method based on digital watermarking is an ex post method. It can not prevent the occurrence of tampering, but it can be used to prove the ownership of intellectual property.

Using digital watermarking can effectively detect whether the software is modified or not. Firstly, use the digital watermarking technology to generate the only digital watermark, hidden in the software executable program code. When the program software is modified, then the software executable program code generated digital watermark, compared with the previous digital watermark, the results are not consistent with the results obtained.

Advantages: software digital watermark with high accuracy, does not produce significant impact on efficiency of program execution, and has strong resistance to attack.

Disadvantages: the need to insert additional code, need to be carefully prepared mute function and its call method, otherwise easy to be experienced by decompile wised, erasing the watermark.

2.2 Consistency detection technology based on the License

The consistency detection technology based on License [2] is the signature of some important program segments in the software by using the technology of cryptography. Then randomly verify the signature in the process of the execution of the program, if the signature verification fails, the program has been tampered with, can interrupt program execution or other measures. It can also encrypt program segmentation, runtime decryption step by step, step by step to clear, so that only run a major program segments can change the next layer of the program code from password into a clear text, ensuring important procedures section cannot be skipped.

Inspection agencies according to the testing software, provided by the customer, with the private key d of the RAS encrypt software features, and in some transform algorithm to generate software check code (should be one to identify whether formal registration code) and key IDEA related to the software source code. The key IDEA with the private key of the RAS d keys encrypted stored in the beginning of the License file. Then software registration code encrypted with a key IDEA in the

License file. In order to further enhance security, in the generated License file can be inserted into a number of redundant information.

Advantages: ensure the integrity and software version, only version match the License can ensure the software properly install for use

Disadvantages: License in the process of use may be forged and tampered with, resulting in software information not consistent, cannot achieve the effect of check.

2.3 Consistency detection technology based on the USB can

USB encryption lock identity authentication [3,4] is a kind of convenient and secure authentication technology developed in recent years. It uses a combination of software / hardware methods, has been widely used in various fields, such as log on the local computer for identity authentication, for all kinds of application software system authentication, etc. For the software deployment version and safety testing version inconsistency problem, using a USB encryption lock identity authentication technology increase the authentication mechanism in the application software can well solve the security problems existing in the process of software version is not consistent.

Advantages: completely avoid the problem of software version inconsistencies, at the same time, due to the complex encryption algorithm of software certificate encrypted, making software in the use of the process of security risks resolved.

Disadvantages: increase the complexity of software use and hardware costs.

Based on the research and analysis of the existing software consistency detecting method, this paper designs and implements a dynamic real-time software consistency detecting system. This system can detect the software version information running on the server regularly, and compare with the version information in the remote system. Finally, design experiment to verify the function and performance of the software consistency detecting system, at the same time, compared with the existing software consistency checking tools, and the results show that the system has good performance. The main characteristic of the system is can real-time detect the software version information running on the remote host, and does not require modify the software program source code, the function of the management system powerful and easy to use.

3 Design and implementation

3.1 Design Principle

In order to make the system more acceptable and practical, the paper fully considers the functional requirements of the existing software conformance testing technology in the system design stage. Its main functional requirements are:

Accuracy

An important configuration file for the software, possible minimal changes will cause the entire software at run time a serious problem. Therefore, software consistency detection technology need more fine-grained, the software structure and the content of small changes can be detected, nip in the bud. [5]

Efficiency and concurrency

With rapid the development of the Internet, the use of software increased dramatically. When a large number of software needs to be consistent, in order to ensure the user experience, it is necessary to complete a large number of software hash generation and comparison in the acceptable time range. A large number of consistency testing equipment (client) at the same time connected to the consistency of the detection system (server), need for network concurrency, and require the system to deal with high efficiency and concurrency.

Real time interactive capability

To monitor the real-time data of the consistency detection equipment (client), so the consistency detection system (server side) must receive the data of the remote data transmission in real time. And taking into account the future and existing consistency detection system for interactive control, so the system needs a good scalability and good system interface.

3.2 The Framework

The software consistency detection system is divided into data acquisition layer, communication network layer, data layer, core business layer and access layer. Client monitoring module can provide online monitoring information, including software configuration file, the core code files; Communication network based on the Netty framework [6] to achieve the information transmission between client and server, to meet the requirements of system communication, high concurrency, real-time interaction and other requirements. To the client to collect information through the communication network, pass to the server, compare to sample library information, and do the core business operations, contains detection client online, consistent detection, send timed tasks, compare for the sample library management operation. Finally, the results are displayed to the user in the access layer. The overall structure of the system as shown below in Fig. 1:

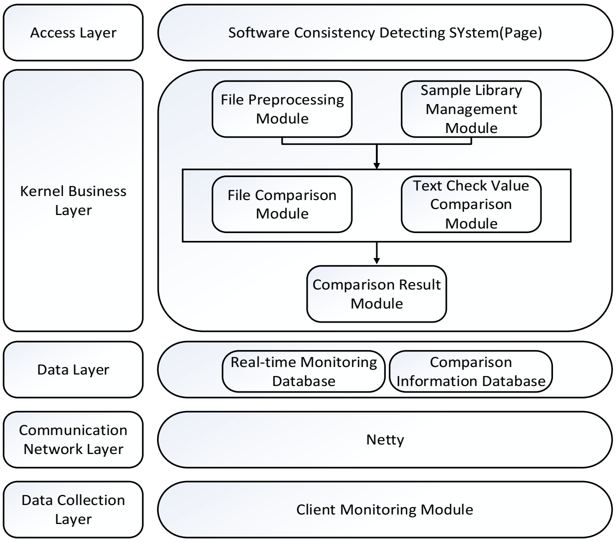

Access Layer
Software Consistency Detecting SYstem(Page)
Kernel Business Layer
File Preprocessing Module
Sample Library Management Module
File Comparison Module
Text Check Value Comparison Module
Comparison Result Module
Data Layer
Real-time Monitoring Database
Comparison Information Database
Communication Network Layer
Netty
Data Collection Layer
Client Monitoring Module

Fig. 1. The software consistency detection system Frame

The implementation of the whole consistency detection technology includes two parts: the consistency checking client and the consistency checking system.

The consistency checking client is used for real-time monitoring of the client, when the system request is received, the request data is sent through the network to the consistency checking system, which is used to complete the consistency checking.

The consistency checking system is based on B/S architecture, which is divided into presentation layer, core business layer and data layer:

The data layer of the system uses the MySQL database to store the sample library and the comparison result. The Hibernate framework is used to complete the connection of the application and the database.

The core business layer of the system includes the detection of the client's online state, consistency checking, sending the timing task, comparing the sample library management and other operations. Detection of online state can detect all the client is online detection. Consistency testing using the client sent the hash file with the server to save the sample library file for comparison, if the results are consistent, that the software is consistent, has not been modified. Sample library management includes the registration documents of the increase, delete, change, search and other operations and sample library of the pretreatment operation. System core module relations as shown below in Fig. 2:

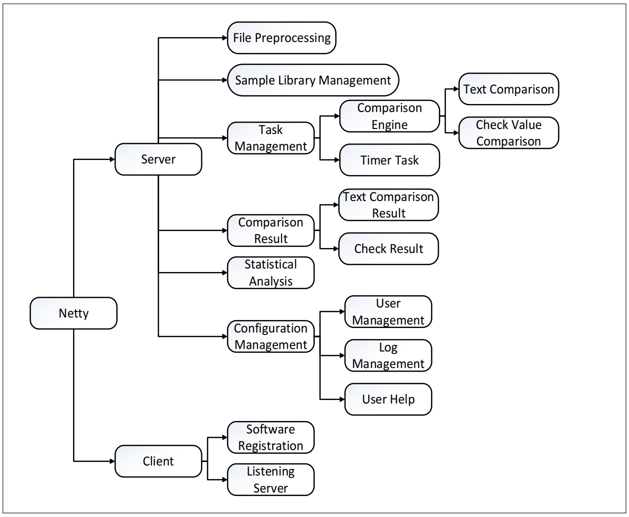


Fig. 2. System core module

3.3 Design And Implementation Of Consistency Detecting

In this paper, we use the communication network layer design based on Netty and the consistency comparison principle based on hash to solve the problem of the network communication processing efficiency and the interaction ability of the system.

Netty is based on NIO, which provides users with asynchronous, event driven development mode, through which developers can quickly and simply develop high-performance, reliable network applications. Netty technology has greatly optimized the process of the development of network application. Not only to ensure the development of convenient and fast, but also to ensure that the performance of the application by the development of its stability and high scalability.

Netty in accordance with the Reactor mode design and implementation, its server communication sequence as shown below in Fig. 3:[9]

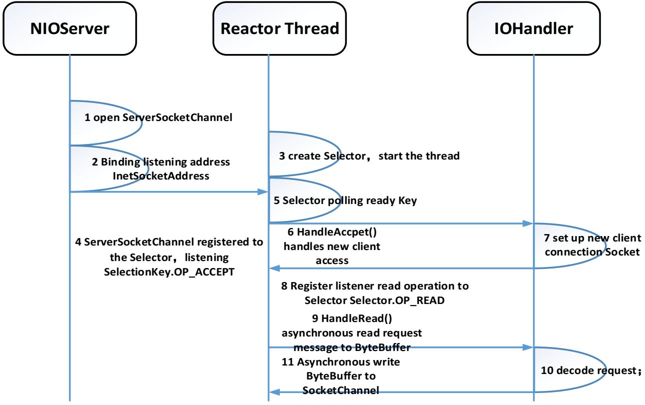

NIOServer
Reactor Thread
IOHandler
1 open ServerSocketChannel
2 Binding listening address InetSocketAddress
3 create Selector， start the thread
5 Selector polling ready Key
4 ServerSocketChannel registered to the Selector， listening SelectionKey.OP_ACCEPT
6 HandleAccpet() handles new client access
7 set up new client connection Socket
8 Register listener read operation to Selector Selector.OP_READ
9 HandleRead() asynchronous read request message to ByteBuffer
11 Asynchronous write ByteBuffer to SocketChannel
10 decode request；

Fig. 3. Server communication sequence

Netty IO threads NioEventLoop due to aggregation of the multiplexer selector can be concurrent treatment of hundreds or thousands of client channel. As the read and write operations are non blocking, which can fully improve the efficiency of the IO thread, to avoid the frequent IO blocking caused by the thread hanging. In addition, due to the Netty using asynchronous communication mode and an IO thread can concurrent processing N client connections for read or write operations, which fundamentally solves the traditional synchronous blocking IO connecting thread model, structure properties, elastic scalability and reliability have been greatly improved.

Network application layer protocol format is mainly the following: [7]

1) message length, packet size of fixed length, enough space complement, the transmitter and receiver follow the same conventions.

2) packet tail add special delimiter, such as each message end add CRLF character (for example FTP protocol) or designated special character as the message separator, the receiver through a special separate character segmentation message distinguish.

3) divide the message into a message header and a message body, the message header contains a field that represents the total length of the message (or message body length).

4) more complex custom application layer protocol.

This paper mainly in the third way to customize the application layer protocol as an example, to achieve the consistency of the detection system based on Netty framework. [8] Use Netty framework for the preparation of the general process:

1. Creates an object for the ServerBootstrap helper class of a NIO server;
2. To create a ServerBootstrap object, to pass the two NioEventLoopGroup thread pool, one called bossGroup, one called workGroup;
3. Call ServerBootstrap class childHandler (childHandler ChannelHandler) method, incoming specific business processing interface implementation class;
4. Using the created ServerBootstrap object to bind a port, the server starts listening.

Through the above operations, system program start, waiting for client connection requests, when the service receives the request data from the client, data from the data decoding module began the transmission and processing between service system modules.

4 Experimental verification

Using the open source server stress testing tool Jmeter, respectively to the traditional multi-threaded non-blocking implementation method, the implementation of the Java NIO and the implementation of Netty framework to carry out the pressure test of the system. From the server CPU occupancy rate, server I/O throughput, system response time 3 aspects for compare and analysis, through the comparison results show that the Netty framework of the system CPU occupancy rate is lowest.

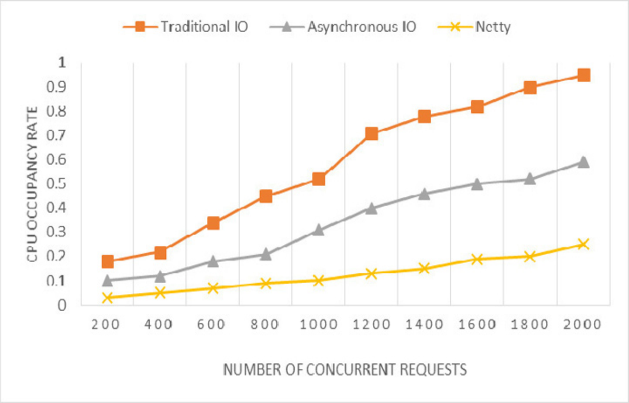


Fig. 4. CPU utilization percentage

In terms of throughput of the I/O system, the experimental results show that, every time each client requests to the server to send 100 bytes of data, not Netty framework for the realization of the consistency detecting system, when the number of clients in about 2000, the client and the server socket connection throw exception; in the Netty framework based on design consistency detecting system, server per second can request processing around 4000 concurrent users of the data. Experimental results show that when 20000 users connect to the server at the same time, the server can still be reliable and stable operation.

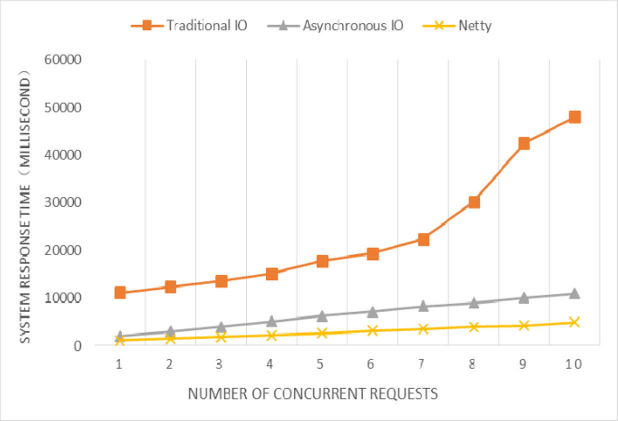


Fig. 5. System response time

5 Conclusion

The paper through the study of software consistency detecting mechanism, based on Netty network communication framework and consistency hash comparison, design and implement a high performance, support high concurrency and real-time monitoring software conformance testing system platform. The system can monitor the consistency of software distributed on several servers in real time, and detect whether the software is tampered with, so as to ensure the safety performance of the software. Compared with other consistency checking tools, the scheme proposed in this paper can not only support for multiple remote server software consistent condition monitoring, but also can give more rich and accurate information than that in fact a very effective distributed consistency detection scheme.

6 Acknowledgement

This work was supported by National Natural Science Foundation of China (No. 61272493, 61502536, U1536122).

References

[1] Huang Y. DIGITAL WATERMARKING TECHNIQUE USING HVS AND STATISTICAL CHARACTERISTIC OF IMAGE[J]. Computer Applications & Software, 2004.

[2] Tao S Q, Chang X R. Software Protection of Power System based on the Encryption Lock[J]. Computer Security, 2010.

[3] Hao Y H, Liu H B, Zheng L, et al. Software Piracy-proof Method Based on USB Encryption Lock[J]. Computer Engineering, 2010, 36(23):119-120.

[4] Huang X L, Zhou L, Li J. A Consistency Detection Method of Power Quality Data Interchange Format File[J]. Applied Mechanics & Materials, 2013, 411-414:1465-1469.

[5] Pimentel V，Nickerson B.G · Communicating and Displaying Real-Time Data with WebSocket.

[6] JBoss.Netty projec〔CP t /OL〕〔. 2012-4-1〕. http://netty.io/.

[7] Java Message Service[EB/OL]. http://java.sun.com/products/jms/.

[8] Jin Z G, Wei L I. Design and implementation of HTTP client based on Netty[J]. Telecom Engineering Technics & Standardization, 2014.

[9] Shang C W, Liu Q T, Zhao G, et al. Study on Mechanism of Consistency Detection in BPEL Activities Authorization Coordination Based on CPN[J]. Computer Science, 2014.

The generation of XSS attacks developing in the detect detection

Baojiang Cui[1,2], Yang Wei [1,2], Songling Shan[3] Jinxin Ma[4]
[1] School of Computer Science, Beijing University of Posts and Telecommunications, Beijing, China
[2] National Engineering Laboratory for Mobile Network Security, China
[3] China Electric Power Research Institute, Beijing, China
[4] China Information Technology Security Evaluation Center, China
Email: cuibj@bupt.edu.cn, fukayang@163.com, shansongling@epri.sgcc.com.cn, majinxin2003@126.com

Abstract. In recent years, the web security events emerge in endlessly, web security has been widely concerned. Cross-site scripting (XSS) attack is one of the most foremost threats which using malicious scripts injected into Web applications and executing the scripts in the client browsers. Moreover, attacker could also combine other means of attack with XSS vulnerabilities to do further attacks, which would lead to disclosure of user privacy and even property damage.

Common detect detection methods include black-box testing and white-box testing. Black-box testing scans faster while it can not locate the specific codes which cause the vulnerabilities. White-box audit tools can locate the specific codes while it spends lots of time to analyze all codes. We propose a novel approach to locate the vulnerabilities which combines Fuzzing test and dynamic taint analysis, and design system prototype, then verification and testing.

1 Introduction

With Web application security events happened frequently, domestic and foreign research organizations and institutions of Web application vulnerabilities do lots of researches which focus on the vulnerability detection and defensing exploits. There are two ways to detect detection which are Black box testing and the white-box testing [1].

White-box testing methods detect detection by analyzing the Web application's code, and due to the complexity of the program, it can not find all vulnerabilities. In this situation, it can run black-box testing which does not need to know the internal structure of the program. Black-box testing input s series of test data to the program, judging the existence of vulnerabilities by the response of the program [2].

Because white-box testing has a high rate of false positives and spends a long time to test, and because the weaknesses of black-box testing in coverage and process of development, it is best to combine these two techniques to detect Web security

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_33

vulnerabilities. Therefore, we propose a method to locate the vulnerabilities which combine Fuzzing test and dynamic taint analysis [3].

Our study focuses on the generation of attack vector library, because attack vector library is a key part of Fuzzing test, which is directly related to the accuracy and coverage of Fuzzing test. Based on the characteristics and attack patterns of XSS, we design XSS attack vector library. In Fuzzing test, using test string from the library to construct malformed URL or form, sending an HTTP request to Web application, judging the existence of vulnerabilities by the consistence of the response of server and vulnerabilities [4][5]. The method of generating test data determines the efficiency of testing process and the capacity of mining vulnerability.

The rest of this paper is organized as follows. In section 2, we introduce our Web application vulnerabilities positioning frames, and explain the whole process of our system. In section 3, some crucial technical method used to generate test data are given. Section 4 presents our empirical evaluation. Finally, we conclude and outline future work in Section 5.

2 Web Application Vulnerabilities Positioning Frames

Our technique consists of spider module, Fuzzing test module and dynamic taint analysis module. Wherein the dynamic taint analysis module comprises taint seed, taint tracker, and exploit analyzer.

Spider module analyzes the Web page content, gets XSS injection points' information, generates test files of sources, and prepares for Fuzzing test. Fuzzing test module determines Web application loopholes' injection points which are dynamic taint analysis's inputs. Dynamic taint analysis module marks taint seeds based on the results of Fuzzing test, traces taint propagations and records the propagation paths by taint propagation rules, ultimately confirms the existence of the vulnerability based on the detection algorithm, and if there are loopholes, it throws taint sources and taint propagation paths in the program.

Vulnerabilities derived from the test results Fuzzing Web application system is not completely accurate, there will be false positives. With the accuracy of dynamic taint analysis, we can verify the candidate vulnerabilities of the Fuzzing test reports, to get the report of real vulnerabilities in Web application. Therefore, the dynamic taint analysis's input will be the candidate vulnerabilities of Fuzzing test, and ultimately the number of vulnerabilities of security report will be the same or less.

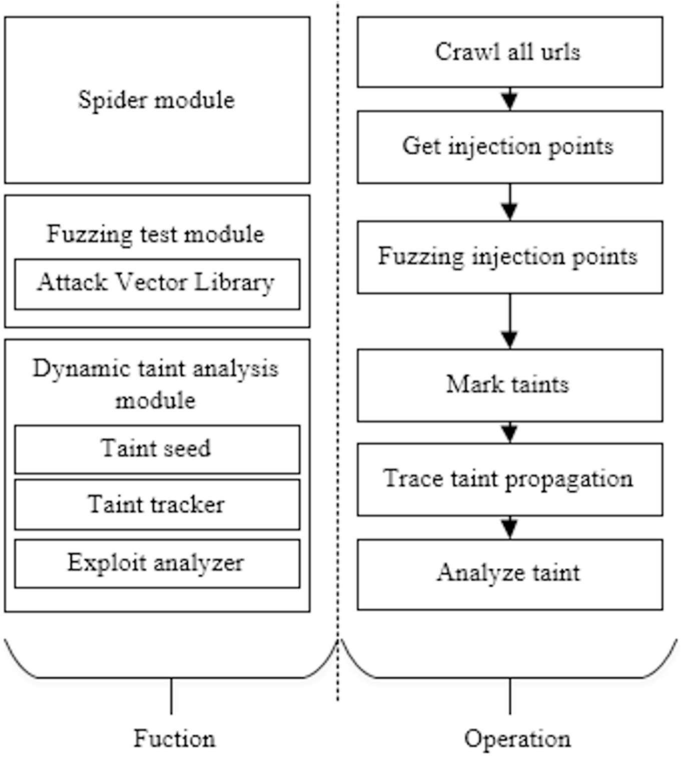


Fig. 1. Web Application Vulnerabilities Positioning Frames

3 Generating the Test Data

Our technique using web server to be the target of Fuzzing test, constantly sending test data to web server, judging the exceptional situation, and ultimately achieving the function that mining potential XSS vulnerabilities.

In the process of the Fuzzing test, the main work processes are repeated to generate data and monitoring software's exceptions. Because the Web server's Fuzzer is different from the other traditional network protocol tester when using test data, in the subsequent sections, we use attack vector instead of the test data in Fuzzing test. This paper presents a technique and an automated tool for finding XSS vulnerabilities in Web applications, and a variety of script codes are generating the attack vectors which we need.

According to the classifications of Fuzzers, this paper respectively adopts two ways to generate attack vector which are Pre-generated attack vectors and Agreement generated automatically attack vectors [6].

3.1 Pre-generated attack vectors

We should focus on interactive data where it is most prone to have XSS vulnerability. The fundamental reason of causing XSS is that there is not deal with input data well, and it generates JavaScript code which should not be executed and makes the browser successfully execute the code, so the input, URL parameters are suspicious of XSS injection point [7]. There are some common injection points as the following tips.

(1) URL. Extracting tags in the page which with parameters in the URL, then replace its parameters by XSS attack code, next constructs the URL that uses HTTP requests' ways. Like: *www.xss.com?Param URL = value*. We need to find the tags like < *a* >, < *iframe* >, < *img* >, etc.

(2) Form. Table will be submitted what user input to the Web server, and Submission methods are divided into the get and post, and it is commonly used post method. The types of input tag are *"text", "password", "submit"*, etc. We need to extract contents of its name and value through regular expression and add attack code in to the content, then re-send the request.

(3) HTML tag attributes. Most HTML tags support the pseudo-protocol form such as *javascript: [code]*. And the protocol represents the body of the attributes can be any JavaScript code that invokes its JavaScript interpreter to run. So users can use the feature to implant XSS attack code [8].

(4) HTML tag event. Interaction between JavaScript and HTML is achieved by the event, and the event that is a browser's or a user's own operation, like *onclick, onmouseover, onload, onerror*, etc. and the response function called an event handler, which executes the JavaScript code. So, since the event itself can execute JavaScript code, then the attacker also can use the feature to execute XSS attacks.

(5) SRCIPT tag. If SCRIPT tag involved in DOM, the parser will perform and modify the HTML page, which may lead to XSS vulnerabilities.

(6) The CSS cascading style form. The function of CSS is used to format the style of HTML tags, like color, position, size, etc. Some properties can be customized to specify, for example: background-image is usually used to change the address of the image, and attacker can also use this feature to insert XSS attack code [9].

In fact, lots of websites often use a variety of filtering mechanisms for processing the input data [10]. In order to ensure the effective operation of the injection of code, our attack vectors need to be deformed to escaping the filtering mechanism. Some common variants like encoding, mixed case, insert a blank string, etc. Using these methods to transform the initial attack vectors, is an important step in the pre-generated attack vectors, and this is an ongoing and continuous improvement process.

3.2 Agreement generated automatically attack vectors

That uses pre-generated attack vectors, through the analysis of the known loopholes, to identify the rules and its using patterns, thereby generating basic attack vector samples; and on this basis, that uses different derivative methods to transform samples; as a result, there is a massive attack vector set for detection. This approach can effectively detect known vulnerabilities, but it has low detection efficiency, and it can not find unknown vulnerabilities. Thus, this paper that uses the way which is Agreement generated automatically attack vectors to generate attack vectors at the same time.

Agreement generated automatically attack vectors requires previous study, understanding and interpretation of the protocol of the sites inputs' filtering mechanism. First according to the priori information to produce a set of data to interacts with web sites, then get the corresponding data filtering mechanism from the feedback information, and analyze and identify different filtering mechanisms of web servers, generate new attack vectors in a heuristic way. Agreement generated automatically attack vectors are divided into two sections which are protocol analysis and data generation.

3.2.1 Protocol analysis

In the production of a variety of attack vectors, there are some important characters has an important role for attack vectors. US-CERT considers that special characters are <>()" '&;{}% etc. If sites do not handle the special characters input, it will bring security problems, but only <>()" '&% of them will lead to XSS vulnerabilities, for the following reasons:

- Symbol < can introduce a new label, such as <img>, <script> and so on.
- Symbol > can end a HTML's label, such as injection scripts in <HL>, <tr> and so on.
- Symbols "and ' can end a tag's attribute, then added an event handler, which insert scripts to attack.
- Symbol % can introduce the form of a hexadecimal encoding of attack scripts.
- Symbol & may introduce javascipt code.

In addition to these special characters, there are some important keywords in the XSS attack. If we don't deal with these keywords will also course potential security problems. For example: script, javascript, document, cookie etc.

- Script and javascript can introduce script code.
- Document can lead script to operate the current document object.
- Cookie can lead script to access cookie information about the current page.

Because of the special role of the above special characters and keywords, common filtering mechanism deal with the special characters like characters encoded, substitutions, deletions and so on. So the website's inputs do not include these special characters. In order to detect whether these sites filtering mechanism is perfect, we need to experiment and test its filtering mechanism, and select the appropriate attack vector for its filtering mechanism or deformed attack vectors for further testing.

This paper uses the following test data to detect site's filtering mechanism.

Table 1. Analyze filtering mechanism's test data

Test data	Special characters
wouldbe < > wouldbe	< >
wouldbe ( ) wouldbe	( )
wouldbe ' " wouldbe	' "
wouldbe & wouldbe	' "
wouldbe % script javascript wouldbe	% script javascript
wouldbe document wouldbe	document
wouldbe Cookie wouldbe	Cookie

In order to facilitate searching input characters which we entered from site's feedback, we use *wouldbe* as the beginning and end tags.

When get the content of the feedback from the site, compare with the original data, thereby obtaining a filtering mechanism.

Currently filtering mechanisms are mainly deletion, encoding and transformation three ways, the following table is an instance of filtering characters.

Table 2. Filtering characters

Filtering method	Original data	Feedback data
Deletion	xDocument	x
Encoding	<	&lt
Transformation	iframe	yiframe

3.2.2 Data generation

Agreement generated automatically attack vectors is analyzing servers' different filtering mechanisms, generate new attack vectors in a heuristic way. This paper mainly searches for delete filtering mechanism, dynamic generated test data by attack vector sample.

Server that uses deletion to deal with key string is a common security measure. For deletion, we designed a context-free grammar to generate suitable attack vectors, and the following is its formal definition.

$$\text{Grammer } G=(\{O\},\{x, y, z, D\}, R, O), \text{ rule } R \text{ is } O\text{->}xyz|DO|OD \quad x\text{->}xD|Dx$$
$$y\text{->}yD|Dy \quad z\text{->}zD|Dz \tag{1}$$

In this grammar O is original attack vector, x y z is a part of O, D is the deleted key characters. Each member of the grammar is use to dynamically generate fuzzing attack vectors.

When analyze site's filter mechanism is deleting string which is document, get a part of attack vector document. Cookie as O, x is doc, y is cument.co, z is okie, then D is document. One of the generating string is docdocumentument.codocumentokie. It's parse tree shown in the figure 2.

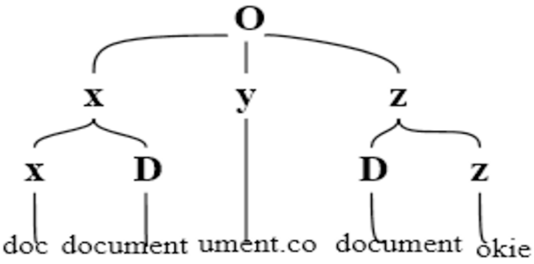


Fig. 2. Parse Tree

According to string which generated by this grammar, it can use the mechanism of deletion to test the server. The above sentence is filtered through the server, eventually converting to the original attack vectors which can be run.

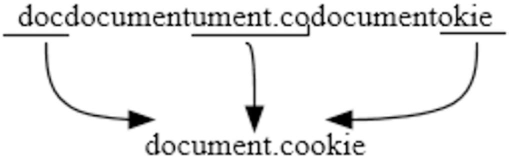


Fig. 3. Attack Vector's Transformation when It though Server

4 Evaluation

To demonstrate the supposed effectiveness of this method, we scan a site to find out its vulnerabilities, select two other mainstream XSS vulnerability detection tools to scan the same site, compare their scan results.

Cross Site "Scripter"(XSSer) is an automatic framework to detect, exploit and report XSS vulnerabilities in web-based applications [11]. It contains several options to try to bypass certain filters, and various special techniques of code injection. Burp Suite is an integrated platform for performing security testing of web applications. Its various tools work seamlessly together to support the entire testing process, from initial mapping and analysis of an application's attack surface, through to finding and exploiting security vulnerabilities [12].

As the results present in Table 3, our method totally detect 6998 urls, and find out 7 vulnerabilities. According to Table 3's data, our method is obviously faster than Burp Suite. We manually test the detected vulnerabilities, and found that XSSer gets a false alarm and two false negatives. Our technique and Burp Suite got no false positive and both got one false negative.

Table 3. Experimental Results

Tool	Time-consuming	The number of detected vulnerabilities	The actual number of vulnerabilities	The number of false positives	The numbe of false negative
Our technique	1035s	7	8	0	1
XSSer	998s	7		1	2
Burp Suite	1264s	7		0	1

5 Conclusion

In this paper, we demonstrate a novel approach to detect XSS vulnerabilities combine Fuzzing with dynamic taint analysis technique. In Fuzzing test, that uses attack Strings from the attack vectors' library to do Fuzzing test, that can get XSS vulnerabilities' injection points quickly. In dynamic taint analysis, by tracking these injection points' propagations, recording the whole process of XSS vulnerabilities in the program from source to sink, thus completing the vulnerability detection. This paper describes the generation of attack vectors for the vulnerability, in order to solve the problems about detect unknown vulnerabilities, that uses pre-generated and agreement generated automatically attack vectors to generate test data and designing a context-free dynamically generated test data grammar based on server deletion filtering mechanism.

Experimental results show that the method can effectively mining sites' vulnerabilities, and it also can effectively reduce the workload of the site safety audit. In our future work, we will analyze the sites' filtering mechanism, and improve the efficiency of generating attack vector algorithm.

Acknowledgments. This work was supported by National Natural Science Foundation of China (No. 61272493 & No.61502536 & No. U1536122).

References

1. Mariani L, Pezze M, Riganelli O, et al. Autoblacktest: Automatic black-box testing of interactive applications[C]//Software Testing, Verification and Validation (ICST), 2012 IEEE Fifth International Conference on. IEEE, 2012: 81-90.
2. Ganesh V, Leek T, Rinard M. Taint-based directed whitebox fuzzing[C]//Proceedings of the 31st International Conference on Software Engineering. IEEE Computer Society, 2009: 474-484.
3. Pai G J. A survey of software reliability models[J]. arXiv preprint arXiv:1304.4539, 2013.
4. Emmi M, Majumdar R, Sen K. Dynamic test input generation for database applications[C]//Proceedings of the 2007 international symposium on Software testing and analysis. ACM, 2007: 151-162.

5. Fan J, Gao P, Shi C C, et al. Research on combine White-box testing and Black-box testing of Web Applications security[C]//Advanced Materials Research. Trans Tech Publications, 2014, 989: 4542-4546.

6. Duchene F, Groz R, Rawat S, et al. XSS vulnerability detection using model inference assisted evolutionary fuzzing[C]//SECTEST 2012-3rd International Workshop on Security Testing (affiliated with ICST). IEEE Computer Society, 2012: 815-817.

7. Martin M, Lam M S. Automatic generation of XSS and SQL injection attacks with goal-directed model checking[C]//Proceedings of the 17th conference on Security symposium. USENIX Association, 2008: 31-43.

8. Wassermann G, Su Z. Static detection of cross-site scripting vulnerabilities[C]//Software Engineering, 2008. ICSE'08. ACM/IEEE 30th International Conference on. IEEE, 2008: 171-180.

9. Wassermann G, Yu D, Chander A, et al. Dynamic test input generation for web applications[C]//Proceedings of the 2008 international symposium on Software testing and analysis. ACM, 2008: 249-260.

10. Hansen R. XSS (cross site scripting) cheat sheet esp: for filter evasion[J]. 2010-01-01]. http://ha. ckers. org/xss. html, 2010.

11. An automatically detect XSS vulnerabilities in web-based applications named XSSer[EB/OL]. https://xsser.03c8.net/.

12. Burp Suite Walkthrough[EB/OL]. http://resources.infosecinstitute.com/burp-suite-walkthrough/.

Searchable Public Key Encryption Scheme with Data Integrity Checking

JunYang[1], Shujuan Li[2]

[1] School of Computer Science, Beijing University of Posts and Telecommunications,
National Engineering Laboratory for Mobile Network Security, China
junyang@bupt.edu.cn

[2] School of Computer Science, Beijing University of Posts and Telecommunications,
National Engineering Laboratory for Mobile Network Security, China
lishujuan011@163.com

Abstract. Searchable public key encryption allows data owners to store the encrypted data in the cloud server and the users can directly search over the encrypted data in the cloud server to obtain the desired data. However, in such situation, for example, in the process of network transmission, the data can be damaged due to the incorrect operation caused by the data owner or other unexpected circumstances, leading to serious consequences that the transmitted data is not consistent with the original data. Thus, the searchable public key encryption scheme with data integrity checking is necessary. In this paper, we implement the data integrity checking and further improve the security and integrity of the scheme. We first propose the concrete scheme and then analyze the security and the performance of the scheme.

1 Introduction

With the development of cloud technology, users can use the services provided by the cloud server more conveniently. Through these existing tools in the cloud server, such as Dropbox, Google Drive and iCloud, users can not only easily store their data in the cloud server, but also can share their data with other users quickly.

Although there are many advantages storing the data in the cloud server, there are still many problems for the development of cloud computing. The problem of data security [1] is the first problem that we should consider, and even to a certain extent, it has become an important factor that restricts its development. To solve the problem mentioned above, the direct idea is to store the data after being encrypted in the cloud server, so that we can ensure the privacy of the plaintext. This, however, raises another question; when the client needs to find the relevant documents containing a keyword, searching the encrypted data in the cloud server becomes a problem. In order to solve this problem, the general method is to use the searchable encryption scheme [3]. In this scheme, the data owner uploads the encrypted documents and the encrypted keywords to the cloud server, and the user submits the trapdoor to the cloud server to search over the encrypted documents.

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_34

In the searchable encryption [2] scheme, data can be stored in the cloud server in the form of the ciphertext, and the user can directly search over the encrypted data [5] instead of decrypting the encrypted data at first. With this method, the cloud manager and the malicious users cannot get any information of the plaintext from the encrypted document and the encrypted keyword, and thus this scheme guarantees the privacy and the security of the data. However, in the cloud environment, due to the hardware, software or human factors, the cloud server sometimes incorrectly modified, or even deleted the users' data. And the users themselves do not store data locally, so users cannot check directly the integrity of the received data. In such a situation, the design and implementation of the searchable public key encryption scheme with data integrity checking becomes necessary. The user check the integrity [7] of the received data to determine whether the received data is consistent with the original data the data owner possesses.

For the above questions, in this paper, on the basis of the original structure of the searchable public key encryption scheme, we realize the function of data integrity checking [4]. We first propose the concrete scheme; and then we analyze the security and the performance of the scheme.

The remainder of the paper is organized as follows. Section 2 presents some related work. And we propose the concrete scheme in Section 3 and analyze the security and the performance of the scheme. Finally, we conclude the paper in Section 4.

2 Related Work

In this section, before we introduce our scheme, we first review the related knowledge used in this paper, including the foundation of mathematics and cryptology concepts.

2.1 Complexity Assumption

2.1.1 Bilinear Map

We assume that G_1 and G_2 are two groups of prime order p and g is a generator of G_1. A bilinear map with cryptography is a map e: $G_1 * G_1 \rightarrow G_2$ with the following properties:

Bilinearity: $\forall$ g, h $\in G_1$, $\forall$ a, b $\in Z^*_p$, $e(g^a, h^b) = e(g, h)^{ab}$.

Non-Degeneracy: $\forall$ g $\in G_1$, $g \neq 0 \Longrightarrow (e(g, g)) = G_2$(e(g, g) generates G_2), in other words: $g \neq 0 \Longrightarrow e(g, g) \neq 1$

Computability: there is an effective algorithm for computing e(g, h) $\in G_2$ for all the g, h $\in G_1$.

2.1.2 Bilinear Diffie-Hellman

We assume that G_1 and G_2 are two groups of prime order p and g is a generator of G_1.

Bilinear Diffie-Hellman Problem(BDH): Fix a generator g of G_1. The BDH problem is as follows: given g, g^a, g^b ∈ G_1 as input, compute $e(g, g)^{ab}$ ∈ G_2. We say that BDH is intractable if all polynomial time algorithms have a negligible advantage in solving BDH.

An algorithm A has advantage ε in solving BDH in G if $\Pr[A(G, p, g, g^a, g^b, g^c) = 1] - \Pr[A(G, p, g, g^a, g^b, g^{ab}) = 1] \geq \varepsilon$, where the probability is over the random choice of generator g in G, and the random choice of a, b, c in Z_p^*.

Definition 1: The BDH assumption holds in G if no algorithm has advantage more than ε in solving the BDH problem in G.

2.2 Searchable Encryption

As shown in Figure 1, a searchable encryption scheme [6] contains four processes:

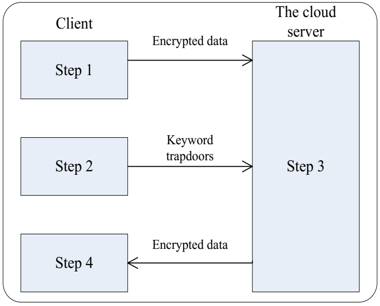


Figure 1: Steps in searchable encryption

- Step1: Encryption process. The data owners use the key to encrypt the document on the local devices, and upload it to the cloud server.
- Step2: Trapdoor generation process. The authorized clients use the key to generate the trapdoor for a keyword, and the trapdoor cannot reveal any information of the plain text.
- Step3: Retrieval process. The cloud server takes as input trapdoor to search [10] over the encrypted data, to return all of the documents containing the specified keyword. The cloud server cannot get more information except for the knowledge if the cipher text contains a specific keyword.
- Step 4: Decryption process. The clients use the key to decrypt the encrypted data returned by the cloud server to obtain the query results.

3 The Proposed Scheme

3.1 Description of the Scheme

We propose a concrete scheme that contains the following algorithms. And we use the graphic to describe the algorithm in a simple manner, as shown in Figure 2.

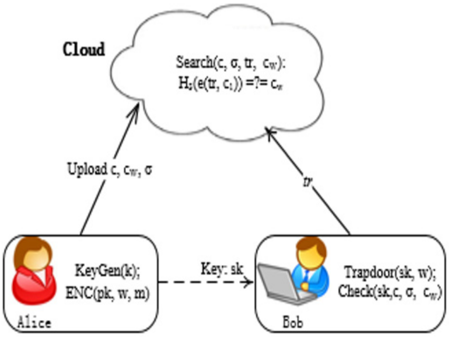


Figure 2: The concrete searchable public key encryption scheme with data checking

1) KeyGen(k): the data owner uses this algorithm to generate the public/private key pair pk/sk. The specific process is as follows:

- Generate a bilinear map group system B= (G1, G2, q, g, e(·, ·)), where q is the order of G_1 and G_2, g is the generator of G_1;
- Select hash functions: H_1: G_1 → $\{0, 1\}^{11}$, H_2: $\{0, 1\}^*$ → G_1; H_3: G_2 → $\{0, 1\}^{13}$; H_4: $\{0, 1\}^*$ → $\{0, 1\}^{14}$.
- Pick a random x ∈ Z_q, and computes $y = g^x$;

Finally, the algorithm outputs: Pk = (q, g, e, G1, Ge, y), sk = (pk, x).

2) ENC(pk, w, m): the data owner uses this algorithm to encrypt data and generate its keyword ciphertexts and the tag σ The algorithm takes as input the public key pk, and:

- randomly picks a r ∈ Z_q, and computes $c_1 = g^r$, $\kappa= y^r$, $K = H_1(\kappa)$;
- for the plaintext m generates its ciphertext $c_2 = K \oplus m$, and generates the encrypted keyword c_w for a keyword w by computing: $h = H_2(w)$, $\mu = e(h, y)^r$, $c_w = H_3(\mu)$.
- for the plaintext generates its tag used to check the integrity of this file by computing: $\alpha= H_4(m, c_2, c_w)$, $\sigma= \alpha \oplus K$.

Finally, the data owner uploads (c_1, c_2, c_w, σ) to the cloud server.

3) Trapdoor(sk, w): the client uses this algorithm to generate a trapdoor for a keyword to perform a keyword search. It takes as input the private key sk and a keyword w and outputs the trapdoor tr for the keyword by computing tr = $H_2(w)^x$. Finally, the client uploads tr to the cloud server.
4) Search(tr, c_1, c_2, c_w, σ): the cloud server uses this algorithm to perform a keyword search over the encrypted data to justify whether the encrypted document contains the specific keyword. This algorithm takes as input the trapdoor tr, the encrypted data c_2, the encrypted keyword c_2 and the tag σand then it judges $H_3(e(tr, c_1))$ =?= c_w. If it is true, the cloud server would send (c_1, c_2, c_w, σ) to the client; otherwise, the cloud server would output "reject".
5) Check(sk, c_1, c_2, c_w, σ): the client uses this algorithm to check [8] whether the received data is consistent with the original data provided by the data owner. This algorithm takes as input the private key sk, the encrypted data c_2, the encrypted keyword c_w and the tag σand

- generates the plaintext m by computing: $\kappa= c_1^x$, $K = H_1(\kappa)$, $m = K \oplus c_2$;

- judges $H_4(m, c_2, c_w)$ =?= $\sigma \oplus$ K. If it is true, the received data is consistent with the original data and the client can accept the plaintext m. Otherwise, this algorithm simply outputs “false”.

3.2 Security Analysis

In order to analyze the security of the scheme, we assume that the cloud server is “curious and honest”, and we assume that the users try to access data within or out of the scopes of their privileges. Moreover, we assume that the communication channel is not safe, that is, an attacker can obtain information from the communication channel.

Theorem 1: The proposed scheme realizes the function of controlled searching.

Proof: The theorem requires that users who have the searchable encryption key sk can search over the specific document, but the users are unable to search over documents out of the scopes of their privileges. Similarly, the user cannot generate other private keys to search for other files based on the known private key. Theorem 1 can deduced from the following lemmas:

Lemma 1: Every authorized user can perform a keyword search successfully.

Proof: Lemma 1 shows the correctness of the proposed scheme. After the data owner has generated the searchable encryption key sk, the key should sent to the authorized user. The authorized user uses the private key sk to generate the trapdoor, and the trapdoor would be submitted to the cloud server. The cloud server would run the search algorithm after receiving the trapdoor. We can see that the authorized user can search successfully based on the known key sk by computing $H_3(e(tr, c_1))$ =?= c_w.

Lemma 2: When a user tries to search for a document out of the scopes of his privilege, or when the cloud server colludes with the user, they are unable to search for documents that they are allowed to search.

Proof: In this case, the malicious user has not only his own private key but also information on the cloud server: trapdoor tr, the ciphertext c_2, the keyword c_w and the tag σ However, the user cannot search for documents out of the scope of his or her privileges. The malicious user has an arbitrary amount of information: $H_2(w)^x$, g, g^x and $(g^x)^r$. According to the discrete logarithm problem, the user cannot guess the value of r and x. A malicious user generates the error sk, and computes the value tr which would be uploaded to the cloud server. The cloud server runs the search algorithm. We can see that there is no c_w to make the formula $H_3(e(tr, c_1)) == c_w$ compute successfully, and the search algorithm would return a “reject” value.

Theorem 2: The scheme realizes the function of query privacy.

Proof: The malicious user knows the private key sk corresponding to the specific document and can search for the documents within the scope of his key. The cloud server also knows the information on the data stored in it: the keyword ciphertext c_w, the trapdoor tr and the pk. However, we can still prove that our scheme achieves the function of query privacy. Theorem 2 can be deduced from the following lemma.

Lemma 3: An attacker cannot obtain the information of the keyword according to the known trapdoor.

Proof: We assume that an attacker A has an arbitrary number of trapdoors, and the attacker tries to obtain the information on the keyword based on this. In this case, the attacker has the information: $H_2(w_1)^x$, $H_2(w_2)^x$, …, $H_2(w_n)^x$. The attacker can obtain the information of the keyword only when he can guess the private key x. According to the information that the attacker has known: g and g^x, we can know that the attacker cannot get the value of x based on the Lemma 2.

Lemma 4: An attacker cannot get the information on the keyword when the attacker knows the keyword ciphertext and other related public information.

Proof: The attacker attempts to obtain the information on the keyword from the keyword ciphertext stored in the cloud server. The attacker knows the following information: the keyword ciphertext $c_w = H_3(e(H_2(w), g^x)^r)$ and the public key pk = (g, g^x, e, G_1, G_2). According to the BDH problem stated in the section 2.1.3, the attacker cannot know the sk x based on the ciphertext c_w and the public key pk. And thus the attacker cannot get the information of the keyword.

3.3 Efficiency

According to the searchable public key encryption scheme [9] with data integrity checking described above, we can see that the algorithms in the scheme mainly involve the pairing computation and the operation of exclusive or. The scheme we proposed is based on the searchable public key encryption scheme which is based on the bilinear pairing. And on the basis of not destroying the original asymmetric nature, we mainly use the operation of the exclusive or to check the integrity of the data. We then discuss the performance of the two operations separately.

In our experiment, we use the language java to implement the operation of the exclusive or. The length of data is 80, and the data type used in byte. We tested the operation on a Lenovo computer with an Intel(R) Core(TM) i5-4590 CPU at 3.30GHZ and 8 GB of RAM running Windows7. We test that when the number of data increases to 1million, the time cost is 27ms as shown in the Figure 3. For the pairing computation, according to the pairing-based cryptography library, the time cost of a pairing computation using the pretreatment for the parings of type ‘a’ is 11ms. According to the java pairing-based cryptography library, the time cost is 7.234ms. Thus, from the point of view of the performance evaluation, the cost time of the operation of the exclusive or is almost negligible compared with the cost time of the pairing computation. So efficiency of the proposed scheme is almost same as the scheme that our scheme is based on.

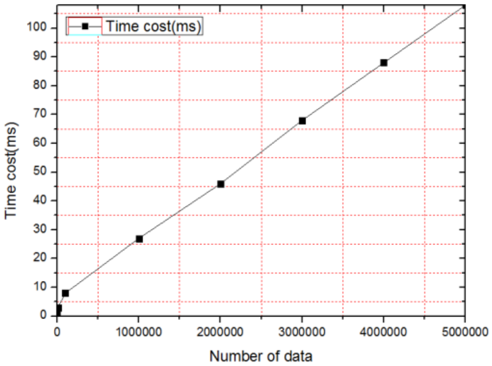


Figure 3: Time cost of the operation of the exclusive or

4 Conclusion

Research on searchable encryption is becoming popular, and the technology guarantees the security of the owners' data in the search process. Based on this concept, for the first time, we construct the searchable public key encryption scheme model with data integrity checking and realize a concrete scheme on the basis of ensuring security. Moreover, we analyzed the performance and security of the concrete scheme in detail. The analysis results show that our work can provide an effective and secure solution to searching data stored in the cloud based on the searchable asymmetric encryption scheme.

In the concrete scheme we constructed, the data owner uploads the encrypted keyword, the ciphertext and the tag $\sigma$to the cloud server. And if the user searches successfully, the cloud server sends the ciphertext and the tag $\sigma$to the user. The user can use the tag $\sigma$to check if the received data is consistent with the original data. Thus the security of the scheme has been further improved. However, in the scheme proposed in this paper, a different key is used for a different document, and if the data owner possesses a large number of documents, a large number of keys are needed, and the data owner needs to send the keys to the authorized users. Moreover, a large number of trapdoors must be generated by users and submitted to the cloud in order to perform a keyword search over many documents. How to reduce the number of encryption keys [11] and the number of trapdoors is the follow-up work we have to consider. Moreover, this scheme provides us a direction to solve the problem of searchable asymmetric encryption, and we can continue to study more problems to be solved in this field.

Acknowledgments. This work is supported by the National Natural Science Foundation of China (No. 61272493，61502536，U1536122).

References

1. Feng DG, Zhang M, Zhang Y, Xu Z. Study on cloud computing security. Ruan Jian Xue Bao/Journal of Software, 2011, 22(1): 71−83 (in Chinese with English abstract)
2. Li JW, Jia CF, Liu ZL, Li J, Li M. Survey on the searchable encryption. Ruan Jian Xue Bao/Journal of Software, 2015, 26(1):109-128 (in Chinese).
3. Shen ZR, Xue W, Shu JW. Survey on the research and development of searchable encryption schemes. Ruan Jian Xue Bao/Journal of Software, 2014, 25(4):880−895 (in Chinese)
4. Ateniese G, Burns R C, Curtmola R, et al. Remote data checking using provable data possession[J]. ACM Transactions on Information and System Security (TISSEC), 2011, 14(1): 12
5. Song D, Wagner D, Perrig A. Practical techniques for searches on encrypted data. In: Proc. of the 2000 IEEE Symp. on Security and Privacy. Berkeley: IEEE Computer Society, 2000. 44−55
6. Shen ZR, Xue W, Shu JW. Survey on the research and development of searchable encryption schemes. Ruan Jian Xue Bao/Journal of Software, 2014, 25(4):880−895 (in Chinese).
7. Chang E C, Xu J. Remote Integrity Check with Dishonest Storage Server[C]//Computer Security-ESORICS 2008. Springer Berlin Heidelberg, 2008: 223-237.

8. Hao Z, Zhong S, Yu N. A Privacy - Preserving Remote Data Integrity Checking Protocol with Data Dynamics and Public Verifiability [J]. Knowledge and Data Engineering, IEEE Transactions on, 2011, 23(9): 1432-1437
9. Boneh D, Crescenzo G, Ostrovsky R, Persiano G. Public key encryption with keyword search. In: Proc. of the EUROCRYPT. Berlin, Heidelberg: Springer-Verlag, 2004. 506−522
10. Baek J, Safavi-Naini R, Susilo W. Public key encryption with keyword search revisited. In: Proc. of the Int'l Conf. on Computational Science and Its Applications. Berlin, Heidelberg: Springer-Verlag, 2008. 1249−1259
11. Cui, B.; Liu, Z.; Wang, L., "Key-Aggregate Searchable Encryption (KASE) for Group Data Sharing via Cloud Storage," in Computers, IEEE Transactions on, vol. PP, no.99, pp.1-1

An Attack Detection System for Multiple Web Applications Based on Big Data Platform

Xiaohui Jin[1,2], Congxian Yin[1,2], Pengpeng Yang[1,2], Baojiang Cui[1,2]
[1] School of Computer Science and Technology,
Beijing University of Posts and Telecommunications, Beijing, China
[2] National Engineering Laboratory for Mobile Network Security
Emails: jinxiaohui@bupt.edu.cn, yincongxian@foxmail.com,
yp1129_@bupt.edu.cn, cuibj@bupt.edu.cn

Abstract. Considering the protection requirements of large organizations for multiple web applications, we design and implement an attack detection system. The system is built on the big data platform, which is highly scalable. It adopts the network-traffic-based detection, capturing, parsing and analyzing the HTTP packets passing by in real time. By analyzing historical data, we are able to get application-specific access patterns, which can help domain experts find out anomalies efficiently. Besides, based on the labels given by domain experts, semi-supervised learning is applied to build attack detection classifier. The system is deployed in the real network of our university and has detected dozens of attacks.

1 Introduction

Large organizations, such as enterprises, colleges and governments, usually have to maintain multiple web applications. The traditional firewall and intrusion detection systems work at network layer, while web application firewall (WAF) just provides rule-based filtering, which cannot fulfill the security protection requirements of various web applications.

The development of big data technology makes it possible to gain massive storage and powerful computing ability at a relatively low cost. It is regarded as an ideal solution for coping with unknown attacks. By combining big data technology and machine learning algorithms in the field of intrusion detection, the protection ability of web applications can be effectively improved.

We design and implement an attack detection system based on big data platform, which can provide attack detection service for multiple web applications. The system is consisted of six subsystems, which are data collection, protocol parsing, big data storage, offline learning, online detection and visualization respectively. The system is deployed in the real network of our university——Beijing University of Posts and Communications (BUPT), providing attack detection service for multiple web applications maintained by the university. The statistic data shows that the system works stably and is able to detect dozens of attacks targeting at web applications.

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_35

The following chapters are organized as follows: Chapter 2 introduces the related work, mainly focusing on the application of machine learning in the field of intrusion detection. Chapter 3 describes the system's overall architecture, and explains the functionality and implementation of each subsystem in detail. The conclusion and future work is in Chapter 4.

2 Related Work

The system described in this paper belongs to the category of intrusion detection system. As an important member of the security appliances, intrusion detection system [1] has a long history. It protects computer systems and networks from abuse, and is another generation of security technology following up the firewall. Intrusion detection is usually regarded as a classification issue since its main objective is to distinguish malicious access behavior. Machine learning is an effective way to solve such problems, and has been applied in the field of intrusion detection for a long period [2]. Supervised and unsupervised learning are two commonly used methods. Supervised learning is introduced in [3][4][5], and maximum entropy, support vector machine (SVM) and random forest algorithms are adopted respectively. In general, supervised learning requires a great deal of labeled training samples. Some researchers label the training sets manually, while a large number of other researchers tend to use some publicly available labeled sets, such as the famous KDD CUP '99 data set. An unsupervised learning method is proposed in [6], while [7] makes attempt to combine the two learning methods. Overall, unsupervised learning is easily affected by the data distribution, whose effect is inferior to supervised learning.

With the development of the Internet, it is now quite easy to collect a large number of unlabeled samples. However, the number of labeled samples is relatively small, and the cost of manual annotation is very expensive. For those reasons, many researchers are turning to semi-supervised learning and active learning. The whole iterative process of semi-supervised learning needs no manual intervention. It is based on a small number of labeled samples and tries to make use of unlabeled data [8][9]. On the other hand, the general idea of active learning is to imitate the learning process of human. It extracts a small number of most uncertain samples, and asks domain experts for a correct label. Active learning classifier is established based on these samples, and is used to classify other unlabeled samples [10]. The system designed in this paper adopts semi-supervised learning. Meanwhile, we also draw lessons from the idea of active learning, and ask domain experts to judge normal access patterns and abnormal access behavior right before the semi-supervised learning starts.

From another point of view, the changing focus of machine learning research reflects the fact that the amount of available data is increasing dramatically. Big data era comes. It can provide adequate fuel for machine learning, so the combination of big data technology and machine learning can effectively improve the ability of existing intrusion detection system. Therefore, we argue that it is quite necessary to design and implement an attack detection system for a variety of web applications based on big data platform.

3 The Proposed System

By utilizing big data and machine learning technologies, we design and implement a system which can provide attack detection service for many kinds of web applications simultaneously. The system's overall architecture is shown in Figure 1. We will introduce the six subsystems respectively.

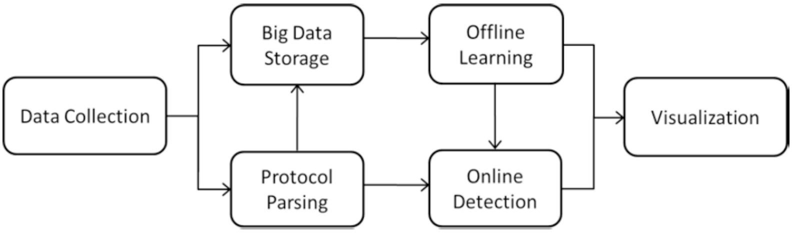


Fig. 1. The system's overall architecture

3.1 Data Collection Subsystem

The data collection subsystem adopts the client/server architecture. A high-performance distributed message queues named Kafka [11] is used in server side. Kafka is intended to cache text log messages originally. However, we make some extension in our system and use it to cache the high-speed network traffic.

The client monitors network traffic on the bypath. It consists of high-performance network traffic capture procedures and the producer client of Kafka. In practice, the server side is usually deployed close to the data source. Because of the client does not do any parsing, and its memory usage is optimized in the code level, the system can collect data at the rate of several gigabit per second.

3.2 Protocol Parsing Subsystem

The protocol parsing subsystem is used to transform the captured network traffic data to an object type, so that the following subsystem can read and process it. We define the HTTPSession class, its members cover all information of a HTTP session, including source IP, source port, destination IP, destination port, protocol type, as well as the main fields in request headers, request body, the main fields in response header, response body and response time. The object of HTTPSession class is the basic processing unit in the whole system.

The protocol parsing subsystem is based on the Spark Streaming [12]. First of all, it reads traffic packets cached in data collection subsystem using Kafka consumer client written in Scala, and uses the 5-Tuple (source IP, source port, destination IP, destination port, protocol type) as the key, packing and restoring the complete HTTP session data with join and reduce operations. Then HTTP protocol analysis module is called by the map operation to map network traffic data to HTTPSession object. To improve performance, we implement HTTP protocol parsing module in C language, and use JNI to interact with Scala.

3.3 Big Data Storage Subsystem

The big data storage subsystem contains two storage platforms——HDFS and HBase.

HDFS is the open source implementation of Google File System (GFS) [13], and is suitable for storing large volume of unstructured data. HDFS is used for storing the original traffic data. The specific implementation is using Spark Streaming tasks to call Kafka consumer client and save the received data to HDFS.

Based on the design principle of Google Bigtable [14], HBase is a column-oriented distributed database. The subsystem uses HBase to storage HTTPSession object sent from protocol parsing subsystem. We build index for key fields, so that the offline learning subsystem can perform effective queries.

3.4 Offline Learning Subsystem

Offline learning subsystem is a machine learning platform based on the Spark MLlib [15], as is shown in Figure 2. By analyzing the historical web access data, we discover following rules:

(1) Normal access takes more than 90% of total access, and they are the high similar to each other.

(2) The access to specific domain and URI shows certain statistical characteristics in several dimensions, including depth, width, number of visits, duration, response status code and response time. The object that deviates from statistical center is more likely to be an attack.

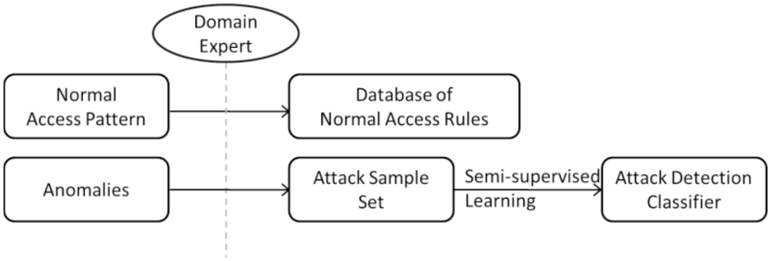


Fig. 2. The Offline Learning subsystem

For (1), we extract the URI and parameters from URL, and generalize it to some normal access patterns. Domain experts make judgment to the extracted normal access patterns to form the database of normal access rules.

The subsystem discovers anomalies according to (2), and submits the HTTPSession object to domain experts for further judgment. The objects marked as attack by domain experts will be added into the attack sample set. The subsystem extracts features from attack sample set and use semi-supervised learning algorithm to train intrusion detection classifier.

3.5 Online Detection Subsystem

The online detection subsystem uses misuse-based and anomaly-based detection methods jointly. Figure 3 shows the work flow.

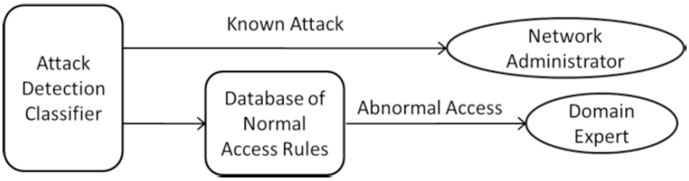


Fig. 3. The Online detection subsystem

Newly-received HTTPSession objects first go through the attack detection classifier. The subsystem informs network administrators if any known attack is detected. Other normal HTTPSession objects will be passed to the database of normal access rules. If an object fails matching the rules, the object will be treated as an abnormal access and submitted to domain experts for further judgment. If domain experts mark it as an attack, then the object will be added to the attack sample set.

3.6 Visualization Subsystem

The main purpose of visualization subsystem is to provide a convenient operation interface to network administrators and domain experts. We will neglect the details here.

4 Conclusion and Future Work

Generally speaking, the system mainly has three advantages. First of all, being built on big data platform, the system can effectively parse web application protocols and extract access patterns leveraging massive storage and powerful processing ability. Secondly, the results of big data analysis make the labeling work of domain experts more specific and efficient. Besides, semi-supervised learning method is very suitable for such application scenarios. Finally, the system uses the misuse-based and anomaly-based detection jointly, which reduces false alarms and improves the recall rate at the same time.

Our statistic data shows that the system processes about 1.8 million web requests per day, and the proportion of detected web attacks in all requests is about 1%. SQL injection, directory traversal and cross site scripting (XSS) are the most common attacks.

Moreover, a lot of work can be done to improve the system. The future work will be carried out in the following two aspects. First, web protocol should be parsed in depth. Currently, the offline learning subsystem can only extract simple request parameters. The parameters coded in JSON or XML format cannot be parsed yet. Besides, IPv6 and HTTPS support should also be involved. Second, the feature selection strategy needs to be optimized. In order to obtain a better performance of classifiers, it is indispensable to do in-depth research jobs on the theoretical knowledge involved in this field.

References

1. J Mchugh，A Christie，J Allen. Defending Yourself: The Role of Intrusion Detection Systems. IEEE Software, 2000, 17(5):42-51
2. D Barbará，J Couto，S Jajodia，L Popyack，N Wu. ADAM: Detecting Intrusions by Data Mining. Proceedings of the IEEE Workshop on Information Security, 2001:11--16
3. Y Gu, A Mccallum, D Towsley. Detecting anomalies in network traffic using maximum entropy estimation. ACM Sigcomm Conference on Internet Measurement, 2005:345-350
4. J Yu, H Lee, MS Kim, D Park. Traffic flooding attack detection with SNMP MIB using SVM. Computer Communications, 2008, 31(17):4212-4219
5. J Zhang, M Zulkernine. A Hybrid Network Intrusion Detection Technique Using Random Forests. International Conference on Availability, 2006, 37(8):262-269
6. G Jia, G Cheng, DM Gangahar，DK Agrawal. Traffic anomaly detection using k-means clustering. In. GI/ITG workshop MMBnet
7. SR Gaddam, VV Phoha, KS Balagani. K-Means+ID3: A Novel Method for Supervised Anomaly Detection by Cascading K-Means Clustering and ID3 Decision Tree Learning Methods. IEEE Transactions on Knowledge & Data Engineering, 2007, 19(3):345-354
8. X. Zhu. Semi-supervised learning literature survey. Technical Report 1530, Department of Computer Sciences, University of Wisconsin at Madison, Madison, WI, Apr. 2006.
9. O. Chapelle, B. Schölkopf, A. Zien, eds. Semi-Supervised Learning, Cambridge, MA: MIT Press, 2006
10. M Almgren, E Jonsson. Using active learning in intrusion detection. Proceedings of the 17th IEEE Computer Security Foundations Workshop (CSFW’04)
11. J Kreps, L Corp, N Narkhede, J Rao, L Corp: Kafka: a distributed messaging system for log processing. NetDB'11, Athens, 2011
12. http://spark.apache.org/streaming/
13. S Ghemawat: The Google file system. ACM SIGOPS Operating Systems Review, 2003, 37(5):29-43
14. F Chang, J Dean, S Ghemawat, WC Hsieh, DA Wallach: Bigtable:a distributed storage system for structured data. ACM Transactions on Computer Systems, 2008, 26(2):205--218
15. http://spark.apache.org/mllib/

Cognitive Countermeasures against BAD USB

Yeunsu Lee[1], Hyeji Lee[1], Kyungroul Lee[2], Kangbin Yim[1]
[1] Dept. of Information Security Engineering, Soonchunhyang University
Asan, South Korea
{yunsu07, gpwl899, yim}@sch.ac.kr
[2] R&BD Center for Security and Safety Industries (SSI), Soonchunhyang University
Asan, South Korea
carpedm@sch.ac.kr

Abstract. Recently, a novel attack technique called BAD USB emerged. This attack injects and executes malicious codes in the firmware that is stored in USB controllers. A serious problem regarding BAD USB, which also manipulates the firmware maliciously, is that the existing anti-virus programs cannot detect it, so the seriousness of this kind of attack is increasing. To solve this problem several countermeasures have been researched, but these are not effective enough. Therefore, in this paper, we propose a way to verify the integrity of the driver or the firmware that is installed by BAD USB proposed. Through the use of this method, solutions for the prevention of the malicious BAD USB behaviors can be formulated.

1 Introduction

One of the interfaces for the connection of the peripheral devices of a computer is the USB (Universal Serial Bus) interface, and the corresponding advantages are faster speed than the other existing interfaces and the possibility of installing it without additional devices through the use of the PnP (Plug and Play) feature. Due to these advantages, it is easy to connect various peripheral devices such as the keyboard, mouse, printer, and storage devices to the computer, and this kind of function is supported by a mounted chipset in the motherboard.

Despite these advantages, however, the security of the USB interface was not considered when it was initially designed, and the interface does not ensure safety for this reason [1]. For instance, the data of USB keyboards are exposed, and the plaintext of encrypted data is exposed by the bypassing of the authentication mechanism in secure USB products [1, 2, 3, 4, 5]. One of the other common interfaces besides the USB interface is the PS/2 interface, and this interface uses the specifically assigned ports 0x60 and 0x64 to communicate with peripheral devices such as the keyboard and the mouse. The security of the PS/2 interface is dependent on the access control of the assigned ports because data such as the status information and the scan code are only transferred through specific ports. As a way to monitor specific ports, when a defender wants to set up specific ports to monitor according to the debug ports, the defender is able to ensure an access advantage by calling the registered protection handler if anyone accesses the ports [6].

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_36

Nevertheless, in the case of the USB interface, and in contrast to the PS/2 interface, the interface does not utilize specific ports, and instead is structures maps to the memory. Due to the feature that is shared memory, the problem here is that the transferred data are exposed; for this reason, the USB interface does not ensure the security of the data that are transmitted from/to USB devices. Some researchers have studied the access-control memory technique, but its effectiveness is insufficient [6, 7, 8].

Secure USB memories are classified according to either the hardware approach or the software approach [1, 2]. The hardware approach provides authentication, encryption/decryption, and access control by adding an exclusive security chip inside a USB device. The software approach supports security on the host side by using software without the additional chip. Like the problem of the USB keyboard, secure USB memories do not provide security in the USB protocol; for example, when the memories are authenticated by users, the exposure vulnerability regarding transferred information such as IDs and passwords is intact. Due to this vulnerability, if attackers steal user-authentication information, they can bypass the authentication mechanism to access data that were originally stored in USB memories safely [1, 2]. To counteract this vulnerability, a novel access-control technique is proposed in this paper whereby the exclusive security chip is utilized. Nonetheless, the problem regarding the exposure of the commands for the access control still remained, and this represented the emergence of a new problem [3, 4]. Attackers analyze commands to access secure areas, and they then bypass authentication mechanisms by replaying authentication-related commands; as a result, they are able to access the data that are stored in the secure area. This security problem of both the USB keyboard and secure USB memories also threatens a variety of devices for which the USB is used.

As described above, the problems regarding the variety of devices for which the USB interface is used have increasingly emerged. Of especial concern is a new attack technique called BAD USB, which only emerged recently. BAD USB first appeared at the 2014 Black Hat Security conference, and this technique injects malicious code into the firmware area of the USB controller; actually, this kind of attack is widely known of, but it is difficult to inject manipulated firmware into this particular area [9]. Generally, the firmware of the device is updated for additional functionality, but an attacker abuses this feature to inject an attacker code into the blank area that does not store the firmware, and the code is then executed by the hooking of the specific routine of the firmware. After the manipulation of the firmware has been completed, if the USB device is inserted into the user's computer, attackers can perform malicious behaviors by executing the malicious code in the firmware. In this way, attackers are able to implement attacks such as keyboard emulation and network-card spoofing [9, 10, 11].

The firmware area that is utilized for the implementation of BAD USB is in the area that is invisible to the user, so it is difficult to detect and prevent this kind of the attack. A prevention technique that accepts or rejects the installation of a device and data communication between the host and USB devices based on a black list and a white list has therefore been proposed, but it is not effective enough [10].

In this paper, novel BAD USB-prevention methods are therefore proposed, whereby an access control for which the information and the driver of the actual installing device are used is applied for the verification of the firmware. For this method, the access control that is merely based on a list of devices is used.

2 Related Works

2.1 Overview of BAD USB

The vulnerabilities of the USB interface have been discussed previously, but the discussion has been limited to only the host side until now, and USB-firmware vulnerability was not mentioned. Nevertheless, a new vulnerability emerged at the 2014 Black Hat, and it means that an attacker can manipulate the firmware that is stored in the controller that supports the USB interface. The device that accomplishes malicious behaviors using the above feature is called BAD USB, and it can manipulate the firmware during malicious activities after the BAD USB is inserted [9]. A more-serious problem is that the existing anti-virus software does not detect BAD USB because most of the anti-virus programs detect malicious actions based on the information of the operating system, so detection method that is based on the firmware information of the USB controller is nonexistent. Because of this, when a manipulated USB storage device is deceived, and when other peripheral devices such as the HID (Human Interface Device), the keyboard, and the mouse are also connected, the device can enact malicious behaviors such as the stealing of keyboard information including the inputted password from a user and the manipulation of the information that is transferred from or to the NIC (Network Interface Controller); furthermore, this malicious code is stored in the firmware of the USB device, so if the code is not removed and the USB device is then consistently inserted into other computers, the attacker would be able to steal the sensitive information of many users.

2.2 Attack Scenarios of BAD USB

The attack scenarios of BAD USB are classified into three categories [9, 10]. First, as described above, if a victim acquires a general USB device wherein the firmware in the USB controller has been manipulated and he/she inserts it into his/her computer, the device recognizes a keyboard device through the running of the malicious code. After that, the keyboard information that is inputted by the victim, including authentication information such as IDs and passwords, and financial information such as banking accounts and passwords, can be exposed; furthermore, the malicious attacker can completely fabricate the inputted keyboard information.

Second, BAD USB can cause a critical problem by penetrating private systems such as stuxnet [12]. In terms of the attacker perspective, they can access public systems to prepare an attack, but they are unable to access private systems; for this reason, private systems are generally safe from security threats. In the private system, however, attackers can attack using the peripheral devices that are used by insiders such as stuxnet, and the authors have considered that this kind of attack could also be enacted with the use of BAD USB. This kind of attack has occurred through the stealing of authentication information for the serious infection of private systems subsequently, it could be propagated to an electric-power system or a nuclear-energy system, where it can cause critical damage such as a malicious operation.

Lastly, BAD USB is able to manipulate the transferred data through a network for the preparation of additional attacks. The data that are to be transmitted to the network include actual information as well as the source-IP address and port, the destination-IP address and port, and so on. Attackers can therefore manipulate the transferred information by spoofing the transmitting data from the NIC; for example, one of these attacks is the inducement of a malicious attacker's server through the falsification of the destination-IP address. Through this inducement, the attackers can download an additional malicious code from the server to a victim's PC, and this action causes damages that leads to second and third attacks.

In addition, besides the injection of malicious code into the firmware, attackers are able to access the bootloader in the controller for the execution of malicious codes during the start-up process of a device, and this is one of the most-powerful attacks.

2.3 Existing Countermeasures of BAD USB

As described above, the number of research studies on the countermeasures against BAD USB is increasing because BAD USB can cause serious damage. Three kinds of countermeasures were introduced in recent studies. The first countermeasure is the making of lists by defenders of the trusted and mistrusted USB devices according to the user; then, the user accepts or rejects devices by comparing the new devices that are inserted into the user's computer. The devices that are inserted into the computer send PIDs (Product ID) and VIDs (Vendor ID) to provide information regarding their identities, and the operating system of the computer loads the existing installed device drivers or installs new device drivers based on the received information. For this reason, defenders make white lists based on the PIDs and VIDs of devices, and then they allow new devices when the PIDs and VIDs of the devices are included in the list. If the PIDs and VIDs of devices are not included in the white list, defenders block the device because it is an unreliable device. Alternatively, the defenders also make black lists based on the PIDs and VIDs of devices, and then they block new devices when the PIDs and VIDs of devices are contained in the list. If the PIDs and VIDs of devices are not included in the list, the defenders ask whether or not the device is a trusted device; as a result, when the device is an unreliable device, the defenders add the device to the black list. Accordingly, when the device is trusted, the defenders add the device to the white list. Representative examples of this kind of countermeasure are the access-control methods that are based on white lists and black lists [10]. But for an analysis of actual malicious code, these solutions are determined by the user's choice rather than an access-control method, and a number of problems are therefore apparent. First, the user does not need to use a trusted device, and second, a malicious device is allowed due to the user's mistakes.

A second countermeasure is the prevention of the modification of the firmware by defenders. This solution is supported by the design of the disable-update feature of the firmware, and also by the addition of a new hardware-security module. The implementation of this countermeasure is possible because the design of the current USB controllers provide update functions for the additional features. But an additional cost is incurred for this solution, and the addition of the new hardware module and the produced controllers will not make up for the drawbacks.

A third countermeasure is the prevention of the modification of the boot loader by defenders. The boot loader loads and runs the firmware so that defenders can prevent the forgery and alteration of the firmware by immobilizing the loaded address and the firmware size. But if the boot loader is not modified by the third party, an attacker can produce BAD USB using the firmware, and in this respect, this countermeasure is insufficient.

3 BAD USB Countermeasures for the Detection of the Manipulated Firmware

As described above, the existing BAD USB countermeasures are not effective enough; for this reason, two countermeasures are proposed in this paper. First, one of the countermeasures uses an access-control method that is based on the actual installed information of the devices and the device drivers. The second countermeasure verifies the firmware by comparing the firmware codes between the original firmware and the installing firmware.

For an understanding of the proposed approach, the installation process of the device driver that is based on the Microsoft Windows platform is subsequently described.

3.1 Driver-installation Process on Microsoft Windows Platform

Fig. 1 shows the device-driver installation process on the Microsoft Windows platform [13].

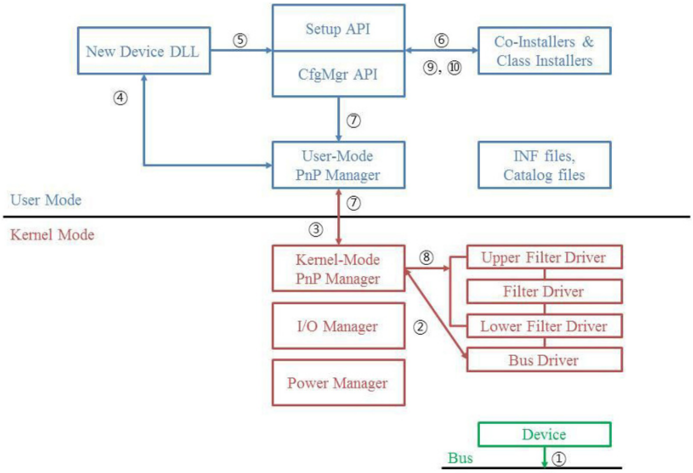


Fig. 1. Installation process of the device driver on Microsoft Windows Platform

Step 1. Users insert USB devices into their computers. If the bus and device of the computer support hot-plug notifications, the users can insert new devices even if the system is running.

Step 2. When devices are inserted successfully into the system, the devices are enumerated and the bus driver receives the hot-plug notifications of the new devices. The bus driver then informs the PnP manager of the kernel mode of the notifications, and the list of bus devices is changed by the calling of the IoInvalidateDeviceRelations function. In this case, the change is a new device on the bus. The PnP manager of the kernel mode queries the list of devices to the bus driver by sending IRP_MJ_PNP with IRP_MN_QUERY_DEVICE_RELATIONS for the BusRelations of the device. The bus driver replies to the IRP with a list of current devices, and the PnP manager recognizes new devices by comparing the lists between the previous devices and the new devices. Hereafter, the PnP manager of the kernel mode collects the information that is related to new devices such as hardware IDs and compatible IDs by sending IRP like IRP_MN_QUERY_ID and IRP_MN_QUERY_CAPABILITIES to the bus driver.

Step 3. The PnP manager of the kernel mode sends a message that the device will be installed by the PnP manager of the user mode. The PnP manager of the user mode attempts the trusted installation process; if it is difficult for the manager to install the trusted installation process, the manager requests the direct usage of the installation drivers.

Step 4. The PnP manager of the user mode creates a new process for the installation using rundll32.exe and starts newdev.dll to install the devices.

Step 5. The newdev.dll (new device dll) calls the Setup API among the device-installation functions and the CfgMgr API from among the PnP-configuration-manager functions to start the installation process. Afterward, the newdev.dll calls the SetupDiBuildDriverInfoList function to generate a possible driver list; in this case, if the operating system does not have an INF file, the function returns an empty driver list. After that, the newdev.dll displays the Found New Device wizard, and the user chooses the location of the device driver such as the floppy disk, the CD, or Microsoft Windows Update. Hereafter, if the location of device driver is inputted, the newdev.dll calls the SetupDiBuildDriverInfoList function again and receives a list that includes the selected device driver.

Step 6. If there are a class installation and co-installations, these can participate in the installation process by handling the DIF request; for example, the newdev.dll calls the SetupDiCallClassInstaller function by sending the installation request of the SELECTBESTCOMPATDRV. Hereafter, the setup process uses both the ClassGuid and the Class of the INF-version section of the devices to determine the device-setup class.

Step 7. The setup process loads the device driver and passes control to the kernel mode to start the device. In this step, the control is delivered to the PnP manager of the kernel mode by a request that is sent to the PnP manager of the user mode in the proper CfgMgr function.

Step 8. The PnP manager loads the selective filter driver and the appropriate functional driver for the devices. Afterward, the PnP manager calls the DriverEntry routine for the necessary drivers that are not loaded. Hereafter, the manager calls the AddDevice routine for each driver, the lower-filter drivers, the function driver, and any upper filter drivers. During this process, the PnP manager assigns resources to the devices, and if necessary, the manager transfers IRP_MN_START_DEVICE to the device driver of the related proper device.

Step 9. The installer can provide the finish-install Wizard page for the installation of an application program and to change the device configuration. After that, the setup transfers the DIF_NEWDEVICEWIZARD_FINISHINSTALL request before the display of the standard complete page.

Step 10. The installer from Windows Vista can support the finish-install actions instead of the finish-install Wizard page for the installation of the application program; to do this, the setup transfers the DIF_FINISHINSTALL_ACTION request after all of the other installation works are complete.

3.2 BAD USB-prevention Method for which the driver integrity is verified

In the case of BAD USB, malicious behaviors are not only enacted in the firmware itself through the manipulation of the firmware, but it is also possible to lead the driver installation that is the inserted malicious code. In this paper, a prevention method for the malicious device driver for which the integrity of the installing device driver is verified is therefore proposed.

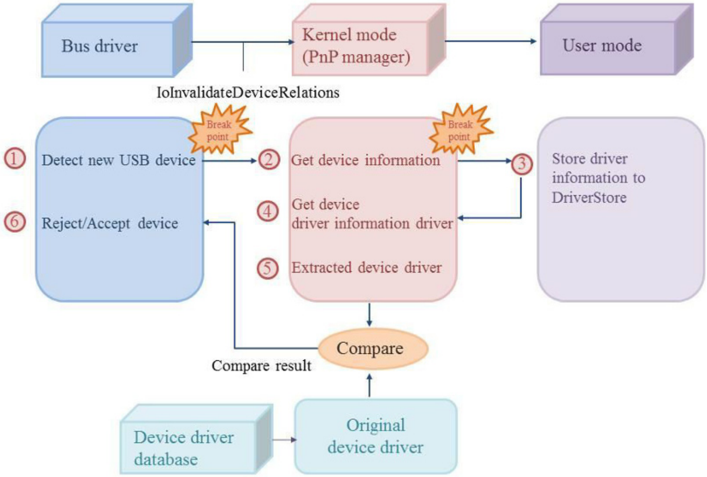


Fig. 2. Proposed prevention method for which the integrity of the installing device driver is verified

First of all, it is assumed that a prevention system consists of a database that stores the original device driver's information for the verification of the integrity of the device driver, and that the database can be in a local host or a remote server. The information of the device driver can comprise the binary form, and it is possible to store hashed values or cipher text to prevent the exposure of the original data.

The proposed prevention method must complete the entire process before the installation and starting of the device driver because the driver could contain malicious code; therefore, the proposed method must verify the integrity of the device driver in the state where the driver is not installed. When the device is inserted into a user's computer, the bus driver detects the inserted device and notifies it to the PnP manager of the kernel mode by calling the IoInvalidateDeviceRelations function. The IoInvalidateDeviceRelations function that is called in this process is called with the several parameters [14] that are the pointers of the DEVICE_OBJECT structure and the DEVICE_RELATION_TYPE; for this reason, when the operating system calls this function, the proposed prevention method extracts the information of a connecting device by using the hooking technique. In particular, the DEVICE_OBJECT structure includes the variety of device data, as well as the DRIVER_OBJECT structure that is one of these data [15]. The DRIVER_OBJECT structure [16] contains the address of the loaded device driver, the size of the driver, and the driver name, and the proposed method collects this information; then, the subsequent data that is for the installation of the device driver are extracted.

When the collection of the information is complete, a prevention program continues the installation by passing the execution control to an original function that is not the hooked function. Next, if all of the installation steps in the user mode are complete, the execution control turns back to the installation of the kernel mode, and the PnP manager loads the installing device driver; therefore, the prevention program obtains the information of the loaded driver based on the collected information during the above steps, and the program therefore verifies the integrity of the device drivers by comparing the data of the extracted driver and the data of the original device driver that are stored in the database. If the installing device driver is manipulated by a third party, because the driver is different from the original driver, the manipulated driver is able to verify the integrity by using the hash operation, and so on. As a result of the verification process, the case of the installing device driver is different from the original driver, and this means that the driver does not ensure integrity. The prevention program quits the installation of the driver due to this lack of integrity; otherwise, the case of the installing device driver is the same as that of the original driver, whereby the driver is not manipulated by a third party. In this respect, the prevention program returns the installation routine ordinarily.

The integrity-verification routine in the prevention program is capable of performing an inspection at the time when the driver is loaded, and the prevention program is also able to examine the data integrity of the other installed drivers when the driver data are extracted in all of the installation processes.

3.3 BAD USB-prevention Method for which the Firmware Integrity is verified

The proposed prevention method for which the verification of the driver integrity is used is capable of preventing the installation of a malicious device driver, but this method does not fundamentally verify the actual malicious code itself. To solve this problem, a prevention method for which the firmware integrity is verified can detect and neutralize BAD USB through the examination of the malicious code inside the firmware, and after the extraction of the malicious firmware that is stored in the controller.

A manufacturer wants to update the firmware for maintenance and the adding or deleting of functions, and this proceeds with the sending of new written complemented firmware in accordance with the updating device, followed by the overwriting of the received firmware to the controller. Hereafter, the inspection step of the firmware is carried out for the normal detection of transferred-firmware errors. In this step, the manufacturer determines that there is no error in the transmission process if it is the same as that of the original firmware after the reading of the written firmware. In this process, specific commands such as a write command to the firmware and a read command to the firmware are transferred between the controller and a host. For this reason, the prevention program is capable of analyzing malicious behaviors based on the received firmware that is sent from the controller after the read command is received. Research studies that analyze malicious behaviors have already been studied considerably, and for these kinds of analysis methods, integrity-verification methods such as the utilization of the signature, the static analysis, and the dynamic analysis are used. In this step, when the prevention program detects malicious code, the program is capable of neutralizing the threat through the rewriting of the original firmware that is stored in the database.

4 Conclusions

In this paper, the surveyed results of BAD USB, which has recently emerged, are described, and the proposed BAD USB-installation prevention methods are based on the surveyed results. The method for the prevention of the BAD USB installation is the prevention of forgery and the alteration of both the firmware and the device driver, and it is based on a verification of the integrity. The existing countermeasures prevent device installation with the use of the access control that is based on black and white lists. When the device and the device driver are not included in the lists, the users can approve or reject the installation of a device, while the device driver determines whether or not the device and the driver will be installed. This method does not analyze malicious behaviors practically, though, so the outcome is inadequate; therefore, in this paper, the forgery and alteration of both the firmware and the device driver are achieved through the verification of the subject that will perform malicious acts such as the installation of a malicious driver or malicious firmware. Through these countermeasures, the proposed methods can prevent the installation of BAD USB. In the future, the proposed prevention methods will be implemented, and the proposed concept will be proved through the implementation of a sample BAD USB.

Acknowledgments. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1D1A1A01057300) and the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2016-R0992-16-1006) supervised by the IITP (Institute for Information & communications Technology Promotion).

References

1. K. Lee, K. Yim, and E. H. Spafford, Reverse-safe authentication protocol for secure USB memories, Journal of the Security and Communication Networks (SCN), vol. 5, iss. 8, pp. 834-845, Aug. 2012.
2. K. Lee, H. Yeuk, Y. Choi, S. Pho, I. You, and K. Yim, Safe Authentication Protocol for Secure USB Memories, Journal of the Wireless Mobile Networks, Ubiquitous Computing and Dependable Applications (JoWUA), vol.1, num.1, pp.46-55, Jun. 2010.

3. J. Kim, Y. Lee, K. Lee, T. Jung, D. Volokhov, and K. Yim, Vulnerability to Flash Controller for Secure USB Drives, Journal of the Internet Services and Information Security (IMIS), vol.3, num.3/4, pp.136-145, Nov. 2013.
4. H. Jeong, Y. Choi, W. Jeon, F. Yang, Y. Lee, S. Kim, and D. Won. Vulnerability analysis of secure usb flash drives. Proceedings of the 2007 IEEE International Workshop on Memory Technology, Design and Testing, (MTDT'07), Taipei, Taiwan, pages 61–64. IEEE, December 2007.
5. S. L. Jewan Bang, ByeongYeong Yoo. Secure usb bypassing tool. Journal of the Digital Investigation, 7(Supplement):S114–S120, August 2010.
6. K. Lee and K. Yim, Keyboard Security: A Technological Review, Proceedings of the Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), pp.9-15, Jun. 2011.
7. K. Lee, K. Bae, and K. Yim, Hardware Approach to Solving Password Exposure Problem through Keyboard Sniff, Academic Science Research, WASET, pp.23-25, Oct. 2009.
8. K. Lee, W. Kim, K. Bae, and K. Yim, A Solution to Protecting USB Keyboard Data, Proceedings of the International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA), pp.108-111, Nov. 2010.
9. K. Nohl and J. Lell, BadUSB – on accessories that turn evil, Black Hat USA, Aug. 2014.
10. S. Neuner, Marshall Plan Scholarship Final Report: Security of the Universal Serial Bus, Dec. 2014.
11. D. J. Tian, A. Bates, K. Butler, Defending Against Malicious USB Firmware with GoodUSB, Proceedings of the Annual Computer Security Applications Conference (ACSAC), pp.261-270, 2015.
12. R. Langner, Stuxnet: Disserting a Cyberwarfare weapon, Journal of the IEEE Security & Privacy, vol. 9, iss. 3, pp.49-51, Jun 2011.
13. Microsoft Developer Network, Device and Driver Installation Example, https://msdn.microsoft.com/en-us/library/ff541158.aspx
14. Microsoft Hardware Dev Center, IoInvalidateDeviceRelations routine, https://msdn.microsoft.com/en-us/library/windows/hardware/ff549353(v=vs.85).aspx
15. Microsoft Hardware Dev Center, DEVICE_OBJECT structure, https://msdn.microsoft.com/en-us/library/windows/hardware/ff543147(v=vs.85).aspx
16. Microsoft Hardware Dev Center, DRIVER_OBJECT, https://msdn.microsoft.com/en-us/library/windows/hardware/ff544174(v=vs.85).aspx

Security Assessment on the Mouse Data using Mouse Loggers

Hyeji Lee[1], Yeunsu Lee[1], Kyungroul Lee[2], Kangbin Yim[1]
[1] Dept. of Information Security Engineering, Soonchunhyang University
Asan, South Korea
{ gpwl899, yunsu07, yim}@sch.ac.kr
[2] R&BD Center for Security and Safety Industries (SSI), Soonchunhyang University
Asan, South Korea
carpedm@sch.ac.kr

Abstract. The mouse device, one of the computer peripherals, is an input device that recognizes mouse movements on the two-dimensional plane of a monitor, and it is possible to use it conveniently or inputting and editing during the running of application programs. The existing password-based authentication methods are vulnerable to keyboard-data exposure, so a new authentication method that is based on the mouse input has emerged. Nevertheless, a security assessment of the mouse-based authentication method has not been practically analyzed; for this reason, the exposure of mouse data is evaluated in this paper based on the mouse loggers that can be easily obtained from Web sites. It will be possible to utilize the analyzed result to formulate a security guideline to prevent the exposure of authentication information that is based on the image-based authentication.

1 Introduction

A mouse device is an input device that recognizes mouse movements on the two-dimensional plane of a monitor, whereby it moves an arrow-shaped cursor that is displayed on the screen and transmits commands using buttons [1]. In addition, along with the other input device, the keyboard, it facilitates the control of application programs and enables a variety of editing commands. In the past, the mouse was connected universally via the PS/2 interface, but the current trend is the connection of the mouse via the USB interface. The PS/2 interface enables the infinite inputting of the keys, and it is advantageous for specific services because of the high maximum number of key inputs; alternatively, the USB interface enables the convenient use of the plug-in structure.

The previous password-based authentication method that was often used involved the input of the password from the keyboard, but the exposure of the keyboard data created a problem [2]; for this reason, a more-secure authentication method is required. To counteract this problem, the image-based authentication emerged, and this method identifies users through a password that is the clicked information from the mouse on the displayed screen image. This method is mainly utilized for high-priority services such as e-commerce and Internet-banking services; nevertheless, the mouse data that are used for the image-based authentication can now also be exposed.

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_37

Moreover, defense techniques have been researched for the protection of keyboard data [3], but the research for these kinds of protection techniques regarding the mouse device is insufficient. In this paper, the security of the mouse data for e-commerce and Internet-banking services is therefore investigated based on the mouse loggers that are easy to obtain from the Internet.

2 Related Works

2.1 Mouse Loggers

The "mouse logger," a new compound word of "mouse" and "logger," displays mouse movements as coordinates, and it records the histories of specific functions such as the click information. In addition, loggers can monitor and record the movements of the mouse and the time of action by replaying the mouse information that is collected from the user with the use of the recording feature [4]. Some mouse loggers provide the recording function for the keyboard information, whereby all of the mouse and keyboard-input information can be recorded; these loggers are Mouse Recorder Pro 2[5], Axife Mouse Recorder [6], Automatic Mouse and Keyboard [7], and Ghost Mouse [8], and Table 1 shows their characteristics.

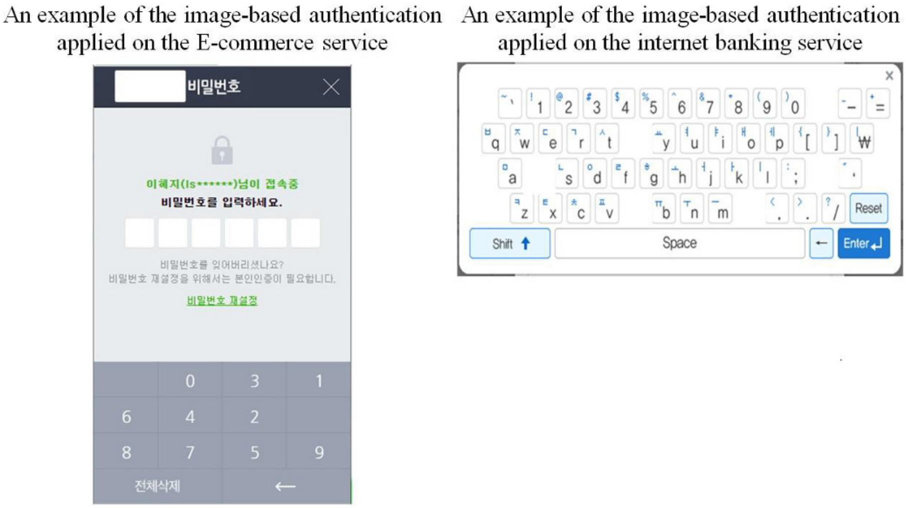
Table 1. Characteristics of mouse loggers

Logger name	Record method	Features
Mouse Recorder Pro 2	Save inputted mouse information to file	• Record coordinates • Record screen images
Axife Mouse Recorder	Save inputted mouse information to file	• Record coordinates • Record keyboard information
Automatic Mouse and Keyboard	Save inputted mouse information to program	• Record coordinates • Record screen images • Record keyboard information
Ghost Mouse	Save inputted mouse information to file	• Record coordinates

As described above, the main functions of the mouse loggers are the recording function for the recording of the coordinates and the production of the screen images for the visualization of the recorded coordinates; however, the features of each mouse logger are different. In the case of the recording-coordinates method, varying methods exist for the storage of a specific file and also for the internal storage of program; moreover, some loggers provide a keyboard-information recording feature. Mouse Recorder Pro 2 and Axife Mouse Recorder provide both the recording coordinates and the screen images, Automatic Mouse and Keyboard provides the recording coordinates, the keyboard information, and the screen images and Ghost Mouse only records the screen images.

2.2 Image-based Authentication

For the image-based authentication, the click information on the displayed image is utilized as the password; that is, a specific image is displayed on the monitor, and the user then inputs the clicked information on the displayed image. When the clicked information is equal to the clicked information that is registered on the authentication server, the user is identified correctly. This method is mainly utilized for high-priority services such as e-commerce and Internet-banking services. Figure 1 shows examples of the application of the image-based authentication.


An example of the image-based authentication applied on the E-commerce service

비밀번호

이혜지(ls******)님이 접속중
비밀번호를 입력하세요.

비밀번호를 잊어버리셨나요?
비밀번호 재설정을 위해서는 본인인증이 필요합니다.

비밀번호 재설정

An example of the image-based authentication applied on the internet banking service

Fig. 1. Examples of the application of the image-based authentication for e-commerce and Internet-banking services

As shown in the figure, this method protects the inputted information that is utilized as the authentication information; nevertheless, the sensitive information can still be exposed to a malicious attacker. In this paper, the security of mouse data that is based on the surveyed mouse loggers that are described in Section 2.1 is therefore evaluated.

3 Security Assessment on the Mouse Data

In this paper, to assess the security of the mouse data, experiments wherein the mouse data of credit-card companies and Internet-banking sites of South Korea are exposed were conducted. The experimental environment consists of the Intel® Core™ i5-2410M CPU @ 2.30GHz with Microsoft Windows 7 installed, 32bit, and a 4 GB RAM. The Internet Explorer program was used to connect the web-sites, and a USB optical mouse was used to input the mouse data.

3.1 Security Assessment on the E-Commerce Services

To derive the exposure of the mouse data for the e-commerce services, the web-sites were surveyed using the image-based authentication of a variety of e-commerce websites, and Table 2 shows the surveyed results.

Table 2. Password-input methods of credit-card companies of South Korea

Credit card company	Payment method	Input method
Company A	• ISP easy payment	Keyboard
Company B	• General payment • Easy payment • One-click easy payment	Keyboard
Company C	• ISP easy payment	Keyboard
Company D	• One-click easy payment • App card • General payment	Keyboard
Company E	• Smart payment • General payment • Password payment	Keyboard, mouse
Company F	• Login easy payment • Company F pay	Keyboard, mouse

The six credit-card companies from Table 2 were surveyed, and two of the companies, company E and company F, utilized the image-based authentication. The exposure of the mouse data is therefore based on two companies for this paper, and the experiment results are shown Tables 3 and 4.

Table 2. Experiment result of company E

	Mouse Recorder Pro 2	Axife Mouse Recorder	Ghost mouse	Automatic Mouse and Keyboard
Coordinate position	O	O	X	O
Cursor record	O	O	O	O
Input window	X	O	X	O
Exposure	X	O	X	O

Table 3. Experiment result of company F

	Mouse Recorder Pro 2	Axife Mouse Recorder	Ghost mouse	Automatic Mouse and Keyboard
Coordinate position	O	O	X	O
Cursor record	X	O	O	O
Input window	O	X	O	O
Exposure	X	X	O	O

In a detailed analysis of the results, the exposure information is classified according to the coordinate position, cursor record, input window, and exposure. The coordinate position denotes the position of the mouse coordinate, and the cursor record and input windows denote the outputs of the cursor and the input window, respectively, when the recording video is replayed. Exposure means that the attacker directly distinguishes the mouse movement through the synthesis of the above-three criteria.

In the case of company E, Mouse Recorder Pro 2 obtained the coordinate position and cursor information, but the input window was not shown. The input window displays each session randomly, so when the attacker does not get it, he or she does not know the password; for this reason, this company is safe from this mouse logger. Ghost Mouse was unable to steal the password because it only obtained the cursor information. Lastly, Axife Mouse Recorder and Automatic Mouse and Keyboard obtained the coordinate position, cursor record, and input window, so they were therefore able to steal the password information.

In case of the company F, Mouse Recorder Pro 2 obtained the coordinate position and input window, but it did not obtain the cursor information. Axife Mouse Recorder also obtained the coordinate position and cursor information, but it did not obtain the input window; for this reason, this company is safe from these mouse loggers. Ghost Mouse did not obtain the coordinate position, but the cursor information and input window were obtained. Lastly, Automatic Mouse and Keyboard obtained all of the information; therefore, this company is not safe from the above-two mouse loggers.

As a result of the security assessment that is based on the analyzed result, Automatic Mouse and Keyboard can steal the passwords of two companies, and Mouse Recorder Pro 2 is unable to steal the passwords of two companies. Moreover, in the case of keyboard input, websites require the installation of a secure keyboard program; however, in the case of mouse input, it is likely that a protection solution is not installed, so most websites do not prevent the exposure of the password by mouse loggers.

3.2 Security Assessment on the Internet-banking Services

To derive the exposure of mouse data for the Internet-banking services, websites with the image-based authentication were surveyed, and six banking sites were selected. Table 5 shows the surveyed results.

Table 4. Experiment results of Internet-banking services

		Company G	Company H	Company I	Company J	Company K	Company L
Mouse Recorder Pro2	Coordinate position	O	O	O	X	O	O
	Cursor record	O	O	O	X	O	O
	Input window	O	O	O	X	O	O
	Exposure	O	O	O	X	O	O
Axife Mouse Recorder Automatic	Coordinate position	O	O	O	X	O	O
	Cursor record	O	O	O	X	O	O
	Input window	O	O	O	X	O	O
	Exposure	O	O	O	X	O	O
Ghost mouse	Coordinate position	X	X	X	X	X	X
	Cursor record	O	O	O	X	O	O
	Input window	O	O	O	X	O	O
	Exposure	X	X	X	X	O	O
Automatic Mouse and Keyboard	Coordinate position	O	O	O	O	O	O
	Cursor record	O	O	O	O	O	O
	Input window	O	O	O	O	O	O
	Exposure	O	O	O	O	O	O

In the cases of companies G, H, and I, Ghost Mouse did not obtain the coordinate position; for this reason, this company is safe from this mouse logger. Mouse Recorder Pro2, Axife Mouse Recorder Automatic, and Automatic Mouse and Keyboard obtained the coordinate position, cursor information, and input window; for this reason, this company is not safe from the above mouse loggers. In the case of company J, Mouse Recorder Pro2, Axife Mouse Recorder Automatic, and Ghost Mouse did not extract any information; for this reason, this company is safe from these mouse loggers. Automatic Mouse and Keyboard obtained all of the information, and the mouse movement during the replaying of the recorded video was confirmed; for this reason, this company is not safe from this mouse logger. Lastly, in the cases of companies E and F, all of the mouse loggers obtained all of the information; therefore, this company is not safe from all of the mouse loggers.

As a result of a security assessment that is based on the analyzed results, most of the Internet-banking websites did not protect mouse data; in particular, Automatic Mouse and Keyboard logger stole all of the mouse-related information from all of the websites. Internet-banking websites do not apply secure solutions to protect mouse data, so this vulnerability can lead to the enactment of financial crimes through the abuse of stolen passwords; therefore, a security solution for the prevention of the exposure of mouse data is urgently required.

4 Conclusions

To counteract the vulnerability of the password-based authentication that is due to the exposure of keyboard data, the image-based authentication emerged; this newer method is utilized for high-priority services such as e-commerce and Internet-banking services. The most-important factor, however, is the mouse data because this method uses authentication information that is a combination of a displayed image and the clicked information from the mouse. Nevertheless, when mouse data is exposed in the same manner as the exposure of keyboard data, the image-based authentication does not ensure security; therefore, in this paper, the security of the image-based authentication that is based on the exposure of mouse data through the use of mouse loggers, which are easy to obtain from the Internet, was assessed. According to the experiment results, most of the e-commerce and Internet-baking websites did not protect mouse data; accordingly, protection solutions regarding mouse data will be studied in the future because the image-based authentication does not ensure the security of mouse data.

Acknowledgments. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1D1A1A01057300) and the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2016-R0992-16-1006) supervised by the IITP (Institute for Information & communications Technology Promotion).

References

1. Wikipedia, Computer_mouse, https://en.wikipedia.org/wiki/Computer_mouse
2. Kyungroul Lee, Youngtae Choi, Hyeungjun Yeuk, and Kangbin Yim, Password Sniff by Forcing the Keyboard to Replay Scan Codes, Proceedings of the JWIS (Joint Workshop on Information Security), pp.9, Aug. 2010.
3. Kyungroul Lee and Kangbin Yim, Keyboard Security: A Technological Review, Proceeding of the IMIS (Innovative Mobile and Internet Services in Ubiquitous Computing), pp.9-15, Jun. 2011.
4. Z Minchev, G Dukov, and S Georgiev, EEG spectral analysis in serious gaming: An ad hoc experimental application, BIO Automation, 13(4), pp.79-88, 2009.
5. LO4D.com, Mouse Recorder Pro 2, http://mouse-recorder-pro-2.en.lo4d.com
6. softonic, Axife Mouse Recorder, http://axife-mouse-recorder.en.softonic.com
7. Download.com, Automatic Mouse and Keyboard, http://download.cnet.com/Automatic-Mouse-and-Keyboard/3000-2084_4-75324350.html
8. Ghost Mouse, Ghost mouse, http://www.ghost-mouse.com

Security Assessment of Keyboard Data Based on Kaspersky Product

Seungho Lee[1], Kyungroul Lee [2], Kangbin Yim[1]

[1] Dept. of Information Security Engineering, Soonchunhyang University
Asan, South Korea
{pods0912, yim}@sch.ac.kr

[2] R&BD Center for Security and Safety Industries(SSI), Soonchunhyang University
Asan, South Korea
carpedm@sch.ac.kr

Abstract. To protect keyboard data that includes sensitive information such as authentication data, secure keyboard programs preempt the extraction of the keyboard data before they are stolen by an attacker. Although these programs are operated at a variety of defense levels, some secure keyboard programs are still vulnerable to hardware-level attacks; therefore, in this paper, the security of keyboard data is evaluated according to the functionality of the Kaspersky "Internet Security" product, whereby secure information is inputted into a variety of websites that comprises SNSs, an email account, and a banking service.

1 Introduction

During the development of modern society into an information-oriented society, the user of a variety of Internet services, such as the SNS (social-network service), the banking service, and email, must input confidential information such as authentication information and personal information via the keyboard; however, the information that is distributed via the keyboard creates a problem whereby the user is exposed to the variety of existing attack techniques. To solve this problem, keyboard-security solutions have appeared whereby the purpose is the prevention of the exposure of confidential user information to third parties [1] so that the user can safely use the variety of Web services.

An introduced keyboard-security solution appears to protect the transmitted keyboard data based on a variety of the technologies from different companies. These techniques involve the encryption of the information of the keyboard that is delivered on the basis of an encryption key, known as the "replace" method, and after the transmitted information of the keyboard is stored in the memory, its value and the knowledge of its removal are acquired [2].

Despite these advantages, the vulnerability of this keyboard-security solution is that it cannot correspond to low-level attack techniques. The existing attack techniques are based on techniques such as procedure-hooking and the DLL injection in the user mode, whereby an attempt is made to hijack the keyboard data; however, with the advances of the attack techniques, the attacks of low-level attack techniques are apparently based on the operating system and the hardware. These techniques are

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_38

a substitution for the interrupt handler [3], but direct polling means that it is possible to preoccupy the keyboard data that are to be transmitted by a specific command transmission that overrides the keyboard-security solution.

Likewise, while the exposure of keyboard data occurs, to complement these problems, authentication technologies such as image-based passwords [4], rather than the keyboard data, which can be implemented to take advantage of the information from the mouse, have been studied; however, these authentication techniques are affected by the same vulnerabilities due to the use of the same interface such as the PS/2 interface and the USB interface. A more-serious keyboard-security problem that does not apply to the variety of services from Korea and most other countries affects SNS services like Facebook and Twitter, the Chase Bank banking service, and the Outlook.com mail service.

For this reason, to protect the keyboard data that are input for a variety of services, the internationally renowned security company Kaspersky developed the "Internet Security" product. For these products, the security of the keyboard is ensured with the application of a password-input window for the websites that provide services like those that are mentioned above; however, a safety assessment of the security technology has not yet been performed. In this paper, an assessment of the safety of the keyboard information that is transmitted for a variety of Web services is performed.

2 Related Research

2.1 Keyboard-data Transmission Process

If keyboard data is entered via the keyboard by the user, the keyboard interprets the information of the pressed keyboard keys, and this is then transferred to the controller that processes interrupts such as the PIC and the APIC. The controller interprets the interrupt, and causes the keyboard interrupt in the host processor, and through the generated interrupt, the interrupt handler that is prepared by the operating system (OS) is called. The called interrupt handler is transferred in the application program to the keyboard data through the driver layer that is associated with a hierarchically structured keyboard for the provision of the keyboard information to the application, and the application receives the keyboard data that are transmitted according to the message [5]. This keyboard-data transfer process is shown in Fig. 1.

2.2 Keyboard-security Program (Kaspersky)

The "Internet Security" products that are developed in the Kaspersy lab are for the protection of personal user information, and the advantage of the Kaspersky lab, worldwide high-level security company, is a high diagnostic rate and a low misdiagnosis rate. These products provide security such as the protection of personal information and Internet banking, the media-control function for the prevention of security threats through removable media, the heuristic-detection (artificial-intelligence detection) technology that makes step-by-step configuration possible, the interface that makes it possible to check the interlocking state of the central management server, and the emergency-repair disk that is based on Linux to provide more-effective security [6].

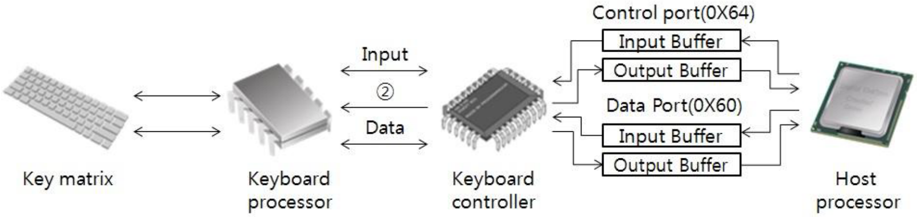


Fig. 1. Process of keyboard-data transmission

2.3 Existing Keyboard-attack Technique

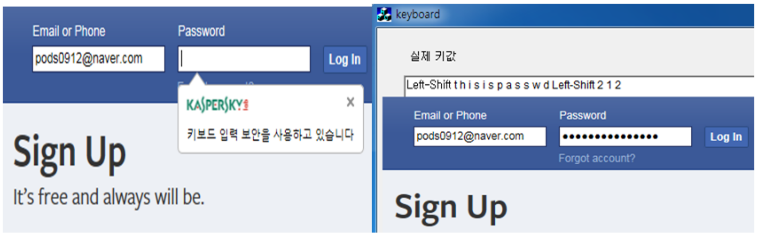
For the extraction of keyboard data, the attack techniques such as direct polling, C/D-bit utilization [7], and RESEND-command utilization [8] are diverse. Direct polling, a technique for the extortion of the keyboard data that is based on the provided information of the keyboard controller, is the attack technique that is used in this paper. The keyboard controller performs a role for the running and transfer of a variety of commands like the information that is inputted into the keyboard, or the information that is outputted from the keyboard. For this reason, once the information from the keyboard is inputted, the status information for the notifying of it will be updated in the OS that checks the status information, and it acquires the keyboard data that are transmitted to the keyboard controller from the keyboard. If the malicious code of the attacker affects the keyboard data before the OS acquires the keyboard data, the safety of the keyboard data cannot be ensured by the OS; therefore, in this paper, the direct-polling attack technique is used to evaluate the ability of the Kaspersky Internet Security product to ensure the safety of the information that is entered for the variety of services.

3 Keyboard-data Safety Rating

As mentioned above, this paper presents an evaluation of the safety that is provided by the Kaspersky Internet Security product regarding the keyboard data for the SNS service, mail service, and banking service. The exposure of the experimental subjects, the Facebook and Twitter SNS services, the Outlook.com mail service, and the Chase Bank banking service was verified. The experimental environment consists of the Intel(R)Core(TM) i5-5200U CPU @ 2.2 GHz processor, the 32-bit Windows 7 OS, and the PS/2 keyboard, and the Internet Explorer Web-browsing application was used.

3.1 Safety Assessment of Keyboard Data for SNS

Based on the Facebook and Twitter SNS services, for which the Kaspersky product was driven and a sample of the direct-polling attack program was applied, the safety of the keyboard data was evaluated. The results are shown in Figs. 2 and 3.



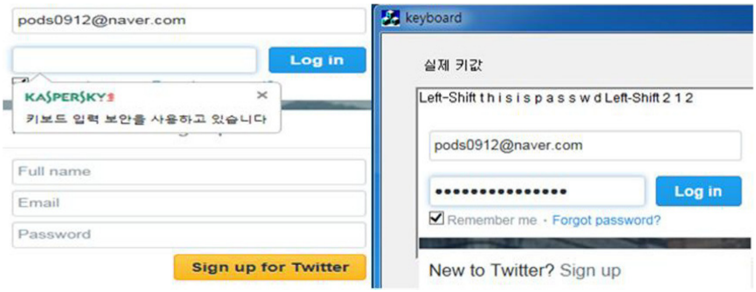
pods0912@naver.com

키보드 입력 보안을 사용하고 있습니다

Left-Shift t h i s i s p a s s w d Left-Shift 2 1 2

pods0912@naver.com ••••••••••••••

Fig. 2. Keyboard-data export in Facebook



pods0912@naver.com

키보드 입력 보안을 사용하고 있습니다

Left-Shift t h i s i s p a s s w d Left-Shift 2 1 2

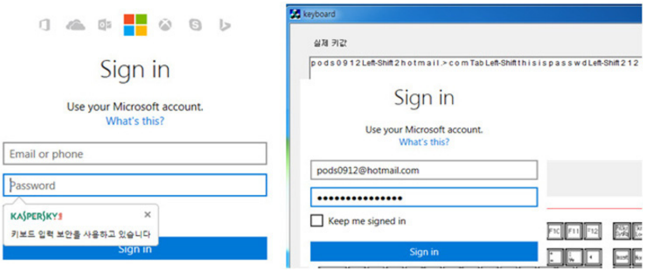
pods0912@naver.com ••••••••••••••

Fig. 3. Keyboard-data export in Twitter

As is shown, for these SNS products, the Internet Security product did not secure the keyboard data.

3.2 Safety Assessment of Keyboard Data for Mail Service

The safety assessment for the Outlook.com service was carried out in the same manner as the SNS-service experiments. Based on the Web site for which the Kaspersky product was driven, the safety of the keyboard data after a sample of the attack program was applied was evaluated. The results are shown in Fig. 4.



```
p o d s 0 9 1 2 Left-Shift 2 h o t m a i l . > c o m Tab Left-Shift t h i s i s p a s s w d Left-Shift 2 1 2
```

pods0912@hotmail.com

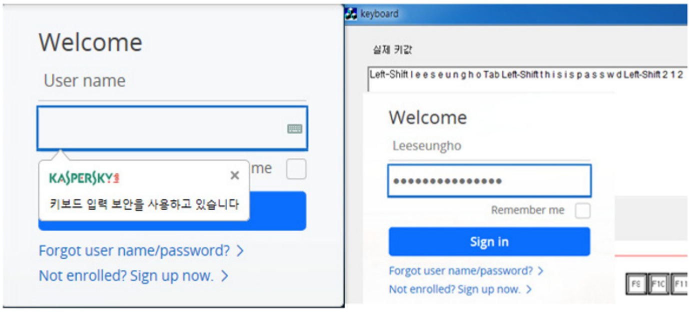
```
•••••••••••••••
```

Fig. 4. Keyboard-data export in Outlook.com

As is shown, although the Kaspersky product was operational, the Outlook.com ID and password were exposed; therefore, keyboard-data security had not been provided.

3.3 Security Assessment of Keyboard Data for Banking Service

Lastly, based on the Chase Bank banking-service website for which the Kaspersky product was driven, the safety of the keyboard data was evaluated. The result is shown in Fig. 5.


Welcome
User name
KASPERSKY
키보드 입력 보안을 사용하고 있습니다
me
Forgot user name/password?
Not enrolled? Sign up now.
keyboard
실제 키값
Welcome
Remember me
Sign in
Forgot user name/password?
Not enrolled? Sign up now.

```
Left-Shift l e e s e u n g h o Tab Left-Shift t h i s i s p a s s w d Left-Shift 2 1 2
```

Leeseungho

```
•••••••••••••••
```

Fig. 5. Keyboard-data export for Chase Bank

As is shown, although the Kaspersky product was operational, the bank ID and password were exposed; therefore, keyboard-data security had not been provided. The results of the experiment are regarding a variety of service websites for which the Kaspersky keyboard-security technology was operational, whereby it was verified that sensitive user information such as IDs and passwords, which were input via the keyboard, was exposed. The results are shown in Table 1.

4 Conclusions

For the safe protection of the information that is distributed from the keyboard, the Internet Security product of Kaspersky can be used; however, the absence of the safety assessments of keyboard-security solutions is reality. For this paper, we therefore evaluated the safety of keyboard data based on a sample attack program for SNS, mail, and banking Web services that had the Kaspersky product driven into them. The evaluation results confirmed the exposure of the IDs and passwords that are required for all of the websites. Based on the results of the experiment, the Kaspersky product does not ensure the safety of keyboard data. In a future research study, the security of keyboard-security technology will be evaluated using a variety of keyboard-attack technologies.

Acknowledgments. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1D1A1A01057300) and the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2016-R0992-16-1006) supervised by the IITP (Institute for Information & communications Technology Promotion).

References

1. Baig, Muzammil M., and Wasim Mahmood. "A robust technique of anti key-logging using key logging mechanism." 2007 Inaugural IEEE-IES Digital EcoSystems and Technologies Conference. IEEE, 2007
2. Kyungroul Lee, and Kangbin Yim ″Keyboard Security: A Technological Review” IMIS (Innovative Mobile and Internet Services in Ubiquitous Computing), pp.9-15, Jun. 2011.
3. "Intel Architechtur Software Developer's Manual Vol.3 System Programming", Intel Corpration, 1999
4. Susan Wiedenbeck, et al., "PassPoints : design and longitudinal ebaluation of a graphicl password system" International Journal of Human- Computer Studies, v.63 n.1-2, p.102-127, July 2005
5. Sanchez, "IBM PC/AT Technical Reference Manual", IBM Corporation, 1985
6. Kaspersky Lab, http://usa.kaspersky.com
7. Taeyoung Jung, Kangbin Yim, "Countermeasures to the Vulnerability of the Keyboard Hardware," Journal of the Korea Information Security and Cryptology, Vol. 18, No. 4, pp.187-194, Aug., 2008
8. Kyungroul Lee, Kangbin Yim et al., "Password Sniff by Forcing the Keyboard to Replay Scan Codes," Proceedings of JWIS 2010, PP.9-11, Aug. 2010

The Detection Technology of LTE based Stratified Fuzz

Jun Yang[1,2], Haixia Yang[1,2], Qinshu Xiao[1,2],
[1]School of Computer Science, Beijing University of Posts and Telecommunications, Beijing, China
[2]National Engineering Laboratory for Mobile Network Security, China
{yangjun, yanghaixia, xiaoqinshu}@ bupt.edu.cn

***Abstract*—Fuzz test usually used in detecting network protocol vulnerabilities, Though that common fuzz test can cover as many as testing cases, its efficiency is relatively low. It may be spend many time to detect an aspect of a protocol. For this problem the paper put forward a more efficient method based on common fuzzing test. This method is applied for LTE protocol because it is raised against the features of LTE protocol. The paper in-depth studied the structure and process of GTP protocol, and designed stratified Fuzz testing process for the detection of GTP protocol to prove that the detection technology of LTE based stratified Fuzz is feasible and more efficient compared to common Fuzzing.**

1 Introduction

In recent years, 4G network has fully penetrated into all aspects of people’s life. At the same time, the army also starting their dedicated LTE network. If the security of 4G network can’t be guaranteed[1], it will cause a serious threat to the people’s life and national security. Excavating exploitable security vulnerabilities of 4G protocol can deter hostile in military, at the same time, make our network more safely.

There are usually two vulnerability mining technology. One is analysis by researchers completely, and the other one is using formal tools. The most common formal tools is fuzzing[3][4]. Li Weiming, Huazhong University of Science and Technology, realized automated detection of fuzz test, and successfully detected many vulnerabilities of EM protocol and ISQLPlus protocol and so on, some of which has been published, but the other haven’t been detected before[5]. Zhang yajun and Li zhoujun, Beijing University of Aeronautics and Astronautics, have brought forward a distributed model for automated white-box fuzzing[6]. Feng shengbo, Beijing University of Posts and Telecommunications, posed an improved fuzzing test methods, and successfully detected the unknown RLC protocol loopholes. Though there are many example of fuzzing, many application of fuzz test are focused on Internet Protocol, particularly in authentication protocol, and research of fuzz in LTE network protocol is still relatively small[7].

On the other hand, traditional fuzzing input random case to test, which improves the comprehensiveness of testing. It is beneficial to detect somewhere researcher difficult to detect by themselves[8][9]. But this is why fuzzing has low efficiency. The improved fuzz test method present are not fully applicable for LTE

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_39

communication protocols, so this paper raise a fuzzing test method for LTE protocol from the hierarchy and priority assignment, based on characteristics of LTE protocol. It can improve the efficiency of fuzzing in mining LTE protocol vulnerability.

2 LTE Protocol Layer Architecture

There are three part will in this paragraph. The first part is the architecture of LTE protocol layer, which will introduce stratified LTE protocol in user plane. The second part will describe GTP protocol which is High-level protocol above TCP/IP. The last part will give an account of the characteristics of LTE protocol.

2.1 Architecture of LTE Protocol Layer

LTE network protocol stack is same to Internet protocol stack that they all adopt the thinking of hierarchical. LTE network protocol stack in user plane is structured as figure 1:

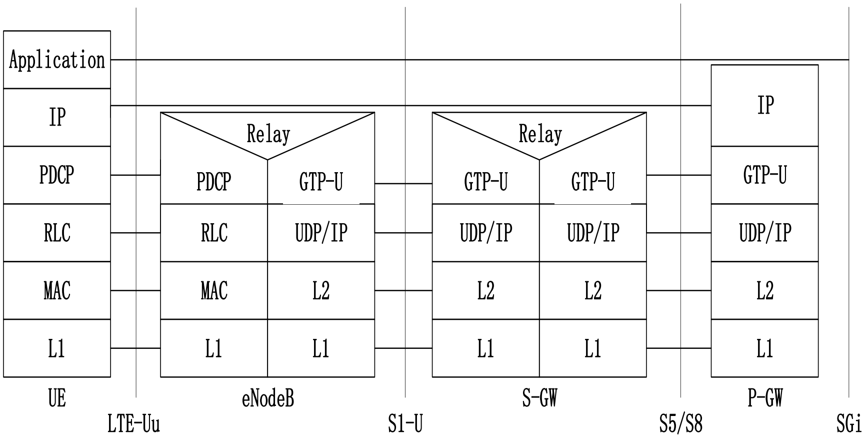


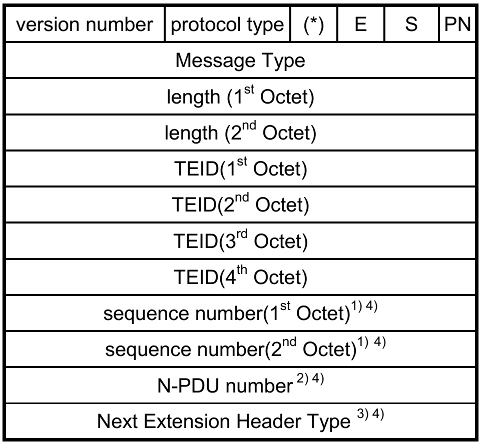
Figure 1. The architecture of LTE network protocol stack in user plane

LTE network protocols can be divided into several layers above. High-level protocols including RLC, PDCP, GTP protocols in LTE network achieve routing through IP address[2]. When transmitting data, the upper layer packets, encapsulated by lower layer protocol and marked the underlying protocol's header, pass down to the physical layer. When receiving data, the upper layer protocol receive lower layer packets to deal and remove the lower layer protocol's header, then transfer them to the upper until they arrive the destination node[10].

The above description outline the overall architecture of the LTE network protocol. For each protocol, they all has a fixed format and field. This paper will bring a brief analysis LTE protocol by an example of high-level protocol GTP.

2.2 GTP Protocol

GTP protocol is the abbreviation of GPRS tunneling protocol. As same as most tunnel technologies, GTP is a high-level protocol above TCP/IP or UDP/IP. And it is transparent for the router when providing the end to end communication between hosts. In TCP / IP protocol stack, GTP can even be understood as an application layer protocol. GTP protocol use tunnel identified TEID multiplex on the network path [11]. GTP protocol can be divided to GTP-U、 GTP-C and GTP` protocols according to different functions. GTP-U is a user-level protocol used to transmit user data, GTP-C is a control plane protocol for the management of GTP tunnel, and GTP’ is for charging.The fellowing picture is GTP protocol format.



version number	protocol type	(*)	E	S	PN
Message Type					
length (1st Octet)					
length (2nd Octet)					
TEID(1st Octet)					
TEID(2nd Octet)					
TEID(3rd Octet)					
TEID(4th Octet)					
sequence number(1st Octet)1) 4)					
sequence number(2nd Octet)1) 4)					
N-PDU number 2) 4)					
Next Extension Header Type 3) 4)					

Figure 2. GTP head thumbnail

- The version number field is used to determine the GTP protocol version. In GTP-C protocol, P-GW will throw away the message when the version number isn’t supposed by GTP-C protocol.
- Protocol type field is used to distinguish the GTP protocol and the GTP` protocol.
- Message type defines a number of message in GTP-C and GTP-U protocol to manage the path.
- TEID field clearly identifies the endpoint of the GTP-U or GTP-C protocol in the other side of the tunnel. It is used to transfer GTP packet multiplex on the tunnel between S-GW and P-GW.

2.3 The Characteristics of LTE Protocol

Based on the above analysis we can see that each protocol has its fixed protocol format, specifically in 3GPP protocol specification[2]. The different functions of each field are defined in the agreement. For example, GTP protocol version number field defines the GTP protocol version, and message category (GTP protocol or GTP 'protocol) are defined by the protocol type field and so on, but there are some fields have correlation, such as field E which defining whether the packet has extension header. If the value of field E is 0 then the following field Next Extension Header Type will not make sense. Field S defines that whether the packet use sequence and if this bit is set as zero, as the same, the following field sequence number will lose significance.

In summary, the characteristics of 4G protocol is:

- It's format is fixed, divided into different fields;
- It's fields can be relevant. Through there are also independent fields.

3 The Algorithm of Stratified Fuzzing

According to analysis for the characteristics of LTE protocol, the paper put forward a new fuzzing algorithm which use thinking of stratification and setting prioritization in the light of LTE protocol's characteristics.

3.1 Policy of Stratification

Depending on the function of each protocol field value, we can put those fields to different levels to test anew.

Specific hierarchical rules:

1) Unrelated fields, which have separate functions, not associated with other fields, can be divided into single layer;

2) Relevant fields, which values may affect whether the value of the second field is valid, are divided into a layer.

The following picture describe in detail the layering strategy of GTP protocol as example.

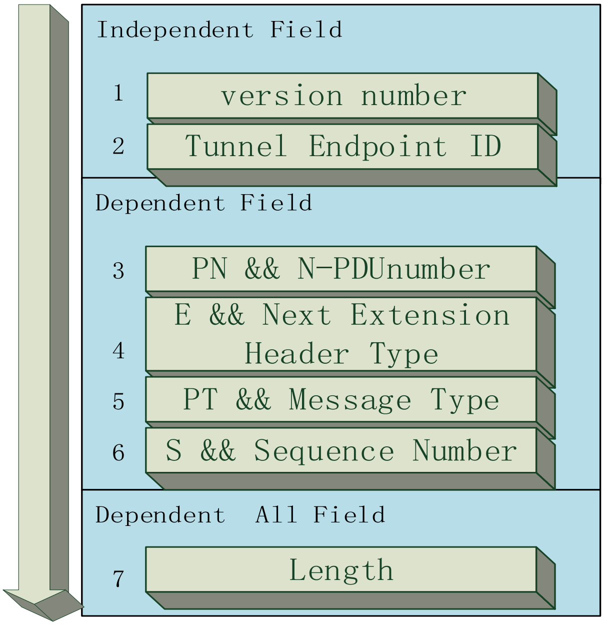


Figure 3. The layering strategy of GTP protocol

As shown as figure 2, in GTP header, version number, TEID can be divided into different independent level. Related Fields include protocol type and message type, field E and the next extension header type, field S and the sequence number, field PN and N-PDU number. The field length is related with all the field of GTP header, so we put it in a single level.

3.2 Policy of Priority

Protocol field values generally have specified range, and the protocol decode data in this range. There are may be many methods to dispose the message when it’s value beyond the range. The most method used is discarding. The method of handling error message is usually to be attacked by hackers. According to the specified range of values, we can construct the abnormal data to avoid more data made by fuzz to be abandoned. So we use a more efficient way of setting priorities to detect every layer divided in previous section.

Specific strategies are as follows:

1) For reasonable data, we mainly test the value of the data within the range, especially values in the boundary of range. For example, if the rang is 1-100, we set the value 1 or 100 to test. Many protocol have a relatively safe method in dealing with the issue of the value of the edge value, and this is easily to lead to issue of pointers cross-border and make system collapse.

2) For unreasonable data, firstly, we detect whether those data can be accepted by protocol. If not, it is unnecessary to test other values beyond the range. If the abnormal data can be accepted, we can continue our test by exhausting random.

We design the processes of fuzz test aimed to GTP protocol :

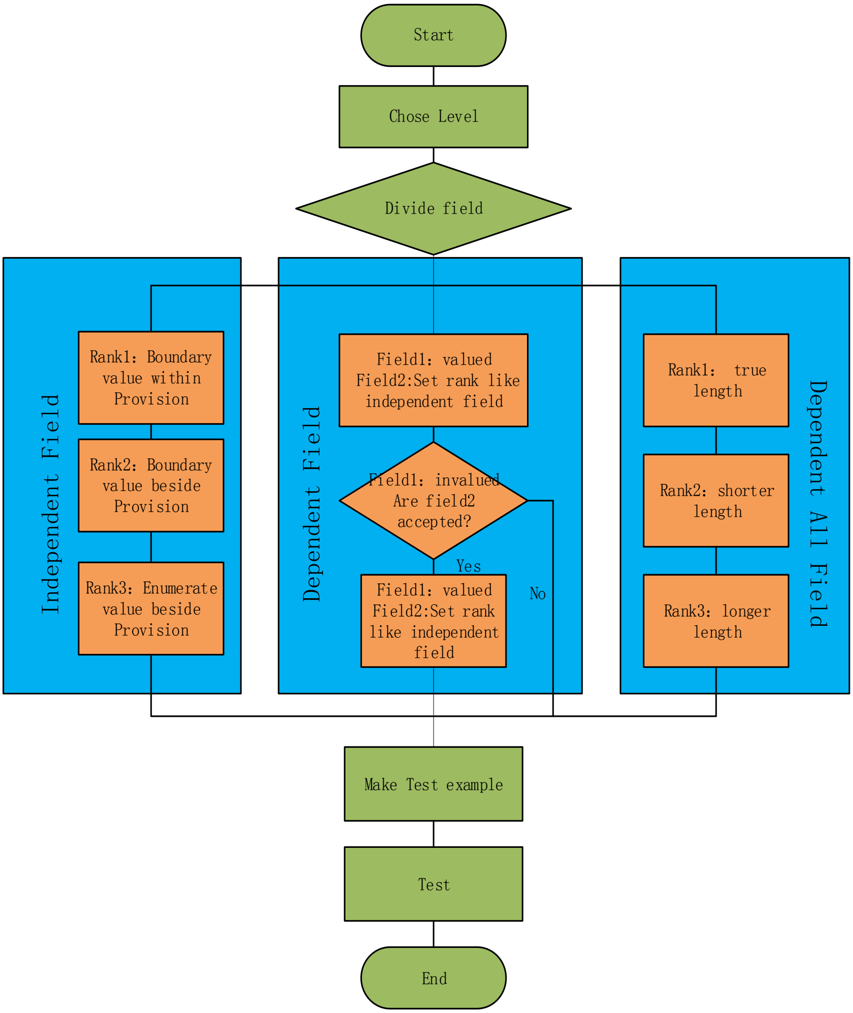

Start
Chose Level
Divide field
Independent Field
Rank1: Boundary value within Provision
Rank2: Boundary value beside Provision
Rank3: Enumerate value beside Provision
Dependent Field
Field1: valued Field2:Set rank like independent field
Field1: invalued Are field2 accepted?
Yes
No
Field1: valued Field2:Set rank like independent field
Dependent All Field
Rank1: true length
Rank2: shorter length
Rank3: longer length
Make Test example
Test
End

Figure 4. Flow diagram of fuzz test to GTP

4 Experiment

4.1 Experimental Platform

In this paper, we use NS-3 simulation environment to experiment for fuzzing method putted in the third paragraph. We will give a brief introduction of NS-3 simulation environment at first.

NS-3 is network simulator driven by discrete event, mainly used in research and education, which is designed to meet the needs of the academic and teaching. NS-3 project is a fully open source development project.

The development of the LTE module for ns-3 was carried out during the Google Summer of Code 2010. The module is built completely in C++. It comprises 89 classes and approximately 9000 lines of code. The module has been merged into ns-3.10.NS-3 including some modules, for example, network mudule， WiFi module，wimax module, LTE module and so on.

We build a custom LTE environment to test the GTP protocol based the LTE template in NS-3. Specific custom protocol stacks as follows:

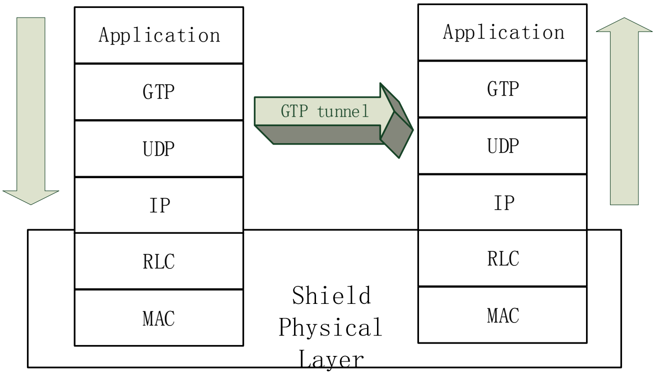


Figure 5. Custom stack of GTP

Construction of custom small-scale LTE network as follows:

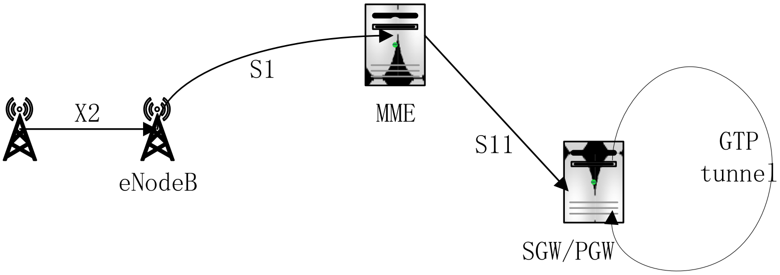
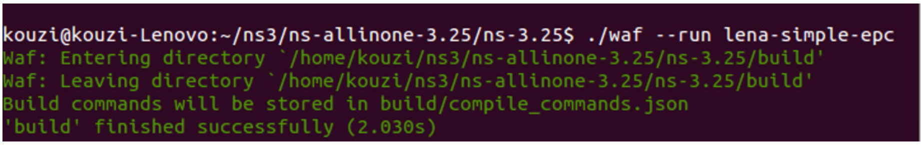


Figure 6. Custom small LTE network

As shown above, in order to simplify the network structure and avoid unnecessary costs, we combine SGW and PGW together using one node. GTP tunnels is implemented in application class nodes.

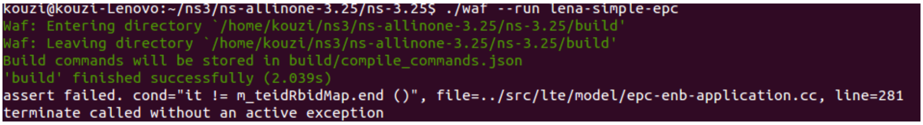
4.2 Result of Experiment

When data in custom scene is normal, the result is as follows:



```
kouzi@kouzi-Lenovo:~/ns3/ns-allinone-3.25/ns-3.25$ ./waf --run lena-simple-epc
Waf: Entering directory `/home/kouzi/ns3/ns-allinone-3.25/ns-3.25/build'
Waf: Leaving directory `/home/kouzi/ns3/ns-allinone-3.25/ns-3.25/build'
Build commands will be stored in build/compile_commands.json
'build' finished successfully (2.030s)
```

When we alter the data randomly of GTP message made by the new method of fuzzing, and then re-compile and run the virtual environment, the result is as following:



```
kouzi@kouzi-Lenovo:~/ns3/ns-allinone-3.25/ns-3.25$ ./waf --run lena-simple-epc
Waf: Entering directory `/home/kouzi/ns3/ns-allinone-3.25/ns-3.25/build'
Waf: Leaving directory `/home/kouzi/ns3/ns-allinone-3.25/ns-3.25/build'
Build commands will be stored in build/compile_commands.json
'build' finished successfully (2.039s)
assert failed. cond="it != m_teidRbidMap.end ()", file=../src/lte/model/epc-enb-application.cc, line=281
terminate called without an active exception
```

After successful compiling environment, the custom environment failed to link. As unreasonable value set in the fields, the virtual environment can’t continue to go smoothly after starting up normal.

4.3 Vulnerability Analysis

In the above case, the altered value is field protocol type whose value is 1. In 3GPP protocol specification, value 1 of protocol type mean the message is echo message. GTP protocol use echo message to maintain connectivity of the GTP tunnel, at the same time, the echo message has the effect of resetting the tunnel connection. IE (information element) of echo message can carry normal node information, as well as, can take along information of reset connection between two nodes. P-GW in the other hand of tunnel will reset the count value of GTP tunnel connection according the message and delete local content relevant context after receiving echo request with the recovery information element.

Therefore, in the echo message of GTP protocol, setting the IE as recovery information can make the GTP session flow cut off.

5 Conclusion

In this paper, we put forward a new method of fuzz test anti LTE protocol based on the LTE protocol layering strategy. This method improve fuzzing in blindness and randomness. And we test this method in NS-3 simulation environment by detecting the GTP protocol of LTE. It proved the method is feasible and efficient relatively in detecting LTE protocol’s vulnerabilities.

Acknowledgment

This work was supported by National Natural Science Foundation of China (No. 61272493，61502536，U1536122).

References

1.3GPP. TS.33.401. V.12.9.0-2013. 3GPP System Architecture Evolution (SAE): Security Architecture (Release12)

2. 3GPP TS 129.060. General packet radio service (GPRS): GPRS tunneling Protocol (GTP) across the Gn and GP interface[s].2005

3. MurphyG. Whltehouse0. Attacks and COUntS measures in 2.5 and 3G cellular IP networks[R]. Cambride MA USA:@stake.Ine.,2004

4. Bavosa. A GPRS security threats and solution recommendations[R]. Sunnyvale CA USA: Juniper Network Inc.,2004.

5. 3GPP TSG-SA2, Security analysis for tunnel establishment[s], Nortel Networks July,2003.

6. Piro G, Baldo N, Miozzo M. An LTE module for the ns-3 network simulator[C]// Proceedings of the 4th International ICST Conference on Simulation Tools and TechniquesICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2011:415- 22.

7. Wang T, Wei T, Gu G, et al. Checksum-Aware Fuzzing Combined with Dynamic Taint Analysis and Symbolic Execution[J]. Acm Transactions on Information & System Security, 2011, 14(2):613-613.

8. Cheng H F, Zhang Y Q. Bluetooth OBEX Vulnerability Discovery Technique Based on Fuzzing[J]. Computer Engineering, 2008, 34(19):151-153.

9. Gtinter Schafer, Research Challenges in Security for Next Generation Mobile Networks, Workshop on Pioneering Advanced Mobile Privacy and Security(PAMPAS), Royal Holloway University of London, Egham, Surrey, United Kingdom. September 2002.

10. Andrei Broder, Michael Mitzenmacher. Network Applications of Bloom Filters: A Survey. Internet Math. Volume 1, Number 4 (2003), pp.485-509.

11. H.N.Hung, Y.B.Lin, “Connection failure detection mechanism of UMTS charging protocol,” IEEE Transactions on Wireless Communication, vol.5, NO.5, pp.1180-1186, 2006

12. Liu Li-Fang, Huo Hong-Wei, Wang Bao-Shu. PHGA-COFFEE:Aligning multiple sequences by parallel hybrid genetic algorithm. Chinese Journal of Computers, 2006, 29(5): 727-733(in Chinese)

13.Liu Qi-Xu, Zhang Yu-Qing. TFTP vulnerability exploiting technique based on Fuzzing. Computer Enfineering, 2007, 33(20):142-144(in Chinese)

14. Makam P. Security vulnerabilities in GPRS networks[R]. Hyder-abad: India Wipro Technologies, 2006.

Secure Multi-Protocol Mapping Gateway in Smart Grid

Kim Jin Cheol[1], Park Sung Wan[1], Lee Yong Gu[1], Lee Seung Won[1]
[1] KEPCO KDN, 105Munjiro, Yuseong-gu, Daejeon, 34056, Rep. of Korea
{shine_1991, starcraft-2, yongkoo_lee.209, swlee_1201}@kdn.com

Abstract. Increasing connectivity, integration with legacy power systems, the proliferation of access points, escalating system complexity and wider use of common operating systems and platforms may contribute to increased risks for the Smart Grid. In this paper, we tried to implement the secure multi-protocol mapping gateway between digital substation IED using the MMS protocol and the SCADA using the DNP protocol.

1 Introduction

Legacy power systems were designed and installed decades ago with limited cybersecurity consideration. Increasing connectivity, integration with legacy power systems, the proliferation of access points, escalating system complexity and wider use of common operating systems and platforms may contribute to increased risks for the Smart Grid.

NERC CIP 002-009 has developed security standards for all utilities with Critical Assets, currently just for transmission, but likely to apply more broadly.[2] IEC 62351 series for utility communications include security for utility-specific protocols (IEC 61850, DNP3), role-based access control, and network and system management.[3] AMI-SEC under the UCA Users Group is addressing security issues for Advanced Metering Infrastructure.[4] IEC TC65C(in conjunction with ISA SP99) is developing security standards for industrial automation. [1]

SEL has developed two types of BITW device SEL 3021-1 and SEL 3021-2 in order to protect the serial communication data between SCADA equipment. In addition, SEL has developed 3025 as a security solution for SSCP (Secure SCADA Communication Protocol) standards in the Hallmark project. SEL 3021-1 provides only confidentiality, but, SEL 3021-2 and SEL 3025 can basically provide data integrity, and optionally provide data confidentiality.

The American Gas Association (AGA) Task Group12 designed the *SCADA Cryptographic Module (SCM)* as a BITW solution that retrofits data authenticity to SCADA communications while maintaining the performance requirements. AGA's SCM provides several cipher-suites to choose from. The most secure ones use AES-CTR for data privacy and HMAC-SHA-1/-256 for data authenticity. Unfortunately, messages must be held back by the receiving SCM using these cipher-suites. [2]

As the connection to the grid system of various shareholders such as DERs (Distributed Energy Resource), EV (Electrical Vehicle) charging stations and digital

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_40

substations, DNP used in legacy power system is mapped to the other protocols such as IEC 61850 protocols and ICCP (Inter-control Center Communication Protocol). In particular, new digital substations are support IEC 61850 protocols, IEEE specifies the standard approach for mapping between IEEE Std 1815 (DNP3) and IEC 61850. In this paper, we tried to implement the secure multi-protocol mapping gateway between digital substation IEDs (Intelligent Electronic Device) using the MMS protocol and the SCADA using the DNP protocol. We propose our DNP-MMS protocol mapping procedure and implement DNP-MMS protocol mapping procedure. We use the mapping rules based new IEEE 1815.1 standard.

2 Relevant Standardization Activities

2.1 IEC 62351 Security Standards

IEC 62351-1 to 9 is being standardized by the ISO/IEC TC 57 WG15 and defines data and communications security for power systems management and associated information exchange. It comprises security definitions for communication protocols, network and system management as well as role-based access control. Consequently IEC 62351 is constantly improved and enhanced to cope with new upcoming requirements. The current coverage of protocols is shown Fig. 1. [3]

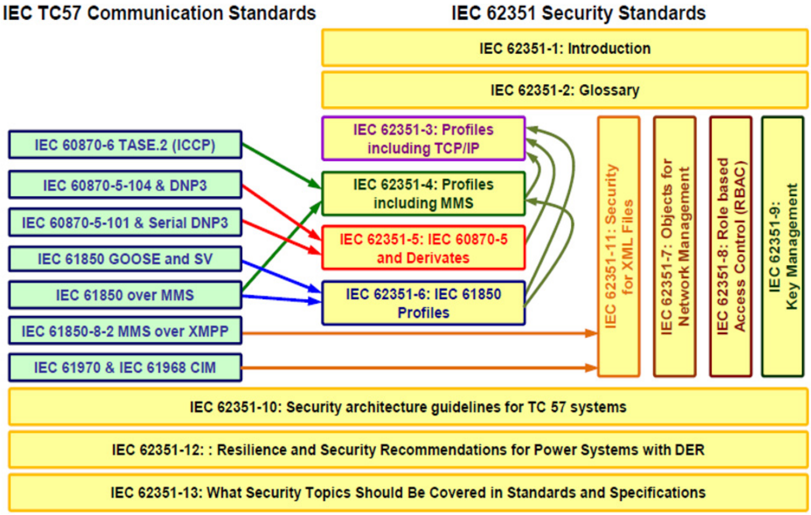

IEC TC57 Communication Standards
IEC 62351 Security Standards
IEC 62351-1: Introduction
IEC 62351-2: Glossary
IEC 60870-6 TASE.2 (ICCP)
IEC 60870-5-104 & DNP3
IEC 60870-5-101 & Serial DNP3
IEC 61850 GOOSE and SV
IEC 61850 over MMS
IEC 61850-8-2 MMS over XMPP
IEC 61970 & IEC 61968 CIM
IEC 62351-3: Profiles including TCP/IP
IEC 62351-4: Profiles including MMS
IEC 62351-5: IEC 60870-5 and Derivates
IEC 62351-6: IEC 61850 Profiles
IEC 62351-11: Security for XML Files
IEC 62351-7: Objects for Network Management
IEC 62351-8: Role based Access Control (RBAC)
IEC 62351-9: Key Management
IEC 62351-10: Security architecture guidelines for TC 57 systems
IEC 62351-12: : Resilience and Security Recommendations for Power Systems with DER
IEC 62351-13: What Security Topics Should Be Covered in Standards and Specifications

Fig. 1. IEC 62351 Series

2.2 IEEE 1815.1 Standard

IEEE 1815.1 specifies the standard approach for mapping between IEEE Std 1815 (DNP3) and IEC 61850. The objective of IEEE 1815.1 standard is to document and make available requirements for exchanging data between IEEE Std 1815 and IEC 61850 protocols using a gateway. Mapping aspects included in the standard are: conceptual architecture; general mapping requirements; the mapping of Common Data Classes, Constructed Attribute Classes and Abstract Communication Service Interface (ASCI); cyber security requirements, the architecture of a gateway used for translation and requirements for embedding mapping configuration information into IEC 61850 System Configuration Language (SCL) and DNP3 Device Profile. This specification addresses a selection of features, data classes and services of the two standards. [4]

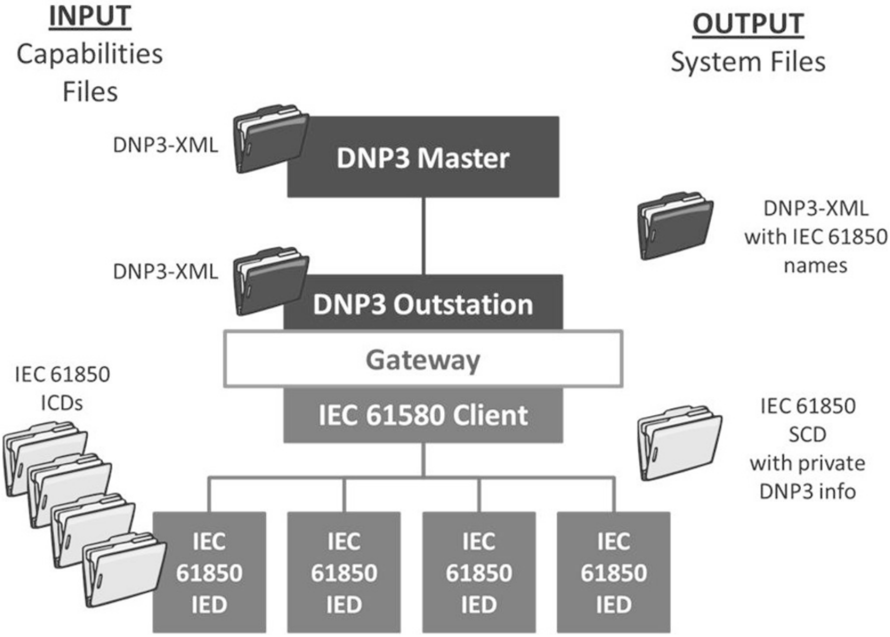


Fig. 2. IEEE 1815.1 Use case

3 Implementation of Secure Multi-Protocol Mapping Gateway

In this paper, we tried to implement the secure multi-protocol mapping gateway between digital substation IEDs (Intelligent Electronic Device) using the MMS protocol and the SCADA using the DNP protocol. We implemented a hardware security module, MCP (Many Core Platform) and secure mapping procedures in order to develop the available secure mapping gateway operating in the control network. The configuration of our secure mapping gateways that we have implemented is shown in Fig 3(a) and (b).

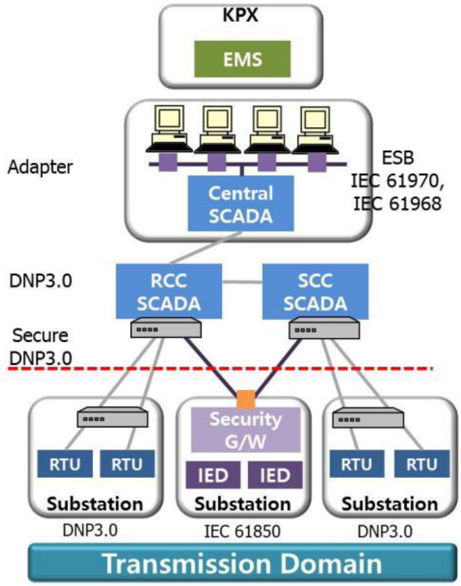


Fig. 3(a). Configuration of Secure Mapping G/W

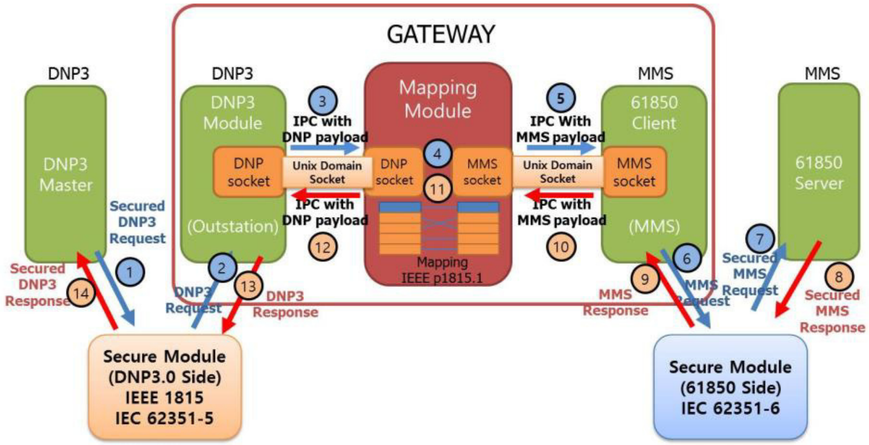


Fig. 3(b). Data flow of Secure Mapping G/W

3.1 Hardware Security Module

In this paper, we implemented a hardware security module to perform a fast encryption/decryption in the control network. The specifications of the hardware security module are shown in Table 1.

Table 1. Specification of Hardware Security Module

Type	Protection function	Remarks
Symmetric	ARIA	Mode : ECB, CBC, CFB(128), OFB(128), CTR Key Size: 128, 192, 256 bit
MAC	HMAC	Hash : SHA-256
Hash	SHA	SHA-256
Random Number Generator	CTR-DRBG	ARIA based
Digital Signature	RSA, ECDSA	ECC, RSA
Key Agreement	ECDH	ECC

The hardware security module performs a fast encryption and decryption using the ‘OPERA’ encryption chip. In addition, we make a temper resistance switch to perform the tempering resistance function and it may be convenient to analyze the status of the security module using the LED. The hardware security module is shown in Figure 4.

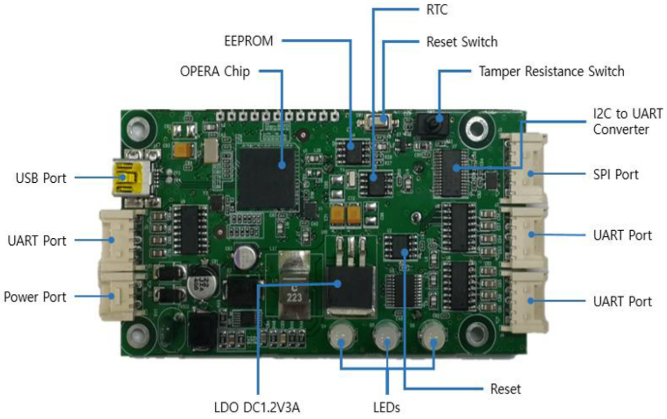


Fig. 4. Hardware Security Module

3.2 Many Core Platform (MCP)

Our security gateway uses the Many Core Platform (MCP). The MCP is a hardware platform with a 32-core CPU and the hardware cryptographic module is linked with MCP via a PCI slot. The MCP performs DNP-MMS protocol mapping process, DNP secure process and secure MMS process in control networks efficiently. The MCP is shown in Figure 5.

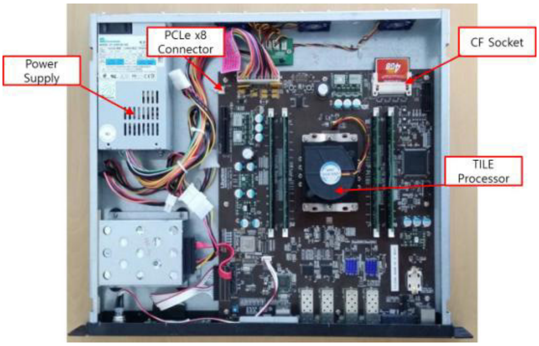


Fig. 5. Hardware Security Module

3.3 Mapping Procedure and Implementation

In this section, we propose DNP-MMS protocol mapping procedure and implement DNP-MMS protocol mapping procedure. Our DNP-MMS mapping procedure is shown in Fig 6. We use the mapping rules based new IEEE 1815.1 standard.

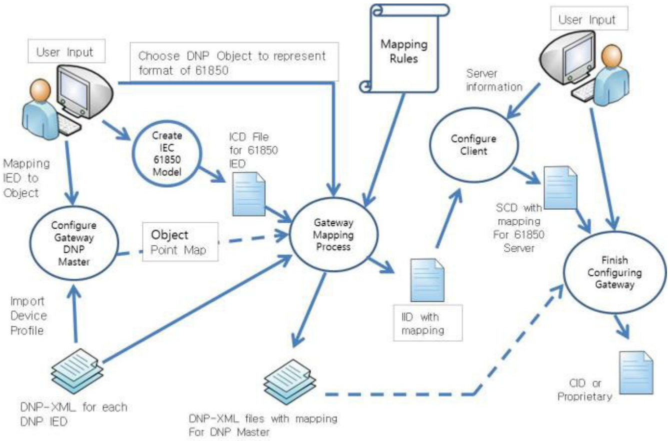
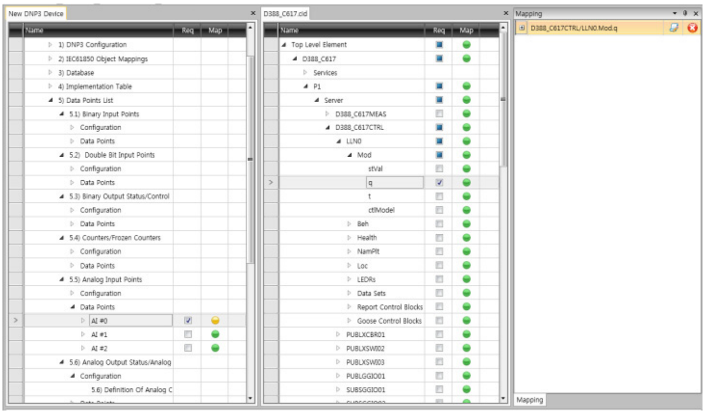


Fig. 6. DNP-MMS Mapping Process

To map DNP-MMS protocol, we first must be able to create and validate DNP XML profiles. We use the DNP3 Forge S/W to create and validate DNP XML profiles. DNP XML profile creating process is shown in Figure 7.



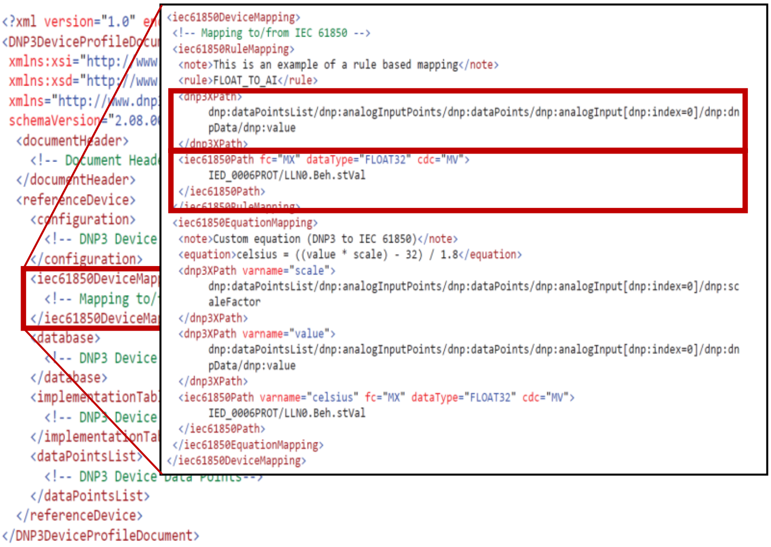
```
New DNP3 Device
Name                                    Req  Map
  ▷ 1) DNP3 Configuration
  ▷ 2) IEC61850 Object Mappings
  ▷ 3) Database
  ▷ 4) Implementation Table
  ◢ 5) Data Points List
    ◢ 5.1) Binary Input Points
      ▷ Configuration
      ▷ Data Points
    ◢ 5.2)  Double Bit Input Points
      ▷ Configuration
      ▷ Data Points
    ◢ 5.3) Binary Output Status/Control
      ▷ Configuration
      ▷ Data Points
    ◢ 5.4) Counters/Frozen Counters
      ▷ Configuration
      ▷ Data Points
    ◢ 5.5) Analog Input Points
      ▷ Configuration
      ◢ Data Points
        ▷ AI #0                         ☑    ●
        ▷ AI #1                         □    ●
        ▷ AI #2                         □    ●
    ◢ 5.6) Analog Output Status/Analog
      ◢ Configuration
          5.6) Definition Of Analog C
```

```
D388_C617.cid
Name                                    Req  Map
◢ Top Level Element                     ■    ●
  ◢ D388_C617                           ■    ●
    ▷ Services
    ◢ P1                                ■    ●
      ◢ Server                          ■    ●
        ▷ D388_C617MEAS                 □    ●
        ◢ D388_C617CTRL                 ■    ●
          ◢ LLN0                        ■    ●
            ◢ Mod                       ■    ●
                stVal                   □    ●
                q                       ☑    ●
                t                       □    ●
                ctlModel                □    ●
            ▷ Beh                       □    ●
            ▷ Health                    □    ●
            ▷ NamPlt                    □    ●
            ▷ Loc                       □    ●
            ▷ LEDRs                     □    ●
            ▷ Data Sets                 □    ●
            ▷ Report Control Blocks     □    ●
            ▷ Goose Control Blocks      □    ●
          ▷ PUBLXCBR01                  □    ●
          ▷ PUBLXSWI02                  □    ●
          ▷ PUBLXSWI03                  □    ●
          ▷ PUBLGGIO01                  □    ●
          ▷ SUBSGGIO01                  □    ●
```

```
Mapping
⊞ D388_C617CTRL/LLN0.Mod.q
Mapping
```

Fig. 7. DNP-XML profile creating process

The implemented DNP-MMS mapping procedure is shown in Fig 8. We use the mapping rules based IEEE 1815.1 standard.



```
<?xml version="1.0" en
<DNP3DeviceProfileDocu
 xmlns:xsi="http://www
 xmlns:xsd="http://www
 xmlns="http://www.dnp
 schemaVersion="2.08.0
  <documentHeader>
    <!-- Document Heade
  </documentHeader>
  <referenceDevice>
    <configuration>
      <!-- DNP3 Device
    </configuration>
    <iec61850DeviceMap
      <!-- Mapping to/
    </iec61850DeviceMa
    <database>
      <!-- DNP3 Device
    </database>
    <implementationTab
      <!-- DNP3 Device
    </implementationTa
    <dataPointsList>
      <!-- DNP3 Device Data Points-->
    </dataPointsList>
  </referenceDevice>
</DNP3DeviceProfileDocument>
```

```
<iec61850DeviceMapping>
 <!-- Mapping to/from IEC 61850 -->
 <iec61850RuleMapping>
  <note>This is an example of a rule based mapping</note>
  <rule>FLOAT_TO_AI</rule>
  <dnp3XPath>
      dnp:dataPointsList/dnp:analogInputPoints/dnp:dataPoints/dnp:analogInput[dnp:index=0]/dnp:dn
      pData/dnp:value
  </dnp3XPath>
  <iec61850Path fc="MX" dataType="FLOAT32" cdc="MV">
      IED_0006PROT/LLN0.Beh.stVal
  </iec61850Path>
 </iec61850RuleMapping>
 <iec61850EquationMapping>
  <note>Custom equation (DNP3 to IEC 61850)</note>
  <equation>celsius = ((value * scale) - 32) / 1.8</equation>
  <dnp3XPath varname="scale">
      dnp:dataPointsList/dnp:analogInputPoints/dnp:dataPoints/dnp:analogInput[dnp:index=0]/dnp:sc
      aleFactor
  </dnp3XPath>
  <dnp3XPath varname="value">
      dnp:dataPointsList/dnp:analogInputPoints/dnp:dataPoints/dnp:analogInput[dnp:index=0]/dnp:dn
      pData/dnp:value
  </dnp3XPath>
  <iec61850Path varname="celsius" fc="MX" dataType="FLOAT32" cdc="MV">
      IED_0006PROT/LLN0.Beh.stVal
  </iec61850Path>
 </iec61850EquationMapping>
</iec61850DeviceMapping>
```

Fig. 8. DNP-MMS Mapping

4 Conclusions

New capabilities for Smart Grid systems and networks, such as distributed intelligence and broadband capabilities can greatly enhance efficiency and reliability, but they may also create much new vulnerability if not deployed with the appropriate security controls. Providing security for such a large system may seem an unfathomable task, and if done incorrectly, can leave utilities open to cyber-attacks.

As the connection to the grid system of various shareholders such as DERs (Distributed Energy Resource), EV (Electrical Vehicle) charging stations and digital substations, DNP used in legacy power system is mapped to the other protocols such as IEC 61850 protocols and ICCP (Inter-control Center Communication Protocol). In particular, new digital substations are support IEC 61850 protocols, IEEE specifies the standard approach for mapping between IEEE Std 1815 (DNP3) and IEC 61850.

In this paper, we tried to implement the secure multi-protocol mapping gateway between digital substation IEDs (Intelligent Electronic Device) using the MMS protocol and the SCADA using the DNP protocol. We implemented a hardware security module, MCP (Many Core Platform) and secure mapping procedures in order to develop the available secure mapping gateway operating in the control network. We propose our DNP-MMS protocol mapping procedure and implement DNP-MMS protocol mapping procedure. We use the mapping rules based new IEEE 1815.1 standard.

Acknowledgments. This work was supported by Korea Institute of Energy Technology Evaluation and Planning (KETEP) grant funded by Korea government Ministry of Trade, Industry and Energy (No. 20131020400660)

References

1. Jincheol Kim, Jungsoo Cho, Seungwon Lee, Implementation of DNP Security in Distribution Automation System, IEEE Big Data Security 2016, pp 1393-1398
2. Jincheol Kim, Y.O. Kim, T.H. Kim. 2013. Implementation of Secure GOOSE Protocol using HSM, Applied Mechanics and Materials Vols. 260-261 (2013) pp 236-241
3. ISO-IEC 62351, Power System Management and Associated Information Exchange – Data and Communications Security Part 1-10
4. IEEE 1815.1-2015, IEEE Approved Draft Standard for Exchanging Information between IEC 61850 and IEEE Std. 1815

An Adaptive DoS Attack Mitigation Measure for Field Networks in Smart Grids

Gunee Lee, Yun-Sam Kim and Jungmin Kang

Abstract The wireless mesh networks is one of the key technologies for field device networks to maximize the effect of smart grid. However, the wireless mesh networks is exposed to DoS attack based on routing misbehavior that cause the network does not work properly. If sensed data cannot be transferred from field network to server side system, we could not get the benefits of smart grid successfully. For protecting the field networks in smart grids from DoS attack based on routing misbehavior, we propose a revised monitoring method that improve the level of security of the wireless mesh networks for smart grid's field device. We also provide the result of experiments.

1 Introduction

Smart grid is an intelligent power grid equipped with information communication technologies. With smart grid, electricity utilities could estimate electricity demand based on customer electricity usage information collected from smart meters, and then they might control peak load situation based on the estimation. Before an electricity peak load occurs, electricity utility has customer reduce electricity usage or makes customer use electricity generated by distributed electricity resource (DER) in the customer premises, which are polar voltaic over the roof, electricity storage

Gunhee Lee
National Security Research Institute, 1559 Yuseong-daero, Yuseong-gu, Daejeon 34044, Korea,
e-mail: icezzoco@nsr.re.kr

Yun-Sam Kim
National Security Research Institute, 1559 Yuseong-daero, Yuseong-gu, Daejeon 34044, Korea,
e-mail: bijak@nsr.re.kr

Jungmin Kang
National Security Research Institute, 1559 Yuseong-daero, Yuseong-gu, Daejeon 34044, Korea,
e-mail: jmkang@nsr.re.kr

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_41

and electricity vehicle. Moreover, customer could delay or bringing forward electricity usage based on the information about the peak load time came from the utility.

Moreover, for maintaining the reliability of power grids, smart grids may have a capability of situation-awareness. With the capability, smart grids could determine the current status of the power grid. It can detect a failure point in power grid near real-time, so the power company can respond to the failure in time. In order to do that, the sensors and the actuators should be deployed over the whole smart grids.

Since the smart grid deployed over the wide and open area, wireless mesh network is one of the highlighted technologies for the field network of smart grids. For example, smart meters form a wireless mesh network in order to transfer their measured value to a utility. A data concentrator connected to a smart meter mesh network collects the measured values and sends the bunch of data to a AMI head-end in a utility.

To achieve the purpose of the smart grids, collecting information from sensors (i.e. smart meters and phasor measurement units) and issuing a command to actuator (i.e. controllers of DER, controllers of the electricity loads and IEDs on protection relays) are the most important tasks. Blocking, intercepting, and dropping the transmission of information and command would be one of the most serious threats against the smart grids. Unfortunately the wireless mesh networks, which is the most spotlighted technology, has high possibility to occur those kinds of threats by intentionally launching the routing misbehaviour attacks against the sensors and the actuators.

One or more compromised node could drop entire or some packets from other smart meters, and then the smart meter network could not provide the functionality of information delivery. The loss of the energy information delivery capability could be expected to have adverse effect on various applications in smart grid, such as demand response, load control, distribution management, and so on.

To mitigate the DoS (Denial of Service) attack based on routing misbehavior, we had introduced a effective method that monitors routing behaviour of neighbour nodes in a wireless mesh network[1]. The previous method, however, has limitations in efficiency and it could not handle a gray hole attack properly. Thus, in this paper, we propose a revised method in order to improve the efficiency of the method and mitigate a gray hole attack properly. To do so, we design a method for building neighbour list and propose an algorithm that adaptively determine a threshold of packet drop ratio. The threshold is the criteria that determine whether a node is compromised by attacker or not.

The rest part of this paper is organized as follows. Section 2 presents the overview of the previous method we introduced and explains its limitations. Section 3 describes a revised and proposed method and algorithm in detail. Section 4 provides the experimental results, and Section 5 draws conclusion.

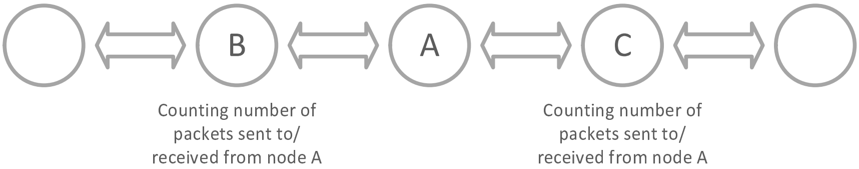

B
A
C
Counting number of packets sent to/ received from node A
Counting number of packets sent to/ received from node A

Fig. 1 Network traffic information monitoring method using cooperating nodes

2 The Previous Measure and its Limitations

Fig. 1 shows the schematic diagram of the previous method to monitor the wireless ad hoc networks and detect routing misbehavior. In this example, nodes in a wireless ad hoc network form a detection cluster. In order to monitor node *A*, neighbor nodes around node *A* collect necessary information continuously. That is, two nodes *B* and *C*, which have node *A* between them, count the number of packets that they have forwarded to node *A* and the number of packets that they have received from node *A*. Any node except node *A* in a cluster could be a watchdog node that monitors behavior of node *A*, and the watchdog node collects and analyzes the counter data from the other nodes in order to determine whether node *A* performs routing misbehavior.

In the previous method, since the watchdog node in the cluster should know its one-hop neighbors and two-hop neighbors, where the two-hop neighbor is all the nodes that could reach in two-hop routing via the monitored node. From the previous example, if the node *B* is a watchdog node, the node *C* is a two-hop neighbor of the node *B*. In the previous method, each node performs authentication process between its one-hop neighbors and two-hop neighbors in order to identify its two-hop neighbors. This process considers not only secure session key exchange between nodes but also mutual authentication between them. Thus when a new node is installed, there would be massive flood of message exchange between the new node and existing node.

In addition, it the previous method, the watchdog node calculate drop ratio of the monitored node, and it checks if the ratio is higher than a threshold. If so, the monitored node could be a attacker node. Thus the determining threshold is a key issue of the method. Higher threshold might increase false-negative ratio of the method and lower threshold might increase false-positive ratio of the method. Moreover, the threshold would be affected by the network performance as well. For example, in rainy day, the packet drop ratio could be increased, so a probability that the normal node would be identified as a attacker could be increased.

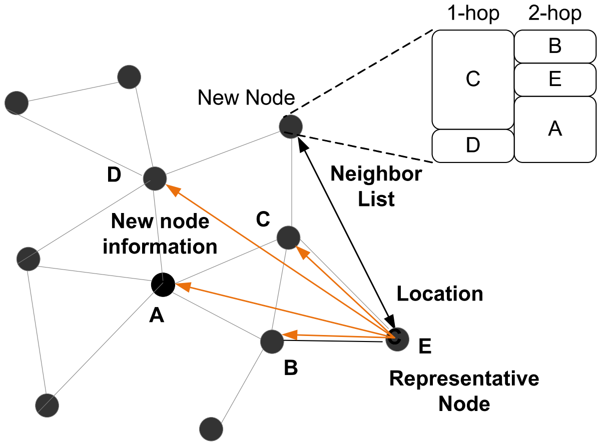


Fig. 2 The schematic view of the neighbor list distribution

3 The Proposed Method

3.1 Building a neighbor list

Fig. 2 shows the schematic view of the neighbor list distribution in the proposed method. When a new node joins a network, it should perform joining process with the representative node, and they should agree with a session key between them. The representative node is the only manager of a wireless mesh network. For example, in the advanced metering infrastructure, a data concentrator, which connects the smart meter networks and the AMI headend, can be a representative node for a smart meter network. After that, the new node should notice its transmission range and its location that is attained through the GPS or the location discovery algorithms [2, 3, 4, 5, 6]. The location is presented with a pair of X-coordination and Y-coordination. It should be encrypted with a session key, and it should be sent to the representative node.

As a new node is registered, the representative node should maintain the nodes information and the location information of the wireless mesh networks. The procedure is defined in the Algorithm 1. When the representative node receives the location information of a new node, it inserts the information of the new node into the *node information table*. The *node information table* is a list of nodes in a wireless mesh network. Each item in the list consists of 4-tuple including node's id, X-coordination of node's location, Y-coordination of node's location and transmission range. Then, it should modify the *location map* according to the information. The location map is a adjacency matrix for the wireless mesh networks. After that, with the location of the new node and the transmission range, the representative node looks up the *node information table* to find every one-hop neighbors of the new node. Based on the well known circle equation, the representative node is able to select every one-hop neighbor of the new node.

Algorithm 1 Modifying the location map of the representative node

```
$N$: a node information table
$n$: number of nodes
$M$: a location map
$L$: a list for the one-hop neighbors of the new node
$Adj$: an adjacency list for two-hop neighbors of the new node
$r$: transmission range of the new node
$(a,b)$: location information on the new node

procedure ModifyLocationMap($a$, $b$, $r$)
    $N[n] \leftarrow (id,a,b,r)$
    $n \leftarrow n+1$
    SelectOneHopNeighbors($a$, $b$, $r$, $L$)
    for $k = 0 \rightarrow n$ do
        $M[n][k] \leftarrow L[k]$
        $M[k][n] \leftarrow L[k]$

procedure SelectOneHopNeighbors($a$, $b$, $r$, $L$)
    for $i = 0 \rightarrow n-1$ do
        if $a-r < N[i].x < a+r$ then
            if $b-r < N[i].y < b+r$ then
                $L[i] \leftarrow 1$
```

The algorithm that builds a *neighbor list* for the new node is given in Algorithm 2. The *neighbor list* is a adjacency list that includes every one-hop neighbors and two-hop neighbors of the new node. Each head node of the list is a one-hop node of the new node and the other nodes of the list are the two-hop nodes of the new node. The algorithm is easy to understand. For each one-hop neighbor of the new node, it is added as a header item of the list and the one-hop neighbors of the header are added at the next of the header. The neighbor list, which is encrypted with the session key, is delivered to the new node.

Algorithm 2 Building a neighbor list for a new node

```
$N$: a node information table
$n$: number of nodes
$M$: a location map
$Adj$: an adjacency list for two-hop neighbors of the new node
$r$: transmission range of the new node
$(a,b)$: location information on the new node

procedure BuildNeighborList($id$, $a$, $b$, $r$)
    $l,k \leftarrow 0$
    for $i = 0 \rightarrow n$ do
        if $M[n][i] == 1$ then
            $Adj[l][0] \leftarrow N[i].id$
            for $j = 1 \rightarrow n$ do
                if $M[i][j] == 1$ then
                    $Adj[j][k] \leftarrow N[j].id$
```

Algorithm 3 Sending a new node information to neighbors of the new node

```
$N$: a node information table
$n$: number of nodes
$M$: a location map
$Adj$: an adjacency list for two-hop neighbors of the new node
$L1$: a list for the one-hop neighbors of the new node
$L2$: a list for the two-hop neighbors of the new node

procedure SendToOneHopNeigbors
    for $i = 0 \to n-1$ do
        if $M[n-1][i] == 1$ then
            $k \leftarrow 1$
            for $j = 0 \to n-1$ do
                if $M[n-1][j] == 1 \wedge j \neq i$ then
                    $L1[k] \leftarrow N[j].id$
                    $k \leftarrow k+1$
            $L1[0] \leftarrow N[n-1].id$
            Send $L1$ to N[i].id

procedure SendToTwoHopNeigbors
    for $i = 0 \to n-1$ do
        if $M[n-1][i] == 1$ then
            for $j = 0 \to n-1$ do
                if $M[i][j] == 1$ then
                    $L2[0] \leftarrow N[i].id$
                    $L2[1] \leftarrow N[n-1].id$
                    Send $L2$ to N[j].id
```

The Algorithm 3 provides the procedure that sends the partial neighbor list adding to the neighbor list of the new node's neighbors. According to the Algorithm 3, the representative nodes sends the new item of the adjacency list for one-hop neighbor of the new node, and it also sends the information of the relay nodes inbetween the two-hop nodes of the new node and the new node.

3.2 Monitoring and Detection of Routing Misbehavior

For detecting the misbehavior of compromised nodes, in the proposed method, we have two steps such as monitoring step and detection step as provided in the previous method[1]. In the monitoring step, each node continuously monitors their neighbors by counting the number of incoming/outgoing messages from/to monitoring node. For the detection step, any watchdog node selected as a detector should collect the counter from two-hop neighbors, which are one-hop neighbor of the monitoring node. Then the detector checks whether the ratio of the incoming and the outgoing packets is in a predefined threshold. If it is in the threshold, the monitoring node behaves normally. Otherwise, it might be compromised node.

The only difference between the previous method and the proposed method in this paer is the detector. As mentioned above, in the previous method, any watchdog node can be a detector. In the proposed method here, however, the only node that can be a detector is the representative node.

3.3 Adjustment of threshold for the packet drop ratio

The attackers can learn or estimate the threshold with his/her knowledge and practices. He/she could achieve the average drop ratio of nodes in the network, and it can select the threshold according to the average value. With the estimated threshold, the attacker can perform gray-hole attack or selective forwarding attack to the networks. Although the attack is launched, the representative node cannot notice that since the value of forwarding ratio does not exceed the threshold.

Thus, in the proposed method, we would employ an automatic threshold revision process. This process would revise the threshold H_rate after a pre-defined time duration. In a normal situation, as the number of revision would be increased, the new threshold value would be stable, so after some time the threshold value is stays at the same value. However, in a abnormal situation, the trend of threshold value's variation would be drifted from time to time. The abnormal situation includes not only attack situation but also accidental events that result in increasing the packet loss ratio.

Fig. 3 shows the schematic view of the revision algorithm. In the figure, there are two groups of forwarding ratio such as group Y and group X. Forwarding ratios of nodes in a local area are categorized into those two groups through any clustering algorithm such as K-means, Fuzzy C-means, Hierarchical clustering, and Mixture of Gaussians [7]. The proposed method will calculate a new threshold H'_{rate} in order to increase the accuracy of the proposed framework. For generating a new threshold, the representative node first calculates the median value of two values such as old threshold H_{rate} and the mean value of group Y as given in equation (1) and (2). Then, it calculates average of deviations that are differences between each value of group X and the median value M according to the equation (3). Finally the new threshold is the sum of the median value M and the average of deviations that are average of differences between each value of group X and M as shown in equation (4).

$$M = Median[H_{rate}, \overline{Y}] \tag{1}$$

$$Y = \{r_d(y) | y \in Y\}, \text{where } r_d(y) = \frac{M_f(y)}{M_r(y)} \tag{2}$$

$$V = \{r_d(x) - M | x \in X, r_d(x) > M\}, \text{where } r_d(x) = \frac{M_f(x)}{M_r(x)} \tag{3}$$

$$H'_{rate} = M + \overline{V} \tag{4}$$

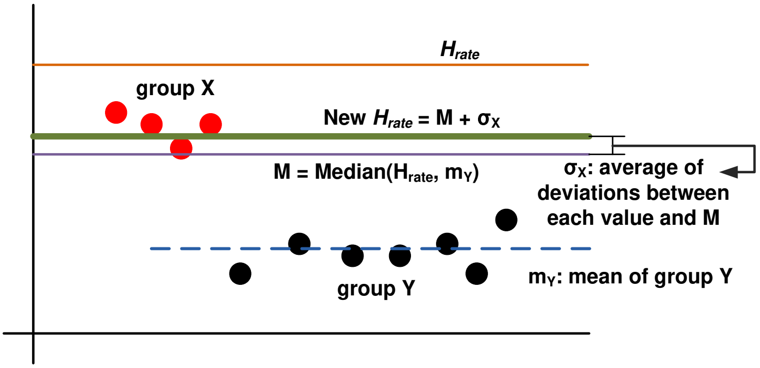

H_{rate}
group X
New H_{rate} = M + σ_X
M = Median(H_{rate}, m_Y)
σ_X: average of deviations between each value and M
m_Y: mean of group Y
group Y

Fig. 3 Adjusting threshold value

4 Experiment results

In order to evaluate the proposed method, we conducted an experiment using network simulator NS-2. In the experiment, an attack that paralyzes network service by filtering out network packets at compromised nodes in the network was launched against the network without any protection method. The same attack was launched against the network employing the proposed method.

In the experiment, 50 smart meters were installed in an area of 1500m300m. Network routing was performed by the DSR (dynamic source routing) method. Twenty of the nodes were CBR sources sending 4 packets per second. In each session, 0 10 attackers were distributed in the network. When communication was active in the network, each attacker dropped a certain percentage of network packets so that network service would be discontinued. Details of the simulation parameters are given in Table 1.

Table 1 Parameters for the experiments

Parameters	Values
Number of nodes	50
Area size (m)	1500 × 300
Traffic model	CBR
Transmission rate (*packets/s*)	4
Maximum number of connections	20
Packet size (byte)	512
Duration (s)	900

Fig. 4 shows the successful intrusion detection ratio and the false intrusion detection ratio of the proposed system according to the number of attacker nodes. The detection ratio increased slightly with the increase in the number of attacker nodes, and was over 97% even when the number of attacker nodes was smallest. This also

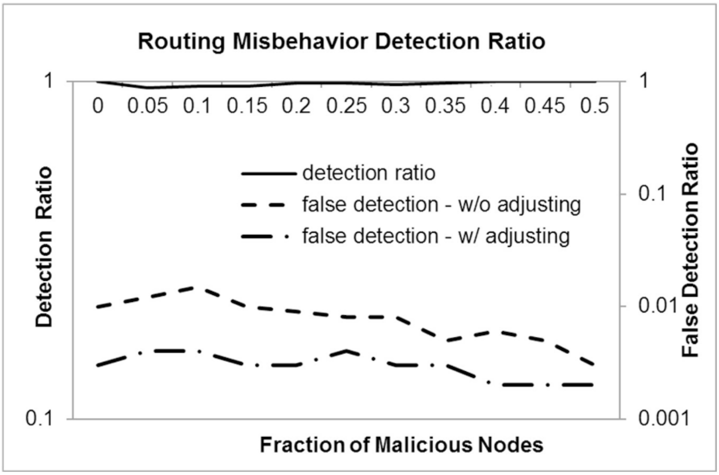


Fig. 4 Changes in the intrusion detection ratio

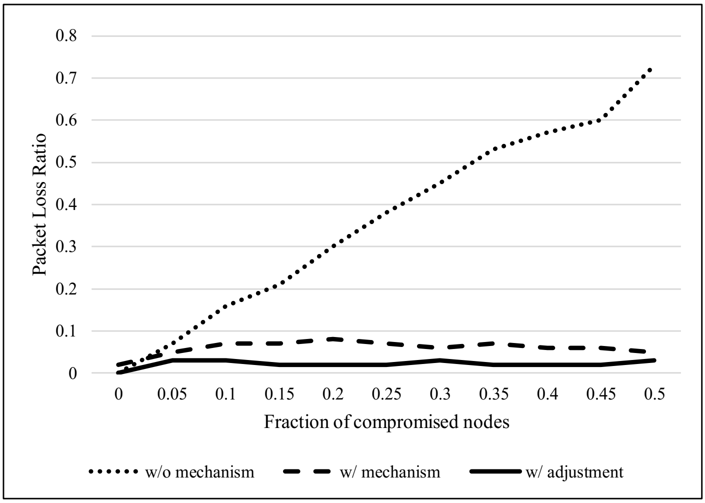


Fig. 5 Change in the packet loss ratio according to the number of attackers

means that false-negative is less than 3%. In the Figure 15, the false detection ratio means false positive ratio to total number of attack detection. When there was no attacker node, false detection ratio was 0.3%. Regardless of the number of attacker nodes, false detection ratio was less than 1%. This suggests that the possibility of false-negative or false-positive is low enough to apply the proposed method to the real environment.

Fig. 5 shows the packet loss ratio according to the change of the number of attacker nodes. When the proposed approach was not employed, the packet loss ratio went up to 73% with the increase in the number of attacker nodes. When the proposed detection method was applied, however, the packet loss ratio was stable within 3%. This ratio includes packet loss caused by attackers and that caused by processing delay resulting from the execution of the routing misbehavior detection process.

5 Conclusion

In this paper, we proposed a revised monitoring method that can detect DoS attacks based on routing misbehavior. In contrast with the previous method we published before[1], the proposed method would remove a burden of authentication process. In addition, the proposed method could detect more intelligent routing misbehavior. The results of experiment in DSR-based network environment showed that the proposed method could prove the efficiency and effectiveness.

Acknowledgements This work was supported by the the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (No. 20141010501870)

References

1. B. Hofmann-Wellenhof, H. Lichtenegger, and J. Collins, Global Positioning System: Theory and Practice, In the Fourth Edition, Springer-Verlag (1997)
2. L. Hu and D. Evans, Localization for Mobile Sensor Networks, In Proc. of ACM MOBICOM, pp. 45-57 (2004)
3. D. Liu, P. Ning, and W. Du, Detecting Malicious Beacon Nodes for Secure Location Discovery in Wireless Sensor Networks, In Proc. of the 25th International Conference on Distributed Computer Systems (ICDCS), pp. 609-619 (2005)
4. W. Du, L. Fang, P. Ning, LAD: Localization Anomaly Detection for Wireless Sensor Networks, In the Journal of Parallel and Distributed Computing (JPDC), Vol. 66, No. 7, pp. 874-886 (2006)
5. N. Sastry, U. Shankar, and D. Wagner, Secure verification of location claims, In Proc. of ACM Workshop on Wireless Security (WiSe), pp. 1-10 (2003)
6. C. H. Romesburg, Cluster Analysis for Researchers, (2004)

Analysis on Attack Scenarios and Countermeasures for Self-driving Car and Its Infrastructures

Dohyun Lim[1], Kitaek Park[1], Dongjun Choi[1], Jungtaek Seo[1] *
1 Department of Information Security Engineering, Soonchunhyang University, Korea
lingdoz@gmail.com, ir0nykt@gmail.com, zzczzc123@naver.com,
sjtgood7@gmail.com

Abstract. Autonomous vehicles collect and process information required for driving autonomously and apply the processed result for vehicle driving thereby driving vehicles automatically by identifying road situations without additional control of brakes, handling, and acceleration pedal by drivers. Since self-driving cars collect information from various sensors and communication is done between various sensors and devices over the infrastructure, they are vulnerable to unexpected accidents due to malicious hacking attacks if self-driving cars are commercialized without ensuring security technologies and establishment of security systems. In particular, securing safety is highly emphasized due to automobile technology that is directly related to human lives and safety. In the present paper, security threats against self-driving cars and infrastructures are analyzed and possible attack scenarios are developed to predict the impact. Furthermore, the current status on research and development of security technology is analyzed and items of technology development to ensure cyber security of self-driving cars and infrastructures as well as R&D strategies are presented for future research.

1 Introduction

The automobile industry has been evolved rapidly into self-driving cars where the state of the art technologies are concentrated. According to Navigant Research, a market scale of self-driving car system in the world will reach $189 billion by 2020 and $1,152 billion by 2035, which indicates that the era of self-driving car and technology are emerging [1]. To keep pace with the global interest, South Korea disclosed a goal of commercialization of self-driving cars at Level 3 by 2020 [2]. The number of deaths by vehicle accidents around the world amounts to 1,240,000 annually and 90% of the deaths are due to driver's faults such as driving while drowsy or drinking driving, which emphasizes the importance of autonomous driving technology positively more and more [1]. However, much attention has also been paid to risk of accidents due to malicious hacking attacks in contrast with the positive prospect on self-driving cars. If autonomous driving technology is commercialized

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_42

without ensuring establishing security systems and techniques, unpredictable accidents due to malicious hacking attacks can occur and enormous disruption on the network can occur during cyber-attacks on V2X. Autonomous driving needs mutual information collection through sensors in order to support V2X (InV: In Vehicle, V2V: Vehicle to Vehicle, V2I: Vehicle to Infrastructure) communication and Advanced Driver Assistance Systems (ADAS). First, it is important to collect correct information through sensors and communicate collected information securely between Electronic Control Units (ECU) over the InV communication environment. If sensors that support the ADAS do not collect correct information or if fabrication and modification of collected information occurs during communication between ECUs to employ the collected information, normal operation of autonomous driving systems cannot be achieved [3]. Second, since V2V and V2I communication environments are Wireless Local Area Network (WLAN) environments that use Wireless Access in Vehicle Environments(WAVE), the communication environments are vulnerable to attacks such as spoofing, Denial of Services (DoS), and Man In The Middle Attack (MITM). Thus, it is essential to secure the security of communication in V2V and V2I, as well as a gateway to server. Accordingly, vulnerability of V2X and autonomous driving technologies and factors of security threat are needed to be analyzed clearly and it is necessary to ensure cyber security over the autonomous driving system in preparation of malicious hacking attacks against autonomous driving systems in the future [4]. To provide a countermeasure, it is essential to analyze possible attack scenarios and impact effect on autonomous driving technology. Thus, this paper investigates the current status of technological development on self-driving cars and infrastructure in Section 2 and presents security threats against autonomous driving technology in Section 3. In Section 4, development and analysis on the countermeasure technology are presented focusing on possible attack scenarios and impact. In Section 5, the current status of development on national and international countermeasure technologies against related attacks is discussed. In Section 6, items of technological development and R&D strategies are presented to secure self-driving cars and cyber security over the infrastructure of self-driving cars in the future and in Section 7, conclusions of the present paper including future research are presented.

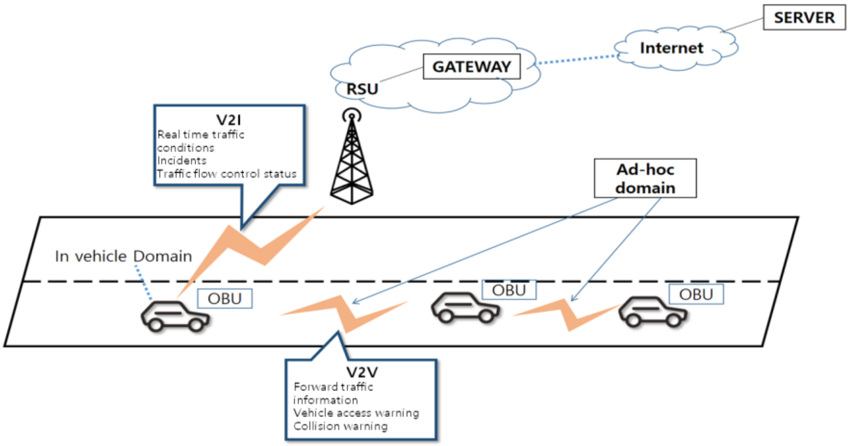


Fig. 1. Conceptual diagram of the self-driving its Infrastrucrue

2 Current status of self-driving cars and infrastructure

2.1 Current status of technological development in South Korea

2.1.1 Current status of technological development in self-driving cars

Hyundai and Kia Motors in South Korea demonstrated the first self-driving car called "Tucson IX" in 2010. Tucson IX demonstrated autonomous driving over nine missions consisting of paved and non-paved roads of 4 km including checkpoints, cross-road and accident-prone location successfully and "Genesis Smart Sense (GSS)", which was a Highway Driving Assist (HDA) system, was launched in the name of Genesis EQ900 in December 2015. The GSS can detect accident occurrence in advance through driving assist technology to help drivers to drive safely and comfortably. Based on the above technologies, the GSS acquired license of autonomous driving in the highway in Nevada State in the USA and accelerated the commercialization of self-driving cars [5].

2.1.2 Cooperative driving technology

As Cooperative Adaptive Cruise Control (CACC) and Automated Queue Assistance (AQA) technologies are developed around the world, the standardization has been underway in South Korea as well. The Telecommunications Technology Association in Korea proposed TTAK.KO-06.0379, which was a message standard at the application level for formation and separation of vehicle platooning during automated queue assist driving. The message standard was made basically by referring the J2735 BSM message proposed by the Federation Internationale de l'Automobile in the USA.

2.2 Current status of technological development in the USA

Vehicles of Ford Motor Company have been equipped with lane keeping system and active park assist in addition to semi-self-driving cars, which are currently commercialized by the CES and research is underway to provide traffic jam assist function for reduction in traffic jam by adjusting a distance between vehicles automatically. Furthermore, the company is now developing universal service including infotainment and vehicle theft prevention system in collaboration with Intel Company [7].

The USA has also started Intelligent Transport System (ITS) related projects as a measure to resolve traffic delay and congestion as well as traffic accidents such as deviation from the lane. Starting from Vehicle Infrastructure Integration (VII) in 2003, the USA changed the VII project into Connected Vehicle project in 2011 to achieve commercialization of Cooperative-ITS(C-ITS) and accelerated R&D on the effect and issue of the connected vehicle technology and commercialization. In the "Safety Pilot" among the project tasks, road tests about various C-ITS services were conducted including interconnection technology test between vehicles over the real driving environment [8].

2.3 Current status of technological development in Europe

The research in Europe has focused on core technology, road safety, and road operation. First, it developed Cooperative Vehicle-Infrastructure Systems (CVIS) from the management viewpoint such as traffic management in cities and open standard-based communication between V2V and V2I for the purpose of traffic stability and efficiency. Second, it developed the safety system (SAFESPOT) that can detect dangerous situations in the road in advance. Finally, it conducted a project for efficient road operation by providing real-time information of specific regions through implementation of two-way wireless communication environment [9].

3 Security threats against self-driving cars and infrastructure

3.1 Security threats against self-driving cars

3.1.1 Physical vulnerability

The Communication Control Unit (CCU) is a unit that can manage multimedia systems inside vehicles by enabling communication with external devices via cellular networks. Since the CCU is not directly connected to the Control Area Network (CAN) bus and an air gap is present between connection and physical parts, communication can be done with other connected components if V850 controller is used thereby being able to control the CAN bus by sending commands through reprogramming.

3.1.2 Encryption and authentication

Since the CAN protocol is suitable to networks for vehicles where real-time property is important, it has been used as a standard for most vehicles among various vehicle control protocols. It is implemented via internal network of various types of vehicles. It does not provide encryption and authentication features although it is a broadcasting communication protocol. In 2010, a research team led by K. Koscher attempted a hacking test using real vehicles and pointed out the problem of internal network in vehicles to show that vehicles can be controlled via replay attack of messages [10].

3.1.3 Access control

The ODB-II port is used to diagnose failure of engines and maintenance of vehicles. Attackers can access the CAN bus by connecting to the ODB-II port and replay attack or DoS attack can be done through packet analysis.

3.2 Security threats against self-driving car infrastructure

3.2.1 Network scale

When vehicles with different specifications and vehicle purpose are driving in the same road in the same region, interference between sensors mounted at different vehicle models, electromagnetic interference occurred at vehicles, and disruption of networks for vehicles can occur.

3.2.2 Operation of devices and systems

A path where attackers can penetrate into infrastructures of self-driving cars can be found due to inappropriate security setup, use of unauthorized mobile storage media, and inappropriate security audit in terms of operation and management of devices and system.

3.2.3 Denial of Service

Attackers can execute an attack of DoS against networks such as the Road Side Unit (RSU) via generation of a large amount of traffic, attempts of multiple accesses, and vulnerabilities to paralyze the traffic control center.

4 Possible attack scenarios and analysis on the impact

4.1 Intrusion scenarios through networks

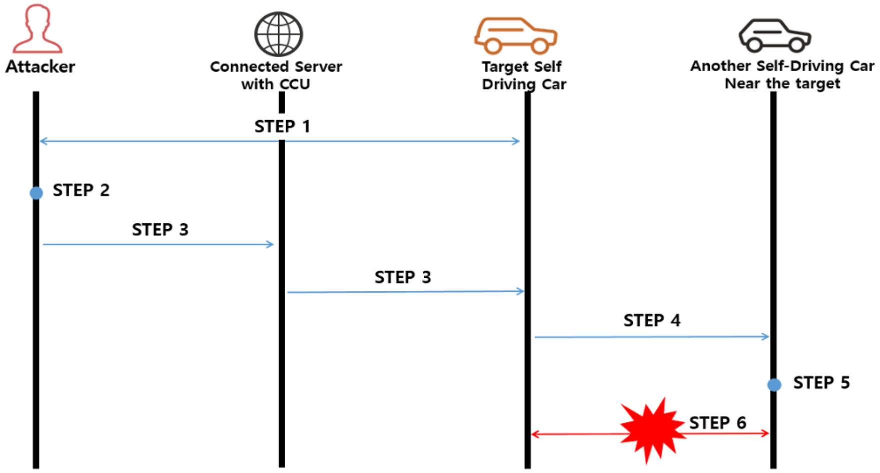


Fig. 2. Attacks on internal communication basis of vehicles

*Precondition 1: Encryption communication is not done in the inside of the CAN bus.
*Precondition 2: There is a vulnerability inside the CCU communication server.

Step 1. Attacker choose a target self-driving car and secures the IP information of the CCU system connected to the Internet through cellular networks.
Step 2. Security vulnerabilities are identified through Internet scanning and remote executable code is developed.
Step 3. The CAN bus communication is analyzed and then manipulation packets related to vehicle driving information is fabricated thereby sending the information through remote executable code.
Step 4. The target self-driving car recognizes the current car speed and acceleration as modified values, and sends Vehicle Safety Communications (VSC) message
Step 5. Another self-driving car that received the VSC message from the infected vehicle recognizes speed and acceleration of the infected self-driving car as modified values.
Step 6. Car collision is induced through re-adjustment of vehicle distance based on the modified value.

Attacker secures the IP address by which the CCU system is connected to the external cellular network and then security vulnerability is investigated through port scanning on the corresponding IP. Based on the identified vulnerability, attacker configures a backdoor to send remote executable command to the CAN bus and produce a manipulation packet related to vehicle driving information in order to send the modified packet through the backdoor. The modified packet is broadcasted through the CAN bus thereby letting the target vehicle recognize the modified driving information as normal value. Here, the modified driving information is included in the VSC that is sent between self-driving cars in real time prior to broadcasting and adjacent self-driving car receives the corresponding information. The adjacent self-driving car performs re-adjustment of vehicle distance via the modified received information, which can be followed by collision or multiple car crashes.

4.2 Intrusion scenarios through physical access

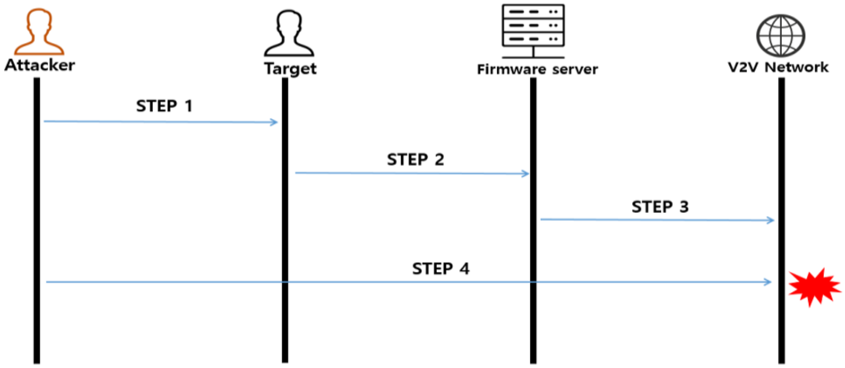


Fig. 3. Attack through firmware update

*Precondition: Penetration process into local networks is not considered.

Step 1. Malicious code is infected through spear phishing or USB thereby inducing malicious behavior desired by attacker.
Step 2. Firmware server manager opens emails or inserts infected USB thereby infecting the firmware server with malicious code.
Step 3. Attacker overwrites the correct firmware with malicious firmware and vehicles updated with malicious firmware are all infected.
Step 4. Vehicles can experience damage due to out-of-control and chain collision according to the command by attacker.

Attacker performs Advanced Persistent Threat (APT) against firmware server in the infrastructure thereby overwriting the update server with malicious firmware. When vehicles update firmware from the update server, those vehicles are infected with malicious code thereby experiencing out of control or vehicle accidents resulting in a large scale of casualties.

4.3 IoT control attack scenario in the home network using V2I communication spoofing

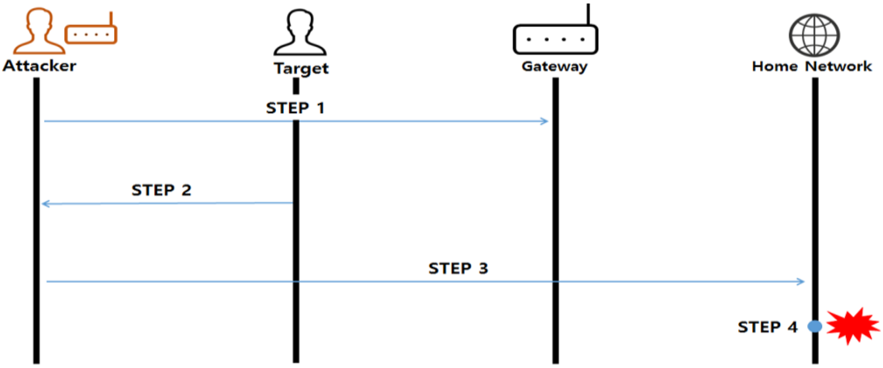


Fig. 4. IoT control in the home network using V2I communication spoofing

*Precondition: Internal authentication and connection of Internet of Things (IoT) in the home network can be done through vehicle and gateway communication data.

Step 1. Self-driving car is camouflaged as gateway through spoofing in the ad-hoc network. The ad-hoc network is characterized by no Access Point (AP) so if communication is needed externally inside the Network Address Translation (NAT), a single node is used as AP to communicate externally.
Here, ARP spoofing is conducted by attacker to be disguised as if attacker were a gateway.
Step 2. Attacker collects information about victim vehicle while being disguised as a gateway. Data by which home network can be connected are extracted from data transferred from victim vehicle to gateway.
Step 3. Through the collected information, home network is connected. Through the collected information, home network is connected and electronic appliances such as gas range, lights, air-conditioning, and personal computer can be controlled.

Step 4. Through the control of electronic appliances, financial damage, privacy information leakage, excessive electric bill charge and fire due to overheating, privacy information leakage through remote connection to PC can occur.

The communication between self-driving car and infrastructure can be done via the Vehicle Ad-hoc Network (VANET). If attacker is within the same network, he/she scans neighbor IPs and selects a victim vehicle thereby spoofing security message containing location, acceleration, and current speed that are transferred to the gateway by victim vehicle after sending its own MAC address. By ARP spoofing the gateway connected to the self-driving car, attacker becomes a gateway and extracts data by which home network can be connected through data sent by victim vehicle. Later, attacker is connected to home network through information extracted by connecting to the external Internet network and controls electronic appliances at home remotely.

4.4 V2V communication paralysis through OCSP server attack at infrastructure data transmission and reception environments

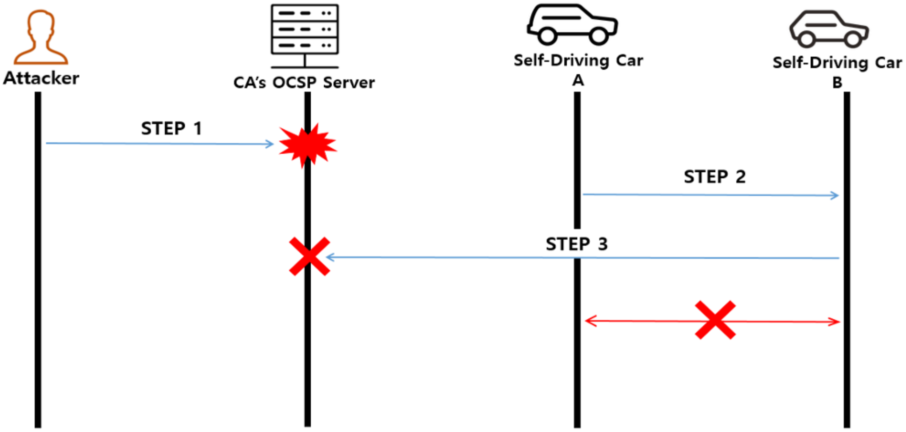


Fig. 5. OCSP server attack over infrastructure data transmission and reception environment

*Precondition 1: Certificate-based authentication environment is already constructed.
*Precondition 2: Attacker has information about Online Certificate Status Protocol (OCSP) server and access path was already acquired.

Step 1. Attacker disables services of OCSP server via various attacks.
Step 2. Self-driving car A signs in data that are transmitted to self-driving car B with private key included in the certificate embedded at the time of manufacture or issued via the Certificate Authority (CA) previously prior to transmission of data.
Step 3. In order to verify data signature received from self-driving car A, a request of validity verification is sent to the OCSP server.
Step 4. Since the OCSP server becomes disabled service status in ***Step 1,*** validity verification cannot be done.

Attacker attempts Distributed Denial of Service (D-DoS) attack through infection of OCSP client application services via malicious code or infection of USB of administrator through spear phishing against OCSP manager to infect OCSP server during maintenance resulting in making services disabled. Accordingly, it is impossible to verify signatures included in data such as accident information during driving and location information that are transferred via V2V communication. Furthermore, certificate is applied not only to self-driving car but also to the RSU thereby making authentication impossible at overall self-driving car environments. Since authentication is not possible, road traffic paralysis or human casualties upon attacking during driving could occur.

4.5 Scenario of communication jamming attack inside V2V environments

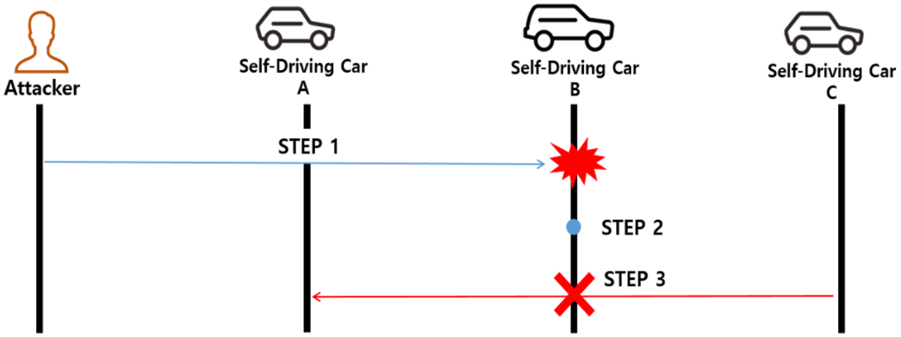


Fig. 6. OCSP server attack over infrastructure data transmission and reception environment

*Precondition: Attacker acquires access right to the CAN network.

Step 1. Attacker gains control of the CAN network inside the target vehicle thereby acquiring access right of the routing table.
Step 2. Attacker changes the routing protocol intentionally.
Step 3. Attacker drops safety message transmitted to surrounding vehicles thereby interrupting the movement to the destination node.

Once attacker gains control of the CAN network inside the target vehicle and acquires access right of the routing table, he/she can control the routing protocol directly. A self-driving car needs control of parameters such as speed by determining situations of surrounding vehicles. Here, messages are exchanged between surrounding vehicles, which are safety messages. If safety message is dropped, it can affect Packet Delivery Rate (PDR) between V2V commun ications and surrounding vehicles can have communication jamming thereby unable to recognize a distance between vehicles at high speed driving situation, resulting in vulnerable human casualties.

4.6 Scenario of vehicle control attack using control App of self-driving car

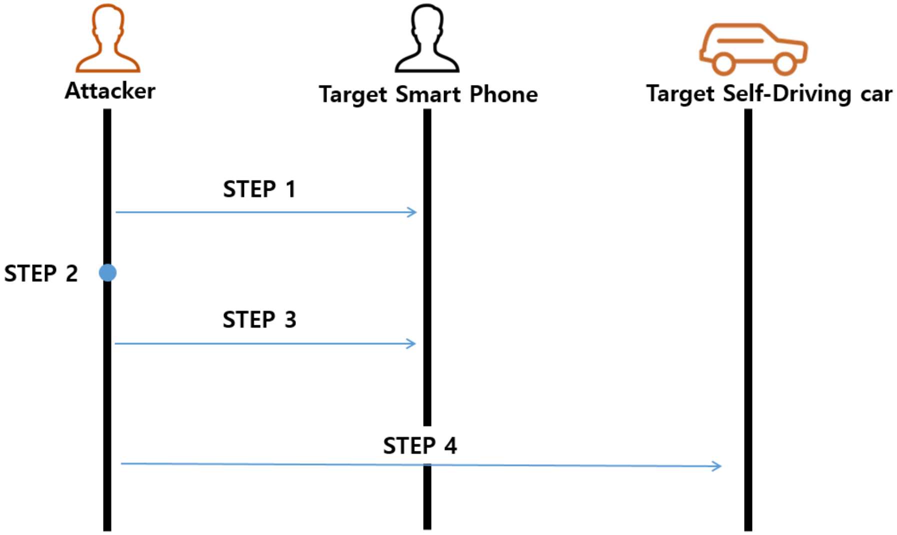


Fig. 7. Vehicle control using control App of self-driving car

*Precondition: Nonce values between message transmission and reception are not applied.

Step 1. A target smartphone is infected by malicious code through SMS phising letters sent by attacker thereby forcing rooting of the smartphone.
Step 2. Attacker acquires root right of the smartphone infected by malicious code thereby installing a rootkit.
Step 3. Attacker monitors control App packet of self-driving car using the rootkit.
Step 4. Based on the acquired information, replay attack is done to steal the vehicle.

A target smartphone is induced to be infected by malicious code through SMS phishing letters sent by attacker. Then, attacker acquires root right and installs a rootkit thereby monitoring control App packets using network packet monitoring tool. After this, attacker retransmits packets through replay attack to steal the vehicle.

5 Current status of countermeasure technology development

5.1 Privacy protective authentication technology

The Electronics and Telecommunications Research Institute in South Korea (ETRI) utilized zero-knowledge proof and encryption techniques for digital signature generation, validation, signature verification, and connection algorithm design. The privacy protective authentication technology can prevent possibility that can easily track driving information of specific vehicle by attacker if no privacy protective measure is provided in vehicle communication. The techniques can be utilized as advanced information exposure control method because it can prove knowledge, right, and qualification justification only without exposure of detailed identification information of users [11].

5.2 Security technology of V2X service integration for self-driving cars

The government in South Korea started a project called "security technology development for V2X service integration for self-driving cars", which was launched in early 2016. In the above vehicular project, Public Key Infrastructure (PKI) for vehicle privacy and information protection and reliability guarantee technology of V2V and V2I communication service for infrastructure technology and self-driving cars, V2N security technology for prevention of hacking remotely and introduction of malicious code against vehicles, and development of remote security update technology for vehicles are performed. Furthermore, it also performs international standardization of security technology for self-driving cars and applicability test and verification of security technology at self-driving environments [12].

5.3 ProtectivX technology

ProtectivX which is a platform for self-driving cars and now under development by Intel and BMW is a platform technology to prevent unauthorized accesses to information systems and IoT networks. The development is now underway to monitor all ECUs that have suspicious activities while residing in the CAN bus inside vehicles. It aims to protect the CAN bus from scan and external threats continuously such as ECU infotainment system, vision safety device, cruise control, electronic key, and remote engine starter [13].

5.4 PRESERVE technology

The PRESERVE technology was designed to prevent the abuse and infringement of privacy information during V2X communication as it summarized projects such as SEVECOM, EVITA, OVERSEE, and PRECISA as shown in Fig. 8 and included other security requirements to complete security and universal applicability tests. This technology is now at the pre-distribution stage after strengthening scalability and reducing the cost more than the previous project [14].

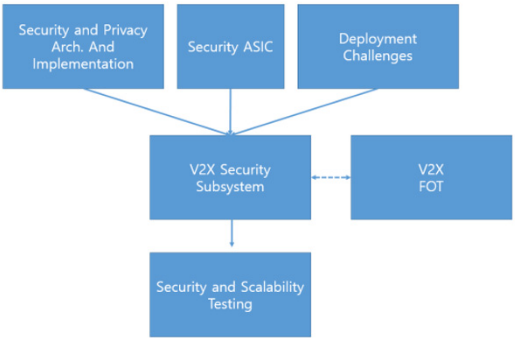

Security and Privacy Arch. And Implementation
Security ASIC
Deployment Challenges
V2X Security Subsystem
V2X FOT
Security and Scalability Testing

Fig. 8. Configuration of PRESERVE

6 Direction and strategy of future technological development

6.1 Implementation of certificate-based certification system

Certification is essential between self-driving car and infrastructure system. For certification between self-driving car and infrastructure, certificate and digital signature issued from the CA can be used. However, identification information is contained in certificate so that moving trajectory and driving time of vehicle can be identified. That is, there is an infringement on privacy information about vehicle location. Thus, it is necessary to have a certification method to protect privacy information that employs anonymous certificates that can replace identification information included in the certificate with non-identifiable anonymous value. Furthermore, it is also necessary to consider a method that requires certification only for a specific section when a vehicle enters that section.

6.2 IDS/IPS embedded with aulighthentication and lightweight encryption algorithm

The CAN bus employs a broadcasting mode that supports data communication in real time inside vehicles but it does not design and implement certification and encryption functions with respect to transmission messages. Thus, it is necessary for the CAN bus to apply certification and lightweight encryption in order to prevent hacking attacks such as packet sniffing or command injection. It is also necessary to apply detection method of white-list mode against hacking attacks that can occur inside vehicles and develop and apply Intrusion Detection System (IDS) and Intrusion Prevention System (IPS) in order to provide real time detection and response by setting a threshold of data value.

6.3 Ultra-lightweight hardware security module for key management and internal encryption operation

The Hardware Security Module (HSM) is not only used in integrity protection of ECU firmware but also in secure flashing, secure boot, run-time tuning detection, and secure debug. However, it requires encryption operation process mandatorily and includes data exchange in real time. Thus, it is necessary for the HSM to include ultra-lightweight and high performance module to provide required functions while maintaining real-time requirements.

6.4 Detection and recovery against ECU security threat

The CAN bus does not provide inter-certification between ECUs so it is vulnerable to spoofing attack. Attackers can broadcast forged or modified packets easily through spoofing attack. Therefore, it is necessary for the CAN controller to overwrite an error frame to the message to notice abnormal ECU to monitoring nodes when the CAN controller detects a spoofing message as a monitoring system to validate integrity of packets. In addition, when abnormal messages are detected through monitoring, it is necessary to perform detection and recovery in real time by combining recovery development technology that can restore information state of ECUs stored in the recovery partition back to the “initial information state”.

6.5 Construction of security testbed

It is necessary for self-driving cars to have an environment where real self-driving car and infrastructures are analyzed with respect to impact and effect scope during accident occurrence as well as security compliance verification and validation inspection, and develop tools that can discover vulnerabilities of components of self-driving cars and environments that can verify vulnerability through simulation hacking and analysis on security vulnerability of service infrastructures and devices. Therefore, a testbed shall be constructed to test various security functions and diagnose possible hacking attacks in advance to complement the system.

7 Conclusion and future research

In recent years, technologies of autonomous driving systems have strived for perfection. Most automobile companies aim to commercialize and run self-driving cars on the road by 2020. For commercialization, sensor parts which play a role as eyes in humans have been advanced continuously. Nonetheless, much attention has been paid to hacking issues in self-driving cars so security has become the main subject. Compared to explosive growth in technologies and industries related to self-driving cars, security system and security technology that support self-driving cars have been fallen behind. Accordingly, various technologies such as security technology development, security standardization, privacy protective certification technology development, hardware-based high-speed encryption module development have been applied to V2X communication around the world.

In the present paper, security threats against self-driving cars and infrastructures were analyzed and possible attack scenarios were developed to predict the impact. Furthermore, a method of R&D on security technology was also provided to ensure cyber security.

For the future research, experiments will be conducted with regard to possible attack scenarios and required security technologies will be studied and developed.

Acknowledgments. This research was supported by the Research Program through the National Security Research (NSR) Institute of Korea. (#2016-038)

References

1. Collaboration of Related Ministries, Support Measures for Commercialization Self-driving Cars, Ministry of Land, Infrastructure and Transport (2015)
2. Lee, Jeonghoon; Kim, Hyunyong, Electricity flows in car. The Collaboration, eBEST INVESTMENT SECURITIES Co., Ltd (2015)
3. Korea communications Agency, Current trend of technologies and prospects of security technology for vehicle communication used in intelligent road systems No. 59 (2014)
4. Shin, Younoh, Precision Map, V2X Self-driving car without road infrastructure, Can it be run well? e4ds News (2016)
5. Hyundai Motor Company, Press release of the current trend on technological development of self-driving cars (2016)
6. Oh, Hyunseo, Technological trend of cooperative driving via V2X communication ETRI (2015) 33-36
7. Kim, Sangguk, Autonomous driving functional systems, Industrial Information Analysis Department (2015)
8. Policy Research Center, V2X communication, Rising core technology for intelligent traffic system, Trend Focus (2014) 45-46
9. Ministry of Land, Infrastructure and Transport, Report on the current trend of standardization on ITS in overseas Vol. 2(2014)
10. Lee, Donghoon, Technology fusion of vehicle and ICT and current trend of security technology, TTA Journal. Vol. 153 (2014)
11. Sohn, Gyungho, ETRI Privacy technology, ISO standard adoption, ZDNet Korea (2013)
12. Lee, Yuji, Launching of technology development for V2X integrated security for safe autonomous driving era, BYLINE NETWORK (2016)
13. Idan, E.: ProtectivX Hacker Detection System Helps Reduce the Threat of a Collision Caused by Hackers. PR Newswire (2016)
14. Norbert, B., Sebastian, M., Jonathan P., Mirko, L., Martin, M., Daniel, E., Michel, S., Michael, F., Rim, M., Marcello, L., Frank, K.:V2X Security Architecture V2. Vol. 1. PERSERVE (2014)

Lightweight IEC 61850 Secure Communication Module for Microgrids

Yun-Sam Kim, Gunhee Lee, JungMin Kang

National Security Research Institute, 1559 Yuseong-daero, Yuseong-gu, Daejeon, Korea
{bijak,icezzoco,jmkang}@nsr.re.kr

Abstract. Microgrid systems use distributed renewable energy resources to reduce dependency of traditional power systems. These energy devices are located at various places because of their characteristic and communicate using Ethernet control protocol like IEC 61850. Because of performance requirement of IEC 61850 standard, most commercial IEC 61850 libraries do not apply security functions even though IEC 62351 standard series define security necessary of IEC 61850. Naturally, IEC 62351 series are not mandatory but optional. However, absence of security functions means energy systems transfer raw data using Ethernet network and are vulnerable to known cyber-attacks like man-in-the-middle attack. In this paper, we suggest IEC 61850 secure communication module to prevent cyber-attacks with satisfaction of IEC 61850 standard performance requirements. Our module is lightweight to satisfy 5m sec reactivity and provides integrity and confidence using bi-directional authentication and X.509 certificates.

1 Introduction

Microgrid system is next-generation grid system using distributed renewable energy resources(DER) like solar energy, wind turban to reduce role of traditional grid system for small area [1]. These energy resources are influenced by environments like weather, location and so on. Wind turban must be placed at slope and solar charger may be set in the roof of houses or buildings. These physical characteristic makes energy resources dispersed. Amount of energy generation of distributed energy resources and energy consumption of microgrid energy consumers determines additional energy necessary. If generation is larger than consumption, microgrid save spare energy to energy saving system(ESS). On the other hand, microgrid get energy from energy plants.

Data of distributed renewable resources is very important. When a hacker modifies control data to DER or information data of DER using cyber-attacks like man-in-the-middle attack, it causes stopping DER or instability of demand and supply of power. At the worst social chaos like blackout may be happened. To prevent these cyber-attacks, protection of data is important issue.

However, most microgrid communication protocols do not supply security functions. Because of physical characteristic of DER, it is impossible to isolate physically

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_43

DER and MG system. It means that communication protocol must supply security functions like bi-directional authentication and encryption. However, most protocols like IEC 61850 consider function and response time as important, they don't have security definitions. Of cause, IEC 62351 standard series defines security necessaries of microgrid protocols, these standards are not mandatory but optional. Due to these reason, many microgrid protocol libraries do not concern security issues.

In this paper, we suggest lightweight IEC 61850 secure module which satisfies both performance and IEC 61850 security requirements. IEC 61850 is most restrict and important protocol to microgrid. When security module satisfies requirements, porting security functions to other protocols is not difficult. Session 2 says that what is microgrid system briefly. And session 3 is our proposed lightweight IEC 61850 module. Session 4 show that how much overhead is happened by our IEC 61850 module. Session 5 is conclusion of this paper.

2 Microgrids

Microgrid system is new and smart bi-directional grid system using distributed renewable energy resources like solar energy, wind turban and electrical vehicle. By U.S. Department of Energy, microgrids is subset of smart grids [2]. Microgrids controls or monitors energy generation and consumption in small area like cities, buildings and campuses. One important key issue of microgrids is generation and consumption happens at the same place [3]. At the view of traditional grid systems, place of energy supplier and consumer is different. Energy is generated at power plants and delivered to houses, buildings and factories. However, at the view of microgrids, a house generates power energy using solar power system and spends energy within the house.

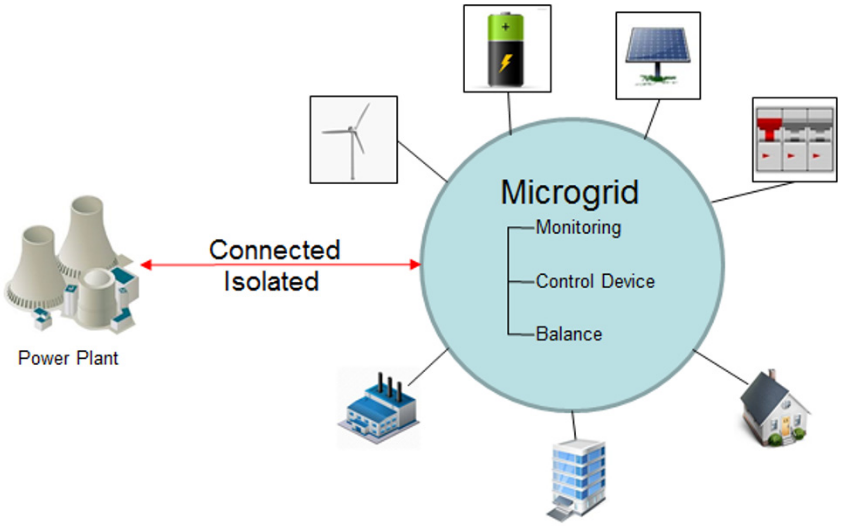


Fig. 1. Concept of Microgrids

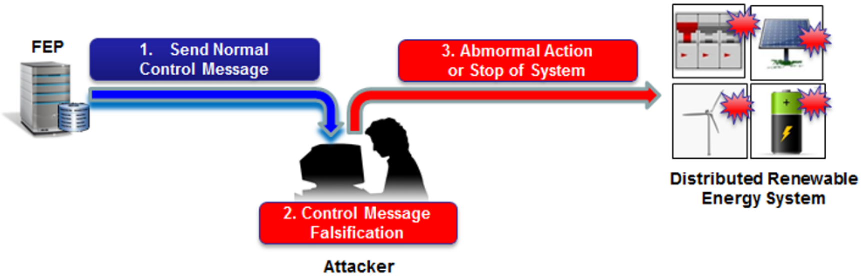

FEP
1. Send Normal Control Message
3. Abmormal Action or Stop of System
2. Control Message Falsification
Attacker
Distributed Renewable Energy System

Fig. 2. Attack Using IEC 61850

When generation of energy is larger than consumption, owner of the house can store spare energy to ESS or sell to other people. When generation is smaller than consumption, microgrid may connect to traditional power grid and get lacked energy from power plants. Figure 1 shows this idea. To control generation and consumption of energy, microgrid system must monitor, control and balance power generation and consumption. It is more difficult and important than traditional grid system because energy generation system is spited at large area. Microgrids can be operated as both connected mode or isolated mode.

Microgrids uses existing communication protocols like IEC 61850, DNP3 and Modbus [4]. These communications protocols are developed to be used by traditional power plants or power transmission and distribution systems. Because of characteristic of traditional power system like response time requirement and enclosed environment, communication protocols focus to response performance and reliability. These protocols are lightweight and simple. Using these raw protocols are not enough because of characteristic of distributed energy resources. Many protocol libraries focus to traditional power grid and do not supply data encapsulation. This means that hackers can see and handle raw data of renewable energy resources which is scattered to wide area easily.

IEC makes security standards to solve this problem. IEC 62351 standard series [5] define security necessaries which are suitable to smartgrid and microgrid systems. However, these defines are not mandatory but recommended options. Applying these defines makes drop of response performance and many protocol libraries do not satisfy this option.

3 Lightweight IEC 61850 Secure Module

IEC 61850 is a communication standard for electrical substation automation systems and key protocol of microgrid [6]. In microgrid, IEC 61850 is used at not only renewable energy generation devices but also microgrid control systems. IEC 61850 defines mandatory response time and IEC 61850 library developers must satisfy this performance requirement. Because of this requirement, many IEC 61850 protocol libraries don’t concern security issues and are weak to cyber-attacks like man-in-the-middle(MITM) described at figure 2. When a MITM attack occurs successfully, an hacker can stop energy generators or give wrong information to microgrid control system. This means the hacker can make an electric power shortage of microgrid on purpose. This happens machine broken and blackout at last. To prevent MITM attacks, it is necessary to apply TLS security standard. However, using pure TLS to IEC 61850 makes drop of performance and dissatisfaction of IEC 61850 requirements. We make secure lightweight IEC 61850 protocol using KEPCO IEC 61850 library. Original KEPCO IEC 61850 library does not have any security functions. This library follows pure IEC 61850 standard. We insert security function to the library like figure 4. We use openssl codes [7] to insert security.

Our secure module is subset of KEPCO IEC 61850 library. However, our module provides secure communication interfaces to have generality. Secure module replaces TCP/IP socket functions of original KEPCO library and take charge of IEC 61850 data transfer. Secure IEC 61850 module is separated as four layer like figure 3. Data transfer layer have a part of TCP/IP socket and TLS socket. TCP/IP socket does common TCP/IP handshake and TLS socket send and get real IEC 61850 data.

Data transfer interface takes charge of making TLS header and transferring IEC 61850 data using TLS socket of data transfer layer. Data transfer interface gets or sends encrypted data and message authenticate code from or to encryption and authentication abstraction interface. This interface also takes charge of TLS handshake, renegotiation and resume protocols. These protocols satisfy IETF TLS standards. Negotiation part is used to make secure channel when two devices are connected. When two devices use negotiation, bi-directional authentication using each certificate. Figure 4 shows bi-directional authentication. When each device verifies received certificate, it uses OCSP protocol to check validity. OCSP information is saved in the

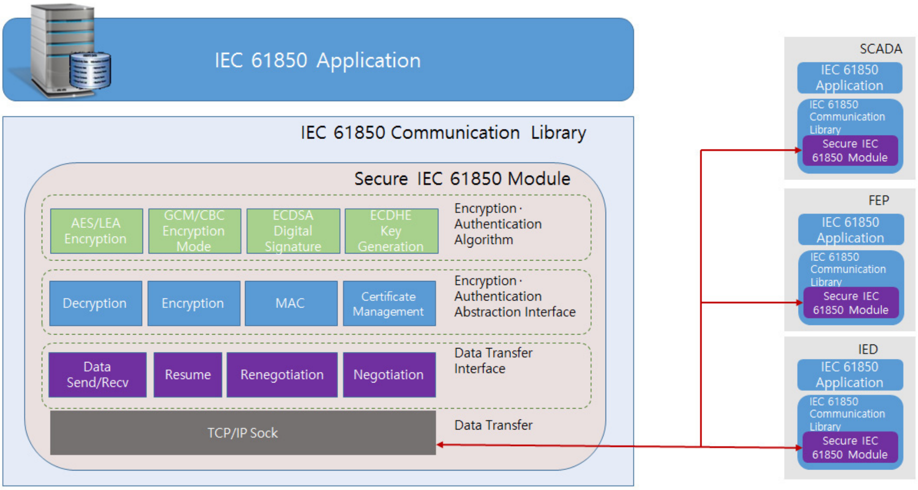
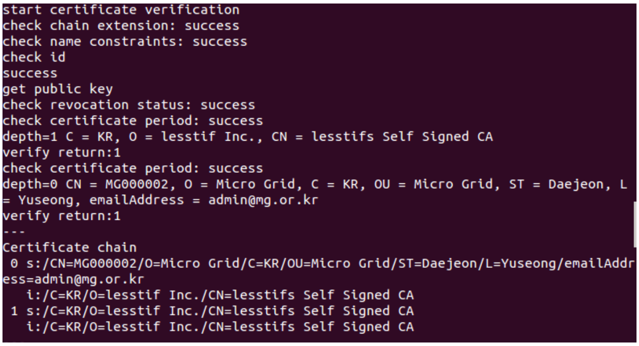
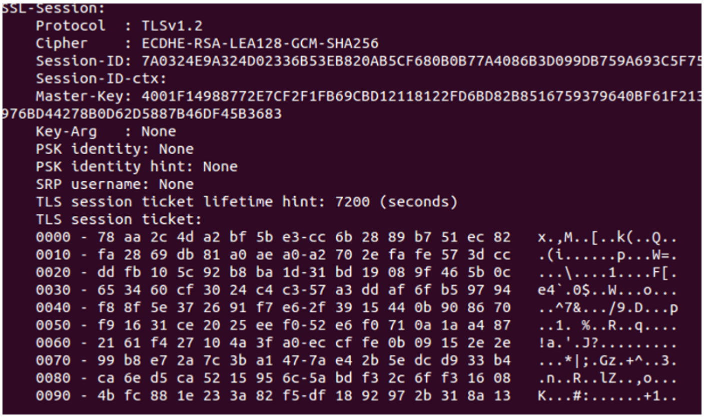


Fig. 3. Overview of Secure IEC 61850



```
start certificate verification
check chain extension: success
check name constraints: success
check id
success
get public key
check revocation status: success
check certificate period: success
depth=1 C = KR, O = lesstif Inc., CN = lesstifs Self Signed CA
verify return:1
check certificate period: success
depth=0 CN = MG000002, O = Micro Grid, C = KR, OU = Micro Grid, ST = Daejeon, L
= Yuseong, emailAddress = admin@mg.or.kr
verify return:1
---
Certificate chain
 0 s:/CN=MG000002/O=Micro Grid/C=KR/OU=Micro Grid/ST=Daejeon/L=Yuseong/emailAddr
ess=admin@mg.or.kr
   i:/C=KR/O=lesstif Inc./CN=lesstifs Self Signed CA
 1 s:/C=KR/O=lesstif Inc./CN=lesstifs Self Signed CA
   i:/C=KR/O=lesstif Inc./CN=lesstifs Self Signed CA
```

Fig. 4. Certificate Verification



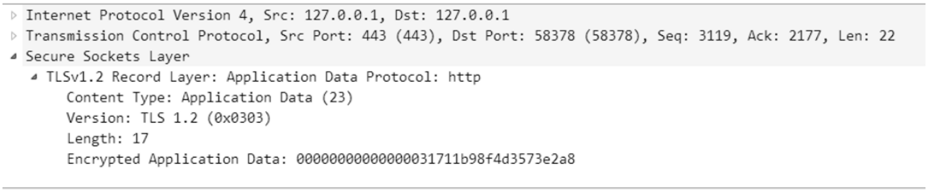
```
SSL-Session:
    Protocol  : TLSv1.2
    Cipher    : ECDHE-RSA-LEA128-GCM-SHA256
    Session-ID: 7A0324E9A324D02336B53EB820AB5CF680B0B77A4086B3D099DB759A693C5F75
    Session-ID-ctx:
    Master-Key: 4001F14988772E7CF2F1FB69CBD12118122FD6BD82B8516759379640BF61F213
976BD44278B0D62D5887B46DF45B3683
    Key-Arg   : None
    PSK identity: None
    PSK identity hint: None
    SRP username: None
    TLS session ticket lifetime hint: 7200 (seconds)
    TLS session ticket:
    0000 - 78 aa 2c 4d a2 bf 5b e3-cc 6b 28 89 b7 51 ec 82   x.,M..[..k(..Q..
    0010 - fa 28 69 db 81 a0 ae a0-a2 70 2e fa fe 57 3d cc   .(i......p...W=.
    0020 - dd fb 10 5c 92 b8 ba 1d-31 bd 19 08 9f 46 5b 0c   ...\....1....F[.
    0030 - 65 34 60 cf 30 24 c4 c3-57 a3 dd af 6f b5 97 94   e4`.0$..W...o...
    0040 - f8 8f 5e 37 26 91 f7 e6-2f 39 15 44 0b 90 86 70   ..^7&.../9.D...p
    0050 - f9 16 31 ce 20 25 ee f0-52 e6 f0 71 0a 1a a4 87   ..1. %..R..q....
    0060 - 21 61 f4 27 10 4a 3f a0-ec cf fe 0b 09 15 2e 2e   !a.'.J?.........
    0070 - 99 b8 e7 2a 7c 3b a1 47-7a e4 2b 5e dc d9 33 b4   ...*|;.Gz.+^..3.
    0080 - ca 6e d5 ca 52 15 95 6c-5a bd f3 2c 6f f3 16 08   .n..R..lZ..,o...
    0090 - 4b fc 88 1e 23 3a 82 f5-df 18 92 97 2b 31 8a 13   K...#:......+1..
```

Fig. 5. Negotiation Result

received certificate. When bi-directional authentication is successful, two devices determine encryption and crypto mode. And then same session key is generated at each device. Figure 5 is handshake result.

Renegotiation module operates when session key needs to be changed. According to IEC 62351 standards, session key must be changed within one day or 10,000 times of communication. We make a configure file to manage renegotiation period. A user can deal renegotiation time and count. However, when the user sets larger value than IEC 62351 standards, renegotiation module ignores user setting and follows IEC 62351 standard. Resume function is happened periodically to remain secure channel.

Encryption and authentication abstraction interface provides common interface about encryption and authentication algorithm. Our module support AES and LEA as encryption algorithms and GCM/CBC as crypto modes basically. However, because of national necessary, additional multiple encryption methods and crypto modes may be supplied. So, secure module must be easy to insert new encryption methods. Encryption and authentication abstraction interface supports transparency in encryption algorithms. These interfaces give guides about adding encryption methods. Also, this



```
▷ Internet Protocol Version 4, Src: 127.0.0.1, Dst: 127.0.0.1
▷ Transmission Control Protocol, Src Port: 443 (443), Dst Port: 58378 (58378), Seq: 3119, Ack: 2177, Len: 22
◢ Secure Sockets Layer
  ◢ TLSv1.2 Record Layer: Application Data Protocol: http
      Content Type: Application Data (23)
      Version: TLS 1.2 (0x0303)
      Length: 17
      Encrypted Application Data: 0000000000000031711b98f4d3573e2a8
```

Fig. 6. Encrypted data with LEA and GCM

abstraction interface also manages loading and verifying X.509v3 certificate and OCSP protocol.

Encryption and authentication algorithm part is charge of real data encryption and message authentication. We support AES and LEA as encryption algorithms and GCM/CBC as crypto mode. Our mode support signature algorithm as ECDSA and RSA. Key generation algorithm is ECDHE. Because of security problem and performance, we do not provide whole algorithms of TLS. Because of same reason, our module support only SHA 256 as hash algorithm. These algorithms are selected based on speed, safety and national requirements. For example, SHA 512 is powerful hash algorithm but is very slow. On the other way, SHA-1 is very fast but does not guarantee safety. Because to safety and performance, our module supports neither SHA 512 nor SHA-1 hash algorithms. Additional algorithms may be inserted to our module if necessary. Figure 6 is capture of encrypted data value '19' using Wireshark TCP/IP packet capture program. Original value of this data is only known to two correct devices. Attackers cannot know original value. Encrypted application data includes encrypted data using LEA and authenticated tag using GCM.

4 Performance Test

Performance of data transfer is one of very important requirements of IEC 61850. IEC 61850 standard says that receiver of data must response within 5ms. To check out performance, we make simple test server and client programs. These programs increase 1-byte data size for every 5 second. Client program is set to send pre-defined static data and prints echo data. Server program echo client data.

Table 1 is our test environments. Secure communication is effected to system hardware due to encryption, negotiation and so on. This means that some embedded device implement case may not satisfy IEC 61850 requirement.

Unlikely to Internet environment, packet size of IEC 61850 and other communication protocol for microgrid is very small. Most of transferring data is sending status or numeric information and their ack messages. Sending large size data or bulk memory data is rarely existing except special cases. So, we limit data size as 20k bytes. Most data of IEC 61850 has less than 1k bytes and we determine that 20k bytes is enough large size in a microgrids. Like (a) of figure 7, applying security rarely influences to overhead to send small size data under 100 bytes. And overhead of encryption of larger data than 100 bytes also is durable. (b) of figure 7 is average of data encryption

Table 1. Test Environments

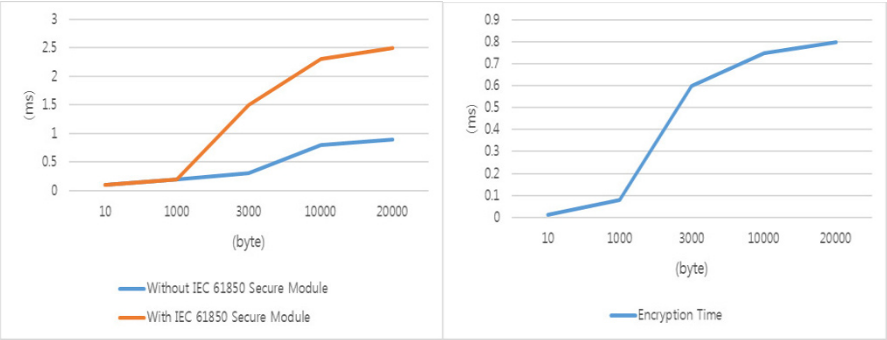
	Type	Value
Client	CPU	Intel Core i-7 5500U
	RAM	8GB
	OS	Ubuntu 14.04
Server	CPU	Intel Core i-7 5500U
	RAM	8GB
	OS	Ubuntu 14.04
OCSP	CPU	Intel Core i-7 3517U
	RAM	4GB
	OS	Ubuntu 14.04

and decryption. This result calculates client's data encryption and server's data encryption and verification by subtract average transfer time. Result says that encryption and decryption speed does not over 0.3 ms.

Result of figure 7 does not include negotiation, renegotiation and resume time. Performance of these protocol does not increase by data size. Their time is static. We average these time like table 2. In our experiment, renegotiation time does not over 3ms. Using result of figure 7 and table 2, performance of our worst case does not

Table 2. Additional Execution Time

Phase	Operation Time
Negotiation	2.53ms
Renegotiation with Sending Certificate	2.61ms
Renegotiation Without Sending Certificate	2.14ms
Resume	0.73ms



(a) Echo Test Performance (b) Encryption Performance

Fig. 7. IEC 61850 Secure Module Echo Test Result

exceed 4 ms and satisfies 5ms IEC response requirement. However, our test is not field test. If internet protocol of microgrids goes bad, performance problem may be happened.

5 Conclusion

Microgrids are one of new and important technologies to manage energy generation and consumption. Widespread distributed energy resources and leak of security of IEC 61850 makes microgrids vulnerable. Moreover, because of performance requirement of IEC 61850 makes insertion of security function to communication libraries harder.

In this paper, we suggest secure communication module to protect microgrids from man-in-the-middle attacks. Our module provides encryption and bi-directional authentication based on IEC 62351. At the same time our module minimizes performance overhead to satisfy IEC 61850 requirement.

However, security issuers of microgrids are not researched enough. Communication protocols of microgrids may have several leaks but researches are not enough also. This means that it is hard to say that our secure module has enough security ability in microgrid environments. It may be large a challenge to show that our module provides enough security functions.

Performance is also important issue. In this paper, we propose lightweight secure IEC 61850 module using KEPCO IEC 61850 library. Using our module, we encrypt IEC 61850 data and prevent man-in-the-middle attacks with IEC 61850 performance requirement. However, it may be hard that our secure module with low computation power embedded device because of renegotiation time. If renegotiation and data transfer is happened at the same time, it may be hard to satisfy IEC 61850 necessary. It is necessary to control renegotiation.

We could prevent some attacks which was reveled at internet environment. However, attack and research using IEC 61850 is rarely existed. Moreover, microgrid system is not spread widely and this technique is now researching. It is very hard to get attack scenarios. We must simulate real attack using IEC 61850 in microgrid environment and check safety of our proposed secure library.

Acknowledgments. This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (No. 20141010501870).

References

1. Lasseter. R. H.: Microgrids. Power Engineering Society Winter Meeting, 2002. IEEE. 2002
2. Office of Electricity Delivery & Energy Reliability, THE ROLE OF MICROGRIDS IN HELPING TO ADVANCE THE NATIONS'S ENERGY. http://energy.gov/oe/services/technology-development/smart-grid/role-microgrids-helping-advance-nation-s-energy-system

3. Lasseter. R. H., Paigi.P.: Microgrid: a conceptual solution. Power Electronics Specialists Conference Vol.6 (2004). 4285-4290
4. Leelaruji. R., Luigi.V.: State-of-the-art in the industrial implementation of protective relay functions, communication mechanism and synchronized phasor capabilities for electric power systems protection. Renewable and Sustainable Energy Reviews Vol.16(2012) 4385-4395
5. Core IEC Standards. http://www.iec.ch/smartgrid/standards/
6. Yoo, B.K., Yang.H.S., Kim.W.Y.: Communication architecture of the IEC 61850-based micro grid system. Journal of Electrical Engineering and Technonogy. Vol.6(2011) 605-612
7. OpenSSL, https://www.openssl.org/

A new approach to building a disguised server using the honey port against general scanning attacks

Hyun Soo Park, Young Bae Jeon, Ji Won Yoon

Abstract The port scan is a well-known technique which malicious people often use before attacking a server. The attackers obtain the fingerprint of the target server by scanning ports and then make an attack scenario. Several approaches including the 'port knocking' and 'Single Packet Authorization' (SPA) have been developed to defense port scanning attack and allow only authenticated users to access ports. However, the approaches have a disadvantage that the attacker can obtain the information about the ports by applying inference techniques given observed patterns. If a router, connecting the server to the outside, is cracked by the attacker, he or she could infer particular ports which authenticated users consistently use to communicate with the server. In this paper, we propose a new defense method, *Honeyport*, which can prevent the attackers from obtaining the information about ports and make them demotivated by disguising the server as peripherals. Furthermore, by adopting packet encryption as in IPSec, the attacker cannot obtain the critical information via packet sniffing in our proposed model.

1 Introduction

Port scanning is one of the major and crucial functions to identify and connect devices for adequate communication in a network. However, this function can be abused by the malicious people, unlike the original purpose. Attackers obtain the fingerprint of a victim server by scanning the server's ports to find the vulnerability and weakness of the victim server. With the scanned fingerprint, the attacker can just check which ports are opened and which services are provided. Afterward, the attacker explores the adequate attack scenario to give a damage the target server or to compromise the victim server.

Hyun Soo Park, Young Bae Jeon, and Ji Won Yoon
Center for Information Security Technologies(CIST), Korea University, Republic of Korea,
Corresponding author is J. W. Yoon e-mail: jiwon_yoon@korea.ac.kr

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_44

That is, since the attacker can find the proper attack types using the port scan, we need to develop a protection system which removes or avoids such unwanted port scanning attack. Several defense approaches including the port knocking[3] and the Single Packet Authorization(SPA)[4, 5] have been developed to hide the information about the server's ports. They are designed to hide the using port about when the server transmits and receives packets over the communication by the authentication. IPSec is an alternative approach which encrypts and encapsulates packets given an IPSec protocol to communicate. Although packets are sniffed from outside, the server can prevent the packet's information from being exposed by the IPSec because it is encrypted. Also, a new scheme called honeypot has been introduced to hide the server's identity. This technique installs a trap server and leads the hacker to attack the trap server, not a real server. The trap server collects and analyzes attackers' logs.

However, there are few techniques which fundamentally prevent hackers' attack trials. While ports are mainly focused in port knocking and the SPA, IPSec considers not the port but data itself using encryption. And the honeypot's point of view is the defense about the only cheating attacker. In this paper, we propose a new technique to integrate all of these conventional approaches and lead the attackers not to intrude the server fundamentally.

This paper is organized as follows. In Section 2, we describe background knowledge about the port scanning, the port knocking, and the SPA. We propose the algorithm about *honey port* in Section 3 and show the results of the experiment related to *honeyport* in Section 4. Finally, we evaluate our approach and conclude with an overall summary of our approach.

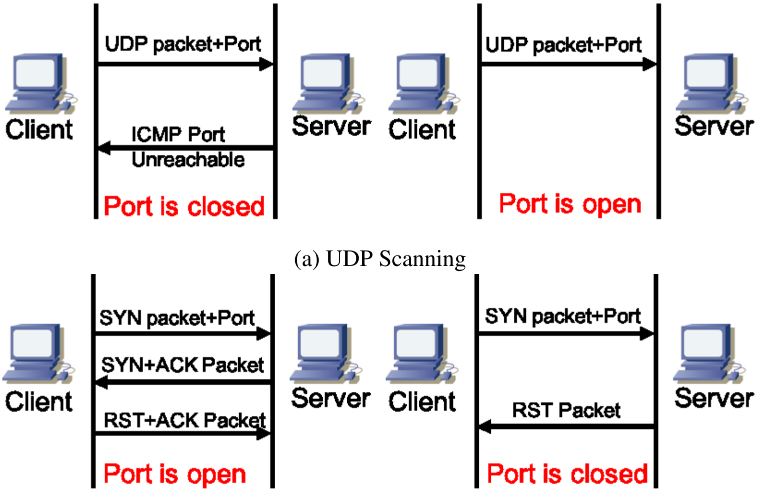
2 Background

2.1 Port Scanning

It is known that the port scan is used to scan a target system in order for exploring its potentially exploitable service. There are mainly two different ways in the port scanning: User Datagram Protocol (UDP) scanning[1, 2] and Transmission Control Protocol (TCP) scanning.

Fig. 1-(a) shows how UDP scanning works. In UDP protocal, "ICMP Port Unreachable" message is generated by the server to inform the client that the destination is unreachable if the destination port is said to be closed. Otherwise, if the client does not recieve any message from the server after sending the the UDP packet with a destination port, we may infer that the destination port of the server is opened.

However, while the client certaintly knows the closeness of the port but cannot exactly know the openness of the server's port since the response packet has been lost in UDP protocol.



(a) UDP Scanning

(b) TCP Scanning

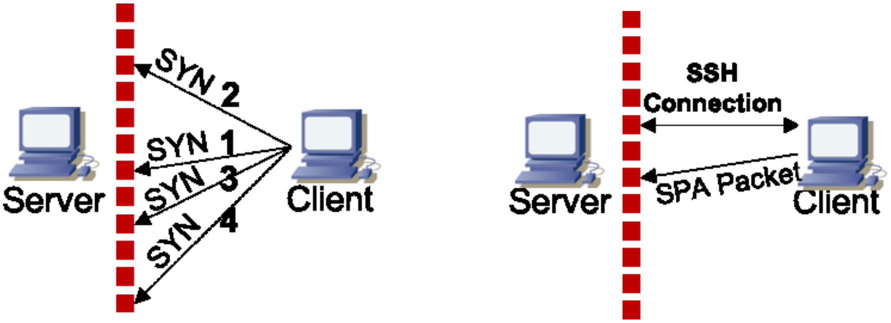
Fig. 1: UDP Scanning and TCP Scanning

TCP is the other port scanning protocol to find the opened ports [1, 2]. Figure 1-(b) shows how TCP Scanning works. When sending a SYN packet to a destination port over TCP Communication, the client receives a *SYN* + *ACK* packet if the port is opened. Otherwise, the client receives a *RST* + *ACK* packet. Thus, the attacker obtains information about which ports are opened.

2.2 Port Knocking

To date, several approaches have been developed to prevent an attacker from the unwanted port-scan.

One of the most well-known approaches is the port knocking[3] which opens ports on a firewall by generating a connection attempt on a set of pre-specified closed ports. When a sequence of connection attempt is correctly received, a server authorizes a user as shown in Fig. 2-(a). Attackers can approach to the server through the firewall but they will fail to connect the server since they do not know the correct sequence of the connection attempt. However, there is vulnerability in this protocol. If the router which connects the server to the outside is compromised by an attacker, the ports are exposed to the attacker.



(a) Port Knocking

(b) Single Packet Authorization

Fig. 2: Two approaches to hide opened ports against port scanning attacks: (a) ‘Port Knocking’ and (b) ‘Single Packet Authorization’ (SPA)

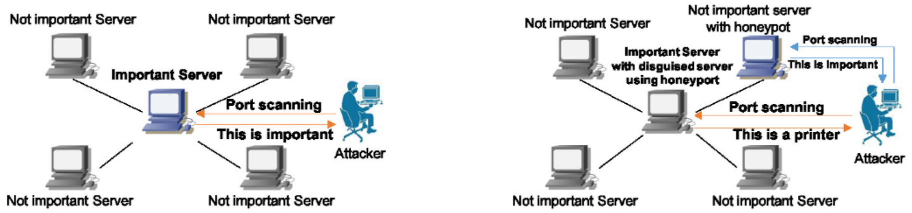
2.3 Single Packet Authorization

Single Packet Authorization(SPA)[4, 5] is a similar way to port knocking in that it requires only a single ‘knock’ for the communication. SPA combines packet filter via drop and packet sniffer. SPA uses packet’s payload to prove the ownership of the information, not the packet’s header in Fig. 2-(b). Clients send only one packet about own identity to the SPA server. This is possible because MTU’s size is hundreds bytes unit in the common network. However, this approach has the same vulnerability with the port knocking.

3 Proposal Algorithm

The background section shows two main approaches to preventing port scan from the attacker: ‘port knocking’ and ‘single packet authorization’ (SPA). However, their effectiveness is rather limited, and the attacker can bypass the approaches using their vulnerabilities. For instance, when ssh is not used in the approaches, the attacker can infer which ports are used because they consistently use the identical ports. In addition, if the routers linking to the victim server are compromised by the hackers, the attacker can monitor the pattern and sequence of the ‘knock’s.

From this point of view, we propose an algorithm which fixes these vulnerabilities in the conventional approaches. Therefore, in order to remove the vulnerabilities, we propose a scheme to disguise the victim system using fake ports. In this paper, we are disguising the target server as a trivial terminal which looks ignorable while an actual dummy terminal is also disguised as a significant server by *honeyport*. In the proposed method, the server falsifies the response which is requested from the port scanning by the attacker and sends it to the attacker. In Figure 3, we can see that the attacker confuses the important server with the printer.



(a) Without the honeyport and the honeypot

(b) With the honeyport and the honeypot

Fig. 3: Administrators can hide their property by opening fake ports and we call this by *honeyport*. The important server can be disguised as an insignificant sever using *honeyport*: we can see that the attacker regards the important server as the printer in this figure.

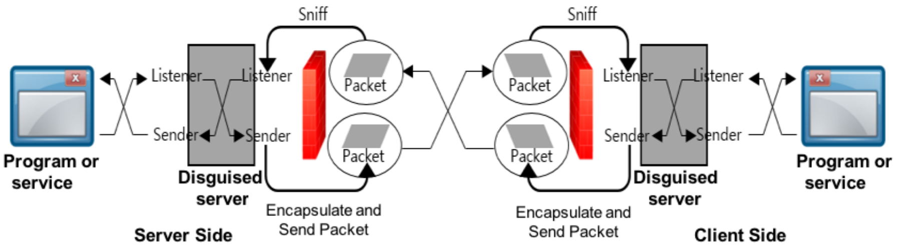


Fig. 4: Flowchart about the disguised server: The server and client communicate each other by the disguised server through *Listener* and *Sender*.

For this, a system administrator can make the disguised server as plotted in Figure 4. The disguised server's function consists of four parts : (a) *listener*, (b) *sender*, (c) *spoofer* and (d) *sniffing packets*.

1. *Listener* catches packets from outside to inside and from inside to outside;
2. *Sender* forwards packets from outside to inside and sends packets from inside to outside;
3. *Spoofer* provides fake information to attacker during port scanning. However, because used ports are closed and hidden by the firewall, programs actually cannot catch packets from outside; and
4. *Sniffing packets*, which is operated in the disguised server, collects all packets blocked by the firewall blocks because the port is closed. The disguised server brings all packets filtered by the firewall and delivers them to the listener.

3.1 Listener

Listener catches two kinds of packets, which move from inside to outside and from outside to inside.

3.1.1 Packets going from inside to outside

Listener sniffs packets going from inside to outside and sends the packets to *Sender* after the series of following processes:

1. *Listener* firstly attaches a source IP address, a source port, a destination port, and a timestamp to packet's payload.
2. *Listener* encrypts it using Advanced Encryption Standard (AES) with a shared key.
3. *Listener* attaches a hash value returned by a hash function, HMAC, which has a parameter of packet's encrypted payload.
4. *Listener* attaches ”Sniffer” to the payload to recognize when packets arrive.

After these processes, the structure of a packet's payload is shown in Figure 5. We can see the tag written as ”Sniffer” in the payload's head and MAC value behind it. Lastly, we can find out the encrypted information of IP address, Port, Timestamp and Original payload in the tail of the packet's payload. *Listener* sends these packets to outside by *Sender*.

The reason for putting additional information into the existing payload is to prepare defense against various attacks. By adding IP address, port, and timestamp into the payload, we prevent replay attack and ARP Spoofing. Besides, by adding MAC into the payload, we prevent that the attacker falsifies packets.

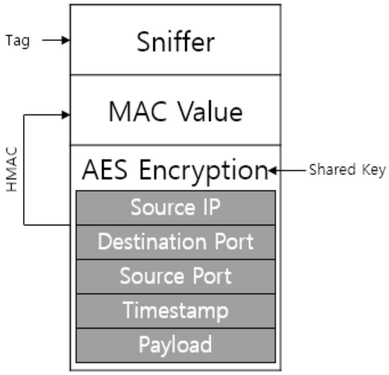


Fig. 5: Payload encapsulation of packet

3.1.2 Packets going from outside to inside

Packets going from outside to inside is processed with the inverse order of processes for packets going from inside to outside. This process is called decapsulation. Because the port which should be used is closed, packets are passed only through the disguised server. In the disguised server, there is much overhead to check every packet. Therefore, we sniff only packets which are labeled as ”Sniffer” in front of packet’s payload. This has the following steps:

1. *Listener* detaches ”Sniffer” in packet’s payload.
2. *Listener* compares MAC in packet’s payload and the value returned by a hash function, HMAC, which has a parameter of the ciphertext in payload.
3. *Listener* decrypts the ciphertext in payload and compare source IP address and own IP address in decrypted text.
4. *Listener* verifies timestamp.
5. *Listener* removes the additional information in the payload except for original payload.

After these processes, the packets are forwarded to inside. Services or programs which have received it from the disguised server can operate without changing the existing protocol since packet recover by decapsulation. The details of the *Listener* is demonstrated in Algorithm 1.

Firstly, the first **if** statement means that if destination IP address and server IP address are equal, and payload includes “Sniffer”, the packets are entered to inside from outside. After making sure the ciphertext forgery by comparing MAC, we decrypt the payload’s ciphertext with a shared key. In this process, if the ciphertext

Algorithm 1 Pseudocode for *Listener*

```
procedure LISTENER(pkt)
    if $pkt[IP].dst = my\_ip$ and $pkt[Raw].load$ include "Sniffer" then
        if $pkt[Raw].load.MAC = \mathbf{HMAC}(pkt[Raw].load.data)$ then
            Try :
                $PlainText \leftarrow \mathbf{AES\_Decrypt}(pkt[Raw].load.data, key)$
            Except :
                Print "Wrong key"
                return false
            if $PlainText.ip = pkt[IP].src$ and $PlainText.timestamp$ is validate then
                $Forward\_pkt \leftarrow pkt$
                $Forward\_pkt[Raw].load \leftarrow PlainText.payload$
                $Forward\_pkt[Raw].dport \leftarrow PlainText.dport$
                $Forward\_pkt[Raw].sport \leftarrow PlainText.sport$
                Calculate Checksum of Forward_pkt
                Sender( $Forward\_pkt$ )
    else if $pkt[IP].src = my\_ip$ then
        $pkt[Raw].load \leftarrow pkt[Raw].load|my\_ip|pkt[TCP].dport|pkt[TCP].sport|timestamp$
        $CipherText \leftarrow \mathbf{AES\_encrypt}(pkt[Raw].load, key)$
        $pkt[Raw].load \leftarrow "Sniffer"|\mathbf{HMAC}(CipherText, key)|CipherText$
        Sender( $pkt$ )
```

is decrypted with wrong key, print "Wrong key". After decryption, we verify the real sender by comparing IP addresses and validate the timestamp. If every process is verified, the payload is decapsulated to original payload and port is changed correctly. The packet is sent by *Sender* to inside after recalculating the checksum. When the packet's source IP address and server IP address are equal, the packets are going to outside from inside. Therefore, we encapsulate packet's payload and attach ciphertext and MAC, and send the packet to outside by *Sender*.

3.2 Sender

Sender takes packets from *Listener* and send them to outside or inside. Following subsections represent how to process the outgoing or incoming the packets in *Sender*.

3.2.1 Packets going from inside to outside

After the series of processes of a packet from *Listener*, *Sender* sends the packet to outside with a random port. Besides, while destination IP address is fixed, the destination port and the source port are set randomly. Through this process, even if the packet has been sniffed from midway, the attacker cannot infer the source port and destination port.

3.2.2 Packets going from outside to inside

Sender does not have to process the packet from outside to inside. It is because *Listener* changes the port of the packet in payload to original port. Furthermore, although the packet's source IP address is same as a public IP address, the packet is not blocked by the firewall because they are sent from an internal program, disguised server. Moreover, the service and program can use original protocol since IP address has not been changed and payload is original. For these reasons, we can simply send the packet which is sent from *Listener* to *Sender*.

We can see the pseudocode for *Sender* in Algorithm 2. As the almost every task has been performed in *Listener*, *Sender* simply changes the packet's port to a random port. If the destination IP address is same as the server IP, *Sender* sends it to loopback or internal network from inside of the firewall as *Listener* has already handled the whole work. In the other case, if the source IP address is same as the server IP address then the packet is going from inside to outside. As the packet will be delivered to outside, we should change the ports to random ports to conceal the information about the port.

Algorithm 2 Pseudocode for *Sender*

```
procedure SENDER(pkt)
    if $pkt[IP].dst = my\_ip$ then
        Send( $pkt$ )
    if $pkt[IP].src = my\_ip$ then
        $pkt[TCP].dport \leftarrow$ Random$(0, 65535)$
        $pkt[TCP].sport \leftarrow$ Random$(0, 65535)$
        Send( $pkt$ )
```

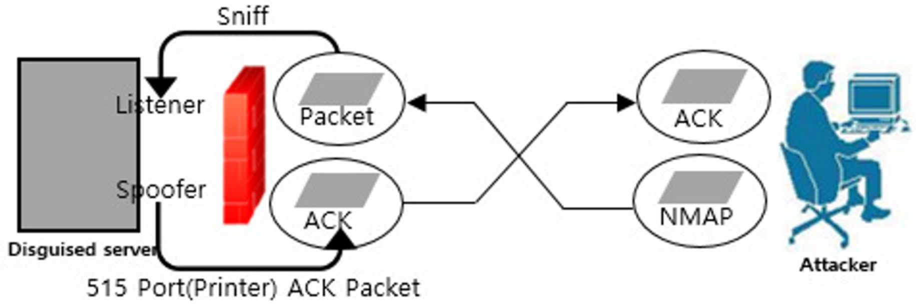


Fig. 6: Port Scanning by an attacker in our approach: If the attacker sends Nmap command to the server for obtaining the information of the server, the server responds to the attacker with ACK packet which makes the attacker regard this as the printer by *Spoofer*.

3.3 Spoofer

Spoofer is a fundamental module of this proposal. If the attackers perform the port scanning to the server and discover that all ports are closed, they will think that the server uses a security program, and then make the scenario to crack the security program. However, as we can see in Fig 6., if the attacker performs port scanning through *Spoofer*, we response with ACK packet for SYN request in 515 port which is usually used in printer. Through this way, the attacker could be led to confuse this server with the printer. As a result, the attacker do not feel the need to attack the scanned sever. In this paper, we will call the port, which response with the falsified information such as 515 port, as the *honey port*.

We can see the pseudo-code for *Spoofer* in Algorithm 3. The general way to find the open ports in port scanning is to make the TCP 3-way handshake. By using this fact, *Spoofer* performs the 3-way handshake to delude the attacker into believing that the port is open when the attacker probes whether 515 port is opened or closed by using command such as *Nmap*. By this method, we make the attacker assume that the server is the printer. To do this, if the packet is received through 515 port, the disguised server creates a new packet, and then set a property of the packet. First, we set ACK value to SEQ value plus one. Second, we set the flag to SYN/ACK.

Algorithm 3 Pseudo-code for *Spoofer*

```
procedure SPOOFER(pkt)
    if $pkt[TCP].dport = 515$ then
        $ACK\_pkt \leftarrow pkt$
        Swap( $ACK\_pkt[TCP].sport, ACK\_pkt[TCP].dport$ )
        Swap( $ACK\_pkt[IP].src, ACK\_pkt[IP].dst$ )
        $ACK\_pkt[TCP].ack \leftarrow pkt[TCP].seq + 1$
        $ACK\_pkt[TCP].flags \leftarrow$ SYN/ACK
        Calculate Checksum of ACK_pkt
        Send( $ACK\_pkt$ )
```

Lastly, we recalculate the checksum and send it to the opponent who performed port scanning with the port used the opponent.

We proceed the experiment with *Server(S)*, *Authenticated Client(AC)*, *Attacker(A)* to build and test our proposed algorithm. We first implement the disguised server, and execute the web server in 80 port with Python, *SimpleHTTPServer* in *Server*. For convenience sake, we will call *S*'s IP address to 1.2.3.4. In a normal condition,

4 Result

Table 1: Procedure of an experiment

```
Server> ./Disguised_server
Server> python -m SimpleHTTPServer 80
Authenticated Client>./Disguised_server
Authenticated Client>curl 1.2.3.4
<html>
<body>
Disguised Server Test
</body>
</html>
Attacker>curl 1.2.3.4
curl: (7) Failed to connect to 1.2.3.4 connection refused
Attacker>nmap 1.2.3.4 -Pn
Starting Nmap 7.01 ( https://nmap.org ) at
Nmap scan report for
Host is up (0.0062s latency).
PORT STATE SERVICE
515/tcp open printer
```

when we use *Nmap* command to the server which does not have the firewall and the disguised server, the result shows that 80 port is open. Then, *AC* also executes the disguised server and use the 'curl' command to access the web server in *S*. Since *A* does not have any right key for the disguised server, *A* executes the disguised server with the wrong key or execute the 'curl' command without the disguised server. Nonetheless, *A* uses *Nmap* command to scan ports and find out the fingerprint. If *S*, *AC* and *A* perform each command, the result is like Table 1.

Table 2: Nmap result of the actual printer

```
$>nmap 2.3.4.5 -Pn
Starting Nmap 7.01 ( https://nmap.org ) at
Nmap scan report for
Host is up (0.0062s latency).
PORT,STATE SERVICE
515/tcp open printer
631/tcp open ipp
9100/tcp open jetdirect
```

As we can see from the result, *AC* uses the right key to the disguised server and accesses to the web page without any problem, while *A* cannot. And also, even if *A* performs *Nmap* for port scanning to *S*, the result shows that only the SYN/ACK packet is received from 515 port which is usually used in the printer. Moreover, we can see that the port scan result is similar to the results which is obtained when the actual printer is scanned as shown in Table 2. For convenience sake, we set the printer IP address to 2.3.4.5. In order to show not only 515 port but also other ports such as 9100 port, we just need to modify the source code of *Spoofer* a little.

Table 3: Calculation of the packet size

Case	**Calculation**	**Bytes**	**Note**
Sniffer Tag	1 byte * 7	7 bytes	7 letters ”Sniffer”
MAC	(160bits/4bits) bytes	40 bytes	Using HMAC 160 bits. Represent one letter instead of hexadecimal
CipherText	$\lceil$256bits/6$\rceil$ + 1 byte	44 bytes	Using AES 256 bits. Dividng by 6 because of encoding Base64
Split Character	1 byte * 2	2 bytes	Split letters for Sniffer, MAC, CipherText
Sum		93 bytes	

Through this, *Attacker* cannot recognize the real web server so the attacker is highly likely to change the target. However, the proposed method in this paper essentially has overhead in encryption and decryption. In this experiment, while an average time is 0.017 seconds in the case without the disguised server, an average time is 0.0690 seconds in the case with the disguised server. The other matter is packet size. The packet which passed the channel of the disguised server contains ”Sniffer”, MAC, encrypted values and split characters so that the size is bigger than the origin. The amount of increased size is proportional to payload size, and the encapsulated packet is 93 bytes bigger than the general one without the payload. The reason why 93Bytes has been increased is in Table 3.

5 Conclusion

In this paper, we propsed a new approach to disgusing hosts using honyport and encryption against port-scanning attacks. In this proposed approach, the server uses the fake ports against the unwanted port scan attack. The attacker who scanned ports port scanning recognizes the device as a printer using honeyport. Furthermore, the packet encryption is also embeded in the approach. Therefore, the user who does not have the right key cannot get any information of the packet and is not permitted to access.

There are issues which have to be improved in this paper. Not only simply sending ACK packets of the printer’s port, but also the specification of the printer is required so that the attacker can be perfectly confused of the target device with a required so that the attacker can be perfectly confused of the target device with a printer. Also, we should confirm that program can protect the packet, and any of information is not leaked by the program. If the router is cracked by the attacker, in theory, our proposed algorithm is secure. Nevertheless, we should simulate that our proposed algorithm is secure in the condition where the router has been compromized.

References

1. De Vivo, Marco, et al. ”A review of port scanning techniques.” ACM SIGCOMM Computer Communication Review 29.2 (1999): 41-48.
2. Lyon, Gordon Fyodor. Nmap network scanning: The official Nmap project guide to network discovery and security scanning. Insecure, 2009.
3. Ali, Fakariah Hani Mohd, Rozita Yunos, and Mohd Azuan Mohamad Alias. ”Simple port knocking method: Against TCP replay attack and port scanning.”Cyber Security, Cyber Warfare and Digital Forensic (CyberSec), 2012 International Conference on. IEEE, 2012.
4. Rash, Michael. ”Single packet authorization with fwknop.” login: The USENIX Magazine 31.1 (2006): 63-69.
5. Michael Rash (March, 2014) Single Packet Authorization with Fwknop Cipherdyn. Retrieved from http://www.cipherdyne.org/fwknop/docs/SPA.html
6. Doraswamy, Naganand, and Dan Harkins. IPSec: the new security standard for the Internet, intranets, and virtual private networks. Prentice Hall Professional, 2003.
7. Davis, Carlton R. IPSec: Securing VPNs. McGraw-Hill Professional, 2001. Ferguson, Niels, and Bruce Schneier. ”A cryptographic evaluation of IPsec.”Counterpane Internet Security, Inc 3031 (2000).
8. Provos, Niels. ”A Virtual Honeypot Framework.” USENIX Security Symposium. Vol. 173. 2004.
9. Krawetz, Neal. ”Anti-honeypot technology.” Security & Privacy, IEEE 2.1 (2004): 76-79.

Password Authentication Using One-Time Key-Based Signature and Homomorphic Encryption

Jong-Hyuk Im, and Mun-Kyu Lee
Department of Computer and Information Engineering, Inha University
Incheon 22212, Korea
imjhyuk@gmail.com, mklee@inha.ac.kr

Abstract. User authentication is a process for a system to verify the identity of a claimed user and to give access permission. Although there are many other authentication methods such as biometrics and physical tokens, passwords are still being used in many applications due to easy deployment. To enhance the security against possible attacks such as an off-line dictionary attack, passwords are usually stored in a hashed form using a random nonce called a salt. However, this does not completely solve the security issue. In this paper, we propose a new password-based authentication method using homomorphic encryption where a password is stored in a remote server in an encrypted form and an input password is compared with the stored one on the encrypted domain. For this purpose, we also propose a new cryptographic primitive called one-time private key-based digital signature.

1 Introduction

User authentication is a process for a system such as a computer, a web server and a smartphone to verify the identity of a claimed user and to authorize the user to access the system. Although there are many other authentication methods such as biometrics and physical tokens, passwords are still being used universally in various applications due to easy deployment. In password-based user authentication, the system compares the current input password in a hashed form with the stored password [1, 2]. This is to prevent password leakage even when the system is compromised. That is, an attacker obtains a hashed value of the original password, but does not know the original password. However, this protection may be defeated by a dictionary attack. When the attacker obtains the hashed value of password, s/he may try to find the original password, i.e., the preimage of the hash, by hashing all strings in a prearranged list, e.g., a list of frequently used passwords such as those in a dictionary and comparing the results with the value obtained from the system. Moreover, the list can be precomputed, which we call an off-line dictionary attack. Let us call the list of tuples (password candidate, its hash value) a dictionary. To enhance the security against an off-line dictionary attack, passwords are usually hashed together with a random nonce called a salt. Because the attacker does not know the salt, s/he has to construct multiple dictionaries each of which corresponds to a salt. However, this solution does not completely solve the security issue because the salt just increases the size of the dictionary for off-line dictionary attack. In addition, in an on-line dictionary attack, the salt does not even lower the attack complexity.

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_45

To solve this security issue, we propose a new password-based authentication method using homomorphic encryption [3-6]. Our proposed method stores password data in a remote server in an encrypted form and a data derived from the current input password is compared with the stored one on the encrypted domain. For our method, we also propose a new cryptographic primitive called one-time private key-based digital signature.

2 Preliminaries

2.1 Traditional Password-based Authentication

Passwords are strings of characters used for user authentication [2, 7]. That is, it is used to verify the identity of a claimed user and to give an access permission. The first step in a password-based authentication is to perform a registration procedure which stores the pair of (user identity, password) for each user. On the other hand, in an authentication procedure, the user is asked to enter his/her password. The target system decides whether or not to allow the user's access according to the comparison result of the stored password and the current input password.

If the stored password file in the system is stolen, the passwords of all users may be leaked. Therefore, when storing a password, the system usually stores a transformed result using a cryptographic hash function. For authentication, the system compares the hash of input password with the stored hash value. The rationale for this is that even if an attacker obtains the hash of a password data, it is very hard to restore the original password. However, many users choose simple passwords such as '1234' and 'password'. It enables several attacks such as dictionary attacks, because the distribution of those passwords is not statistically uniform. When the attacker obtains the hashed value of a password, s/he may try to find the original password, i.e., the preimage of the hash, by hashing all strings in a prearranged list, e.g., a list of frequently used passwords and comparing the results with the value obtained from the system. In contrast to a brute force attack where all random passwords are considered, a dictionary attack considers only frequently used passwords.

Dictionary attacks may be classified into two types, off-line and on-line dictionary attacks. In the former, an attacker constructs a dictionary, i.e., a list of hashes of frequently used passwords, in advance to obtaining the actual hash value of a target password. To prevent this attack, passwords are usually hashed with an additional random nonce called a salt. The attacker then has to construct multiple dictionaries, that is, a distinct dictionary should be constructed for each salt because s/he does not know the salt before obtaining the target password file. This approach increases the attack complexity. Another way to increase the complexity is to increase the complexity of hash function [7]. However, this also degrades the performance of normal password hashing.

In an on-line dictionary attack, an attacker constructs the dictionary after obtaining the target hash. Because a salt will be available to the attacker at this point, the attacker constructs a single dictionary relevant to the salt. Therefore, a salt is useless as a countermeasure to an on-line dictionary attack.

2.2 Homomorphic Encryption

Homomorphic encryption is an encryption scheme which enables arithmetic operations on ciphertexts [3-6]. Let us denote an encryption operation for a plaintext m using key K as $Enc(K; m)$. If an encryption scheme satisfies $Enc(K; m_1 * m_2) = Enc(K; m_1) * Enc(K; m_2)$ for some arithmetic operation $*$ with any plaintext messages m_1, m_2, it is called a homomorphic encryption scheme under operation $*$. If the arithmetic operation is addition (multiplication), the encryption scheme is called an additive (multiplicative) homomorphic encryption scheme. A fully homomorphic encryption (FHE) scheme is an encryption scheme which enables an arbitrary number of homomorphic additions and homomorphic multiplications on ciphertexts [3]. A somewhat homomorphic encryption (SHE) is an encryption scheme which enables only a limited number of homomorphic operations on ciphertexts.

Although FHE and SHE schemes are highly useful in various applications that compute on the encrypted domain, they are rather less efficient than additive homomorphic encryption (AHE) schemes such as the Paillier cryptosystem [6]. Recently, homomorphic encryption schemes have been used for various applications such as statistical operations [8, 9], biometric verification [10] and implicit authentication [11]. In this paper, we only need an additive homomorphic scheme. We denote a homomorphic addition by $\boxplus$. That is, $Enc(K; m_1 + m_2) = Enc(K; m_1) \boxplus Enc(K; m_2)$.

3 Proposed Method

3.1 System Model

To help readers understand the proposed method, we only explain the authentication scenario for smart devices, although our proposal can be applied to any situation where passwords are used. In our password-based authentication system, the user's password data are stored in a remote server instead of the target device. That is, there are three parties; user, device and remote server. To protect the user's password, a combination of a password hash and an additional digital signature will be stored in an encrypted form in the remote server and arithmetic operations required to password-based authentication will be done homomorphically on an encrypted domain also in the remote server. To be precise, the role of each party is as follows:

- A user is a party who remembers his/her password and wants to be authenticated by his/her device.
- A device is a party which verifies the identity of claimed user using the user's password. For this purpose, it stores cryptographic keys for homomorphic encryption and does all required encryption and decryption operations. In addition, it stores the public key for an additional digital signature used in authentication.
- A remote server is a party which stores the encrypted data required for authentication. It also stores the public key for homomorphic encryption. It performs homomorphic operations on the user's encrypted data. The server is

assumed to be honest but curious. Namely, it follows the protocol although it may try to learn information on the user's original password from stored ciphertexts.

3.2 Proposed Password-based Authentication Method

Fig. 1 compares traditional password-based authentication methods using cryptographic hash function and our proposed method. Unlike the traditional methods, our method stores encrypted password data and performs operations to verify the identity of the claimed user on ciphertexts.

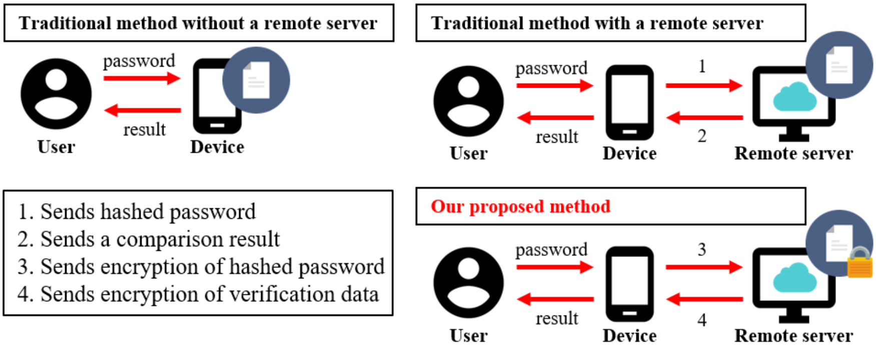

Traditional method without a remote server
password
result
User
Device
Traditional method with a remote server
password
result
1
2
User
Device
Remote server
1. Sends hashed password
2. Sends a comparison result
3. Sends encryption of hashed password
4. Sends encryption of verification data
Our proposed method
password
result
3
4
User
Device
Remote server

Fig. 1. Comparison of traditional password-based authentication and the proposed method.

The proposed method consists of two phases; registration and authentication, which are shown in Fig. 2 and Fig. 3, respectively. The details of these phases are as follows:

Registration (Fig. 2)**.** When the user starts a registration process, his/her device generates keys for homomorphic encryption, i.e., encryption key K_e and decryption key K_d. It also generates a key pair for digital signature, i.e., public key Pk_{sig} and private key Sk_{sig}. When the user enters his/her identity ID and the password PW into the device, the device performs a cryptographic hash function h using PW, and gets a k-bit output $p = (p_1, \dots, p_k)$. The device then performs a signing operation on authentication data u using the private key Sk_{sig}, and gets a k-bit signature $s = (s_1, \dots, s_k)$. The authentication data u is supplementary data that will be used in the authentication phase. Because a user's password will be verified by comparing the stored value of u, it should be sufficiently long, e.g., 128 bits. It may be device-dependent, or many devices can share the same value. In any case, u should not change over time.

After signature generation, the device removes the private key, which is different from a traditional digital signature. We will call this new type of signature a one-time private key-based signature. Removing the private key makes it impossible for the device to generate a signature any more. On the other hand, signature verification is still possible because the public key has not been removed.

The device combines p and s and produces $m_i = p_i \oplus s_i$, for $1 \le i \le k$. After encrypting each bit m_i with the homomorphic encryption key K_e, the device sends the ciphertexts to the remote server as well as K_e and the server stores them. If the Paillier cryptosystem is used as the underlying homomorphic encryption, the remote server should blind each ciphertext by adding a random even number for security. That is, the server updates C_i to $C_i \boxplus Enc(K_e; 2r_i)(1 \le i \le k)$ using random integers r_i.

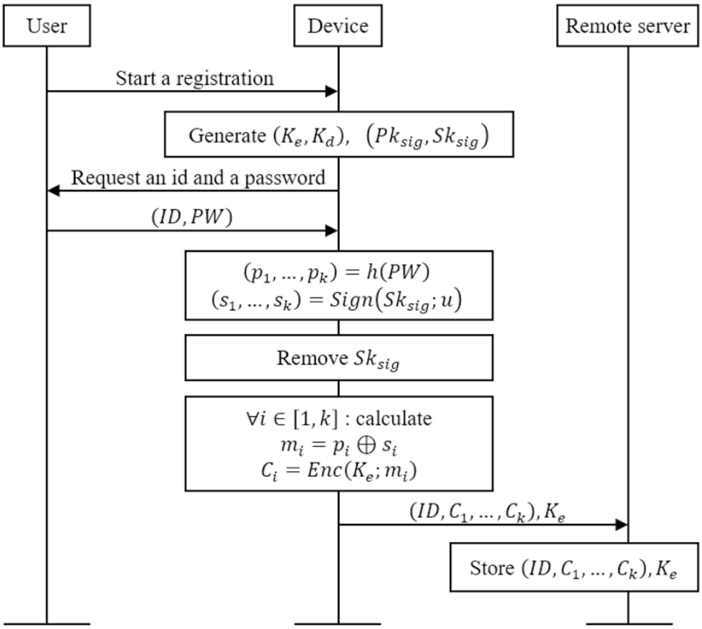

User
Device
Remote server
Start a registration
Generate (K_e, K_d), (Pk_{sig}, Sk_{sig})
Request an id and a password
(ID, PW)
$(p_1, \dots, p_k) = h(PW)$
$(s_1, \dots, s_k) = Sign(Sk_{sig}; u)$
Remove Sk_{sig}
$\forall i \in [1, k]$: calculate
$m_i = p_i \oplus s_i$
$C_i = Enc(K_e; m_i)$
$(ID, C_1, \dots, C_k), K_e$
Store $(ID, C_1, \dots, C_k), K_e$

Fig. 2. Registration phase of our proposed method.

Authentication (Fig. 3). When the user starts a process of authentication, the device requests his/her identity and password from the user. When the user enters his/her identity ID' and password PW' into the device, the device computes a k-bit hash value $p' = (p'_1, \dots, p'_k)$ from PW'. The device encrypts each p'_i using K_e, and obtains ciphertexts of p', $(C'_1, \dots, C'_k)$. The device sends $ID', C'_1, \dots, C'_k$ to the remote server. After receiving those ciphertexts, the remote server searches its database for a record such that $ID = ID'$. If this record does not exist, the remote server terminates the authentication and notifies the device of authentication failure. Otherwise, the server calculates $R_i = C_i \boxplus C'_i$ $(1 \le i \le k)$ on the encrypted domain, and sends the results to the device.

After receiving $R_1, \dots, R_k$, the device decrypts them using K_d, and gets $r_1, \dots, r_k$. If the Paillier cryptosystem is used as the underlying homomorphic encryption, each r_i is updated to $r_i \bmod 2$. Then, $r = (r_1, \dots, r_k)$ is a k-bit string. If everything is normal, r should be the same as signature s in the registration phase. (The reason for this will be explained in the next section.) Next, the device performs a verifying operation on the "signature" r and "message" u using the signature verification key Pk_{sig}. If the verification is successful, the device permits the user's access.

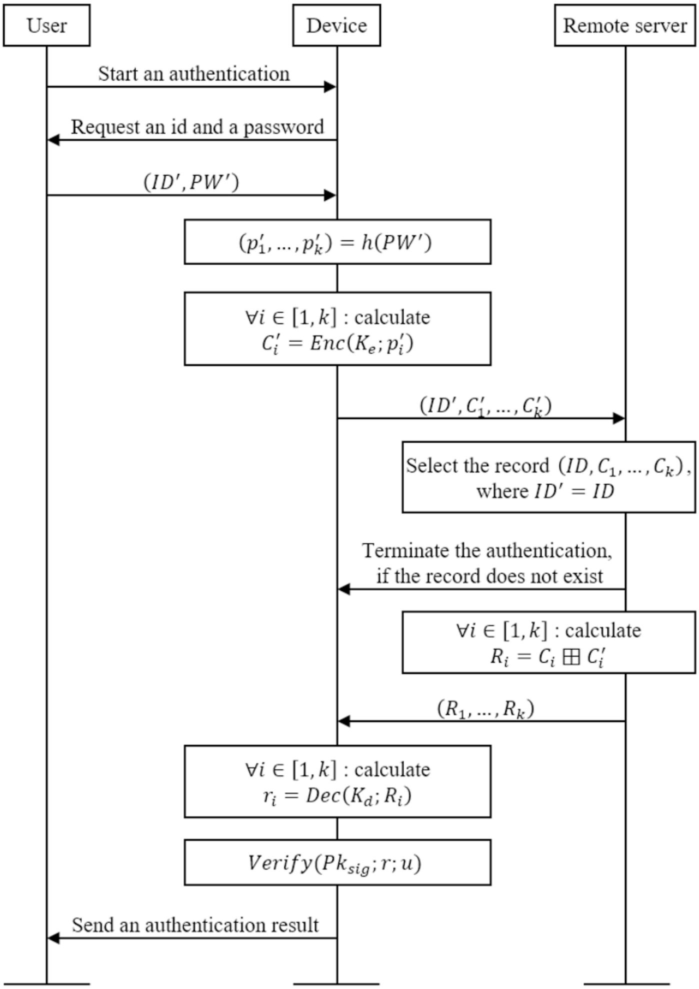

User
Device
Remote server
Start an authentication
Request an id and a password
(ID', PW')
$(p'_1, \ldots, p'_k) = h(PW')$
$\forall i \in [1, k]$: calculate $C'_i = Enc(K_e; p'_i)$
$(ID', C'_1, \ldots, C'_k)$
Select the record $(ID, C_1, \ldots, C_k)$, where $ID' = ID$
Terminate the authentication, if the record does not exist
$\forall i \in [1, k]$: calculate $R_i = C_i \boxplus C'_i$
$(R_1, \ldots, R_k)$
$\forall i \in [1, k]$: calculate $r_i = Dec(K_d; R_i)$
$Verify(Pk_{sig}; r; u)$
Send an authentication result

Fig. 3. Authentication phase of our proposed method.

4 Soundness and Completeness of Proposed Method

In this section, we show the correctness of the proposed method and analyze the security. First, we show that a legitimate user is always successfully authenticated. If a user enters a proper password, the equation $p_i = p_i'$ $(1 \leq i \leq k)$ is satisfied. Then,

$$r_i = Dec(K_d; R_i) = Dec\big(K_d; Enc(K_e; m_i) \boxplus Enc(K_e; p_i')\big)$$

$$= Dec(K_d; Enc(K_e; p_i \oplus s_i \oplus p_i') = Dec\big(K_d; Enc(K_e; s_i)\big) = s_i,$$

for $1 \leq i \leq k$. Therefore, the signature r is verified successfully.

Second, we see what happens if the device is compromised. Note that the device stores a homomorphic encryption key K_e, a decryption key K_d, authentication data u, and a signature verification key Pk_{sig}. With this information, the attacker may try to either recover a password hash or pass the authentication test even without the exact information on the hash. First, we examine the first possibility. Note that what the remote server does in the authentication phase is to return $Y_i = C_i \boxplus X_i$, given some value X_i. Because C_i contains the information on a password, an attacker may try to extract this information by generating X_i appropriately and decrypting Y_i. If $Dec(K_d;\ Y_i) = y_i$ and $Dec(K_d, X_i) = x_i$, then $Dec(K_d, C_i) = x_i \oplus y_i$. (For simplicity, we abused notations. The exact equation is $Dec(K_d, C_i) \equiv x_i + y_i$ mod 2.) Therefore, $p_i \oplus s_i = x_i \oplus y_i$, i.e., $p_i = (x_i \oplus y_i) \oplus s_i$, where s_i is one of the bits in $Sign(Sk_{sig}; u)$. If the attacker knows Sk_{sig}, s/he can compute s_i. Then, p_i is easily computed. However, because Sk_{sig} and s are not stored in the registration phase, there is no way for the attacker to find s_i. If s/he has no information on s_i, $p_i = (x_i \oplus y_i) \oplus s_i$ may be 0 or 1 with equal probability, 1/2. As a result, the attacker obtains no information on p_i. We remark that the communication channel between the device and server should be protected. If not, the attacker may obtain R_i by eavesdropping a normal authentication session. Then, by hacking into the device and obtaining K_d, s/he can recover s_i, and thus p_i. Now we examine if it is possible for an attacker to pass the authentication test without the exact information on the password hash. Note that the only way for him/her to do this is to know R_i. Because this is not possible according to the above assumption, an attacker cannot pass the test without a correct password hash.

Finally, we see what happens if the server is compromised. Note that the only information the server has is C_i $(1 \leq i \leq k)$, i.e., homomorphic ciphertexts instead of passwords or their hashed values. Because the attacker does not know the decryption key, s/he cannot recover the content, m_i. This also proves the fact that the honest-but-curious server does not learn any information on the password.

If the attacker hacks into both of the device and the remote server, the attacker can easily recover the password hash. However, it is hardly possible in real world.

5 Conclusion

In this paper, we proposed a password-based authentication method using one-time key-based digital signature and homomorphic encryption. We showed that our method is secure unless a very unusual event such that both a device and a remote server are compromised happens. It will be an interesting future work to implement our method for a real world application.

Acknowledgments. This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2016-H8501-16-1008) supervised by the IITP (Institute for Information & communications Technology Promotion).

References

1. Provos, N., Mazieres D.: A Future-Adaptable Password Scheme. In: USENIX Annual Technical Conference '16, FREENIX Track (1999)
2. Burr, W., Dodson, D., Newton, E., Perlner, R., Polk, W., Gupta, S., Nabbus, E.: Electronic Authentication Guideline. In: NIST Special Publication 800-63-2 (2013)
3. Gentry, C.: Fully Homomorphic Encryption Using Ideal Lattices. In: STOC '09, 169-178 (2010)
4. Dijk, M., Gentry, C., Halevi, S., Vaikuntanathan V.: Fully Homomorphic Encryption over the integers. In: EUROCRYPT '10, 24-42 (2010)
5. Brakerski, Z., Gentry, C., Vaikuntanathan, V.: (Leveled) Fully Homomorphic Encryption without Bootstrapping. In: ITCS '12, 309-325 (2012)
6. Paillier, P.: Public-Key Cryptosystems Based on Composite Degree Residuosity Classes. In: EUROCRYPT '99, 223-238 (1999)
7. Turan, M., Barker, E., Burr, W., Chen, L.: Recommendation for Password-Based Key Derivation. In: NIST Special Publication 800-132 (2010)
8. Graepel, T., Lauter, K., Naehrig, M.: ML Confidential: Machine Learning on Encrypted Data. In: ICISC '12, 1-21 (2012)
9. Naehrig, M., Lauter, K., Vaikuntanathan, V.: Can Homomorphic Encryption Be Practical?. In: CCSW '11, 113-124 (2011)
10. Im, J., Choi, J., Nyang, D., Lee, M.: Privacy-Preserving Palm Print Authentication using Homomorphic Encryption. In: IEEE DataCom '16, 878-881 (2016)
11. Shahandashti, S., Safavi-Naini, R., Safa, N.: Reconciling User Privacy And Implicit Authentication for Mobile Devices. Computers and Security, 53, 215-233 (2015)

On-line Voting System with Illegal Ballot Filtering Using Homomorphic Encryption

Mun-Kyu Lee, and Jong-Hyuk Im
Department of Computer and Information Engineering, Inha University
Incheon 22212, Korea
mklee@inha.ac.kr, imjhyuk@gmail.com

Abstract. A simple on-line voting protocol using homomorphic encryption is proposed. In addition to the basic properties required of a voting system, e. g., a voter's privacy, the system has additional functionalities such as automatic filtering of illegal ballot. Moreover, it is also possible that a voter who inadvertently cast a spoilt vote may vote again without revealing its secret ballot.

1 Introduction

Electronic voting, or on-line voting is a voting method where a computer or automated voting equipment is used [1] instead of a paper ballot. It aims at providing faster, more cost-effective, and more accurate voting. There have been a number of voting systems and protocols in the literature [2-9], each of which defines and satisfies various security goals. In this paper, we propose a new protocol for on-line voting with the property that an illegal or faulty vote can be filtered in real time without revealing its content and the voter of this vote can have the second opportunity to cast a correct vote. The proposed protocol uses homomorphic encryption schemes where degree-2 equations can be evaluated on the ciphertext domain.

2 Preliminaries

2.1 On-line Voting

A voting protocol is composed of voters, candidates, and the authority [1]. Voters are the participants that cast votes, candidates are the choices that voters can select, and the authority is an entity responsible for conducting the voting. There could be some adversary who attempts to manipulate the voting. An adversary may try to breach the privacy of voters or modify the result of voting. An adversary could be an internal one, or it could be an external one who cooperates with a voter.

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_46

There can be many different types of voting. For example, in a 1-out-of-2 (yes/no) voting, a voter chooses his/her answer from yes and no. In a 1-out-of-t voting, a voter chooses one of t candidates, while in a k-out-of-t voting, k candidates are chosen. In 1-out-of-t and k-out-of-t voting types, the candidate(s) chosen by the voter obtains 1 point, while other candidates obtain 0 point. The accumulated value over voters will be the final score for a specific candidate. Our proposal covers all the above types, but do not cover an ordered voting (multiple choices with preference) and a write-in voting where no candidates are given.

There has been an extensive research on on-line voting [2-9], and various security requirements for voting systems have also been proposed in previous works. The details for these requirements may be found in e.g., [1]. However, here we summarize the common requirements as follows:

- Completeness: the voting system should record and collect the votes correctly.
- Soundness: no voter can vote more than what is allowed. That is, s/he can only give 1 or 0 to candidates. S/he cannot give a value > 1 to his/her favorite candidate, nor a value < 0 to competitors.
- Privacy: no one, including even the authority, should be able to know any voter's choice.

The other important requirements contain authentication, integrity, non-reusability, auditability (accountability), etc. But we do not explicitly consider these properties in this paper because they can be achieved using legacy cryptographic techniques.

2.2 Homomorphic Encryption

Let $Enc(m)$ and $Dec(c)$ denote encryption and decryption operations over a plaintext message m and a ciphertext c, respectively. That is, $Dec\big(Enc(m)\big) = m$. A homomorphic encryption scheme is an encryption scheme where arithmetic operations can be properly done on its encrypted domain. For example, if an encryption scheme satisfies $Dec\big(Enc(m_1) + Enc(m_2)\big) = m_1 + m_2$ for any pair of plaintexts (m_1, m_2), then we say that this scheme is additively homomorphic. Similarly, if $Dec\big(Enc(m_1) \times Enc(m_2)\big) = m_1 \times m_2$, this scheme is multiplicatively homomorphic. If an encryption scheme is both additively and multiplicatively homomorphic for an arbitrary number of operations, it is called a fully homomorphic encryption scheme [10]. On the other hand, if it allows only a limited number of operations, we say it is somewhat homomorphic.

The idea of homomorphic encryption was first proposed by Rivest et al. [11] in 1978. Paillier proposed an additive homomorphic encryption scheme which provides homomorphic addition operations on an encrypted domain [12], and Boneh et al. proposed a homomorphic encryption which provides an arbitrary number of additions and depth-1 multiplications on ciphertexts, enabling evaluation of degree-2 equations [13]. In 2009, Gentry first proposed a fully homomorphic encryption scheme which provides an arbitrary number of additions and multiplications on an encrypted domain [10]. Since the fully homomorphic encryption scheme of Gentry, many works on fully homomorphic encryption schemes have appeared. While Gentry's scheme is based on the sparse subset sum problem over lattices, homomorphic encryption schemes based on the approximate GCD (Greatest Common Divisor) problem over the integers [14-16] and the (ring) learning with errors problem [17, 18] have been proposed.

Recently, Catalano and Fiore proposed a technique to transform an arbitrary additive homomorphic encryption scheme, e.g., the Paillier system, into a homomorphic encryption scheme which can evaluate degree-2 equations on ciphertexts [19].

3 Proposed Voting Protocol

The proposed voting protocol is composed of n voters ($n \geq 1$), t candidates ($t \geq 2$), and the authority as the traditional voting systems. In addition, it requires m intermediate collector(s) ($m \geq 1$). The protocol will be described for the 1-out-of-t voting, but it can be easily modified to the yes/no case, by regarding 'yes' and 'no' as the first and second candidates, respectively. It is also applicable to the k-out-of-t case, although we do not explain the detail in this paper. The voting protocol is composed of three stages; initialization, collection, and decision stages.

3.1 Initialization Stage

1. Let R be the authority. R generates a key pair composed of a private key S_R and a public key P_R for an underlying homomorphic encryption scheme, and sends them to voters as well as the information on t candidates. The homomorphic encryption we use is a somewhat homomorphic encryption with multiplication depth 1. That is, it should be able to deal with equations up to multiplicative degree 2 on its encrypted domain, e.g., $Dec(Enc(a) \times Enc(b) + Enc(c) \times (Enc(d) + Enc(e))) = ab + c(d + e)$. However, it is not guaranteed that $Dec\big(Enc(a) \times Enc(b) \times Enc(c)\big) = abc$, because the degree of this equation is 3. For this purpose, we may use the BGN scheme [13]. We may also use the somewhat homomorphic version of recently developed fully homomorphic encryption schemes such as integer-based schemes [14-16], LWE-based schemes [17, 18], and lattice-based schemes [10]. However, the most efficient one is to combine an additive homomorphic scheme such as the Paillier system with the recent general transformation technique which transforms an additive homomorphic scheme to a somewhat homomorphic scheme with multiplication depth 1 [19].
2. R decides the mapping between voters and collectors, i.e., which voter to send the ballots to which collector, and informs the corresponding voter and collector of this information.

3.2 Collection Stage

1. Each voter i $(1 \leq i \leq n)$ assigns 1 for his/her favorite candidate and 0 for all the other candidates. That is, if we denote voter i's vote on candidate j as v_{ij} and voter i's choice is candidate x, $v_{ix} = 1$ and $v_{ij} = 0$ for all $j \neq x$. If the voter does not want to choose anyone of the candidates, s/he may set $v_{ij} = 0$ for all j $(1 \leq j \leq t)$. The voter then encrypts each v_{ij} using the authority's public key P_R. As a result, s/he obtains

$$V_{ij} = Enc(P_R, v_{ij})\ (1 \le j \le t), \quad (1)$$

and sends them to the corresponding intermediate collector.

2. The intermediate collector $k \in \{1, \dots, m\}$ assigned for voter i computes the verification vector for voter i as
$$E_i = ((V_{i1} + \cdots + V_{it}) \times (V_{i1} + \cdots + V_{it} - 1),$$
$$V_{i1} \times (V_{i1} - 1), V_{i2} \times (V_{i2} - 1), \dots, V_{it} \times (V_{it} - 1)) \quad (2)$$
and sends it to the authority.
3. The authority collects all E_i for $1 \le i \le n$ and decrypts each element in $E_i = (E_{i0}, E_{i1}, E_{i2}, \dots, E_{it})$ using its private key, producing
$$c_{ij} = Dec(S_R, E_{ij}) \quad (3)$$
for $0 \le j \le t$. If there is any nonzero c_{ij}, this means that voter i cast an illegal vote. This could be a failed trial for breach, or it could be a just inadvertent one, which are indistinguishable. Anyway, the authority notifies the intermediate collector corresponding to voter i of this fact.
4. The corresponding collector may inform voter i to vote again to prevent a spoilt vote, if it is the policy of the authority.
5. Each intermediate collector $k\ (1 \le k \le m)$ aggregates the encrypted votes (except the spoilt ones), and computes
$$U_1^{(k)} = \sum_i V_{i1},\ U_2^{(k)} = \sum_i V_{i2},\ \dots,\ U_t^{(k)} = \sum_i V_{it}. \quad (4)$$
The collector then sends theses values to the authority.

3.3 Decision Stage

1. Now the authority has m sets of encrypted aggregation,
$$(U_1^{(1)}, U_2^{(1)}, \dots, U_t^{(1)}),, \dots, (U_1^{(m)}, U_2^{(m)}, \dots, U_t^{(m)}).$$
2. It computes
$$W_i = \sum_{k=1}^{m} U_i^{(k)} \quad (5)$$
for $1 \le i \le t$.
3. It then decrypts each W_i using its private key S_R and obtains
$$w_i = Dec(S_R, W_i). \quad (6)$$
4. Finally, it finds the index z such that w_z is the maximum among w_i. Candidate z is elected.

3.4 Example

Let us assume that there are $t = 3$ candidates, $n = 100$ voters, and $m = 4$ intermediate collectors. Let's say, voter 5 is assigned to collector 2, and he wants to vote for candidate 3. Then, he may set $(v_{51}, v_{52}, v_{53}) = (0,0,1)$ and send its component-wise encryption to collector 2. (This choice will pass the filtering procedure in step 3 of the collection stage, because $c_{50} = c_{51} = c_{52} = c_{53} = 0$. For more details, see section 4.2.) Let's assume that collector 2 has 20 voters assigned to it, including voter 5. Then it will receive 20 tuples of (V_{i1}, V_{i2}, V_{i3}). If no illegal vote was found in steps 2 to 4 of the collection stage, it will compute $U_1^{(2)}$, $U_2^{(2)}$, and

$U_3^{(2)}$ by accumulating those 20 tuples. In step 2 of the decision stage, the authority will compute $W_1 = \sum_{k=1}^{4} U_1^{(k)}$ using the values from four collectors. It also computes W_2 and W_3 in a similar manner. Finally, it decides the winner by finding the W_i decrypted to the maximum.

4 Security Analysis

In this section, we show that the proposed voting protocol satisfies the security requirements mentioned in section 2.1.

4.1 Completeness

According to the additive homomorphic property of the underlying encryption scheme, combining (1), (4), (5) and (6), we obtain

$$w_j = Dec(S_R, W_j) = Dec\left(S_R, \sum_{k=1}^{m} U_j^{(k)}\right) = Dec\left(S_R, \sum_{all\ i} V_{ij}\right)$$

$$= \sum_{all\ i} Dec(S_R, V_{ij}) = \sum_{all\ i} v_{ij}$$

for $1 \le j \le t$. Because $v_{ij} = 0$ (for preference) or 1 (for non-preference), w_j represents the exact number of voters who voted for candidate j. It is easy to see that $\sum_{j=1}^{t} w_j \le n$, where the inequality is for the case where (1) either some of the voters did not choose any candidate, i.e., gave 0 to all candidates, or (2) some of the votes were not counted because it was filtered in step 3 of the collection stage. The latter only happens if the authority's policy does not give another chance for voting to a voter who cast a spoilt vote.

4.2 Soundness

It is sufficient to show that a voter cannot give a value > 1 to his/her favorite candidate, nor a value < 0 to competitors. We achieve this goal by establishing the following lemma.

Lemma 1. If all $c_{ij} = 0$ for $0 \le j \le t$, one of the following two cases holds:

1) all v_{ij} are 0 for $1 \le j \le t$, or

2) $v_{iJ} = 1$ for some $J \in \{1, \dots, t\}$, and $v_{ij} = 0$ for all $j \ne J$.

Proof. Because we assumed that the underlying encryption scheme provides a depth-1 multiplication, (2) and (3) implies $c_{i0} = Dec(S_R, E_{i0}) = (v_{i1} + \cdots + v_{it}) \times (v_{i1} + \cdots + v_{it} - 1)$. Thus, $c_{i0} = 0$ implies that $v_{i1} + \cdots + v_{it}$ is either 0 or 1. Similarly, for $1 \le j \le t$, $c_{ij} = 0$ implies that v_{ij} is either 0 or 1. Because v_{ij} is either 0 or 1 for all $1 \le j \le t$, we can enumerate all 2^t combinations of $(v_{i1}, \dots, v_{it}) \in \{0,1\}^t$.

Among these, any combination with more than two nonzero v_{ij}, i.e., more than two 1's, produces $v_{i1} + \cdots + v_{it} > 1$, which contradicts the condition that $c_{i0} = 0$. Then, the only remaining possibilities are the two cases mentioned in the lemma. Finally, it is easy to see that these two cases make $c_{ij} = 0$ for $0 \leq j \leq t$. To be precise, the first case makes $v_{i1} + \cdots + v_{it} = 0$ and the second makes $v_{i1} + \cdots + v_{it} = 1$. This proves the lemma.

The first case in lemma 1 is for the case that voter i did not choose any candidate, which is reasonable if s/he cannot make a decision. The second case stands for the normal situation where the voter selected only one candidate. If it is not guaranteed that a voter can only give 1 or 0 to candidates, the election may be corrupted. For example, s/he might try to give a huge score to let her favorite candidate elected irrespective of other voters' choices, or s/he might give a negative score to let her enemy fail to be elected. In addition, it can also be detected if a voter tries to choose multiple candidates. However, by slightly modifying the equation for E_{i0}, we can also support the k-out-of-t voting where a voter can select up to k candidates. Finally, we remark that the detection of illegal votes based on the values of c_{ij} only reveals that there is something wrong, but it does not discriminate which is the case among the above three possibilities; a huge score, a negative score, or multiple choices.

4.3 Privacy

The privacy of voters is guaranteed by aggregation. In step 5 of the collection stage, each intermediate collector sends the authority only the aggregated value of the votes that it collected. Even when the authority tries to separately decrypt $\left(U_1^{(j)}, U_2^{(j)}, \ldots, U_t^{(j)}\right)$ for $1 \leq j \leq m$ in step 3 of the decision stage instead of their aggregated values $W_1, W_2, \ldots, W_t$, the authority cannot know the choice of a specific voter. What the authority learns is the sum of votes for each candidate from all voters connected to a specific collector. For example, by decrypting $U_1^{(k)}$, the authority will recover $\sum_{(for\ all\ i\ connected\ to\ collector\ k)} v_{i1}$. That is, the collector finds out how many people in the community managed by collector k voted for candidate 1, but not who did. This situation is exactly the same as the way the secret ballot principle is observed in a local polling station for an off-line voting. We remark that, however, the privacy may not be protected if a collector and the authority colludes.

4.4 Support for Second Voting

If the authority adopts the policy that it allows for failed voters to vote again, it may raise the ratio of valid votes without harming the completeness and soundness. Moreover, the failed voter's privacy is guaranteed even in the spoilt vote and the second vote.

4 Security Analysis

We presented a new protocol for on-line voting with the property that an illegal or faulty vote can be filtered in real time without revealing its content. Then, the voter of this vote may have the second opportunity to cast a correct vote. Our protocol uses somewhat homomorphic encryption schemes where degree-2 equations may be evaluated on the ciphertext domain. The proposed protocol guarantees the voters' privacy, and at the same time, it prevents voters from giving illegal scores to candidates.

Acknowledgments. This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2016-H8501-16-1008) supervised by the IITP (Institute for Information & communications Technology Promotion).

References

1. Mursi, M., Aggassa, G., Abdelhafez, A., Sarma, K.: On the Development of Electronic Voting: A Survey. In: International Journal of Computer Applications, 61(16), 1-11 (2013)
2. Chaum, D.: Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms. In: Communications of the ACM, 24(2), 84-88 (1981)
3. Sako, K., Killian, J.: Receipt-Free Mix-Type Voting Scheme: A Practical Solution to The Implementation of A Voting Booth. In: EUROCRYPT '95, 393–403 (1995)
4. Benaloh, J.: Verifiable Secret-Ballot Elections. In: Ph.D. thesis, Yale University (1987)
5. Cramer, R., Gennaro, R., Schoenmakers, B.: A Secure and Optimally Efficient Multi-Authority Election Scheme. In: EUROCRYPT '97, 103-118 (1997)
6. Schoenmakers, B.: A Simple Publicly Verifiable Secret Sharing Scheme and Its Applications to Electronic Voting. In: CRYPTO '99, 148-164 (1999)
7. Fujioka, A., Okamoto, T., Ohta, K.: A Practical Secret Voting Scheme for Large Scale Elections. In: AUSCRYPT '92, 248–259 (1992)
8. Park, C., Itoh, K., Kurosawa, K: Efficient Anonymous Channel and all/nothing Election Scheme. In: EUROCRYT '93, 248–259 (1993)
9. Juels, A., Catalano, D., Jakobsson, M.: Coercion-Resistant Electronic Elections. In: WEPS '05, 61-70 (2005)
10. Gentry, C.: Fully Homomorphic Encryption Using Ideal Lattices. In: STOC ' 09, 169-178 (2010)
11. Rivest, R., Adleman, L., Dertouzos, M.: On Data Bank and Privacy Homomorphisms. In: Proceedings of the 19th Annual Symposium on Foundations of Secure Computation-FSC 1978, Academic Press, 169-180 (1978)
12. Paillier, P.: Public-Key Cryptosystems based on Composite Degree Residuosity Classes. In: EUROCRYPT '99, 223–238 (1999)
13. Boneh, D., Goh, E., Nissim, K.: Evaluating 2-DNF formulas on ciphertexts. In: Theory of cryptography, 3378, 325-341, (2005)
14. Dijk, M., Gentry, C., Halevi, S., Vaikuntanathan, V.: Fully Homomorphic Encryption over the Integers. In: EUROCRYPT '10, 24-42 (2010)
15. Coron, J., Naccache, D., Tibouchi, M.: Public Key Compression and Modulus Switching for Fully Homomorphic Encryption over the Integers. In: EUROCRYPT '12, 446-464 (2012)
16. Cheon, J., Coron, J., Kim, J., Lee, M., Lepoint, T., Tibouchi, M., Yun, A.: Batch Fully Homomorphic Encryption over the Integers. In: EUROCRYPT '13, 315-335 (2013)

17. Brakerski, Z., Gentry, C., Vaikuntanathan, V.: (Leveled) Fully Homomorphic Encryption without Bootstrapping. In: ITCS '12, 309-325 (2012)
18. Clear, M., McGoldrick, C.: Multi-Identity and Multi-Key Leveled FHE from Learning with Errors. In: CRYPTO '15, 630-656 (2015)
19. Catalano, D., Fiore, D.: Using Linearly-Homomorphic Encryption to Evaluate Degree-2 Functions on Encrypted Data. In: ACM-CCS '15, 1518-1529 (2015)

Real-Time Malicious Script Blocking Technology at the Host-Level

SangHwan Oh, HanChul Bae, Seongmin Park, HwanKuk Kim (Author)
Security R&D Team 2
Korea Internet & Security Agency
Seoul, Korea
osh1, hcbae, smpark, rinyfeel@kisa.or.kr

Abstract. Due to the diversity of mobile devices, interests have been increased towards HTML5, the next generation's web standard which pursues cross platform. To play media files or process 3D graphics in previous HTML environments, users had to install non-standard plug-ins such as Silverlight or Active X. On the other hand, HTML5 provides new tag functions of audio, video etc and new java script functions of Websocket, Geolocation API etc to substitute non-standard technologies such as Active X. To use such functions of HTML5, web browser developers are competitively applying HTML5 to their browsers, making the active conversion to HTML5 a global trend of today. Along with such trend, however, the risk of new cyber attacks taking advantage of java scripts, the key function of HTML5, is also increasing. Cyber attacks based on scripts can trigger vicious actions when the user just accesses web pages inserted with vicious scripts, and thus there are limits in detection using previous security technologies. This paper proposes a technology which collects and analyzes HTTP traffic generated through web browsers at host level to detect and block vicious scripts.

1 Introduction

Services provided through the online environment are growing more diversified due to the fast dissemination of the Internet. Traditional HTML could provide only static services, and non-standard plug-ins like Active X were required to provide dynamic service, which can cause security. The most powerful method to resolve the issue is the next-generation web standard HTML5, which was announced in October 2015. [1] Various functions were added to HTML5 that could replace non-standard plug-ins but retained compatibility with traditional HTML. For instance, HTML5 controls media using new tags: video, audio and Canvas that replace Adobe Flash. Moreover, it enables implementation of multi-thread (Web Worker), web socket communication and utilization of location information by using additional JavaScript APIs. In addition, HTML5 supports cross-platform, and Gartner has selected HTML5 as one of its top 10 mobile technologies. [2] In line with the trend, the world's leading web browser developers are rushing to upgrade their browsers to support HTML5.

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_47

Google has set its policy to make mandatory use of an HTML5 player to play video on its browsers, Chrome and FireFox. In such an environment in which conversion to HTML 5 becomes mandatory, however, security threats to HTML5 are increasing. XSS (Cross Site Scripting) attack using new tag uses video and audio that can pass through XSS attack detection filter, which uses pattern matching. In particular, JavaScript, the core function of the new features, is facing more threats. The cyberattack on the Korean government and public institutions on June 25, 2013, was an example of a "script-based cyberattack" using JavaScript. Additionally, the script-based attacks took place in the Chinese video sharing site SOHU TV in 2014 and the open source share site GitHub in 2015. In a traditional malicious code attack, the codes were required to be downloaded and installed on a user PC to perform malicious behavior and existing detection technologies could detect and block those malicious codes. For a script-based cyberattack, however, since JavaScript is running on web browsers only, once the browser is closed, no trace remains on a user PC, which makes difficult for existing security technologies to deal with it. And since most JavaScript is obfuscated to protect developers' ideas or improve performance, attackers also obfuscate their malicious script, which can easily pass through traditional signature-based static detection technology. Thus this study describes technology to detect a script-based cyberattack at the host level. Chapter 2 describes earlier research on technologies to detect web attacks, Chapter 3 features technology to detect and block malicious script through local proxy, Chapter 4 analyzing the result of applying the processed technology, and Chapter 5 presents the conclusion and future direction of research.

2 Paper Preparation

Script-based cyberattack is web-based attach which is made through web. Responsive method to the web-based cyberattack can be mainly categorized into two; network level and client level method. The paper describes client level method which is related to this study, and JavaScript obfuscation technology which is one of the biggest issue in script-based cyberattack.

Client level technology

There are currently a number of on-going researches on client level technologies to respond to web-based cyberattack. The most responsive technology among those is one used in WebCheck[3], SiteAdvisor[4], etc that inspects URLs connected on browser. The technology blocks access to malicious web sites by matching URL to access and black list of URLs with history of malicious code distribution. However, there is still difficulty to deal with the attack through only this method as attackers frequently change their URLs.

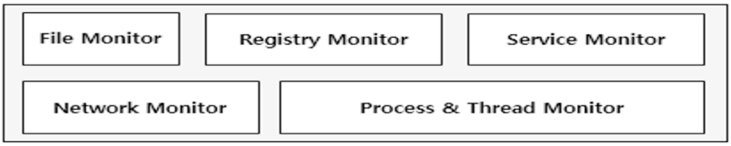


Fig. 1. Configuration of Observer

To come over with the limits, technology to monitor web browser has been developed as it is start-point of attack. [5] As show in [Fig 1] the technology suggested method using malicious behavior detection module called Observer. As show in [Fig 2] It applies process unit in browser via Observer module consisting of File Monitor, Network Monitor, Process & Thread Monitor, etc and collects events from the processes via API hooking to detect malicious behavior. As monitoring web browser itself, while the technology is useful to respond to web-based attack, it makes high loads as it creates Observer and monitors all processes in browser. On the other hands, the technology requires high specification client device which is major weak point in client level which should be operated in various environments.

Therefore, the paper describes technology with light in accordance with client environment, but powerful detection function.

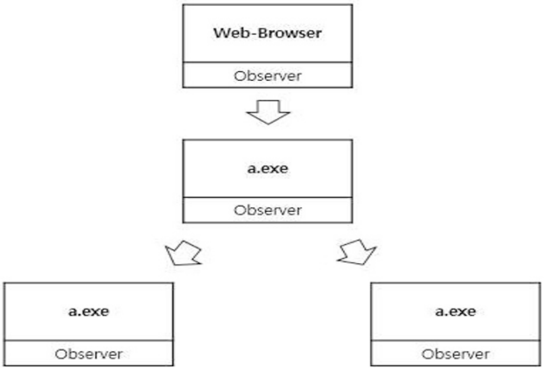


Fig. 2. Observers attached to the each of the processes

JavaScript obfuscation

Along with diversification of web-based cyberattack, various technologies responding to the attack have been developed. The most frequently used technology is signature-based static analysis technology which extracts pattern of malicious codes and analyzes them via pattern matching with analysis target. Attackers are trying to pass through the pattern matching based static analysis technology by generating new malicious codes via obfuscation.

In script based attack, they evade signature-based method by obfuscating JavaScripts in various ways. The most common methods are string split which splits malicious code into a number of string variables and combines them when executing and method to convert character set to ASCII code by using JavaScripts' own functions (escape, and unescape). IN addition to those methods, attackers try to disarm the existing security technologies by using XOR encoding or applying their own encoding functions. Therefore, the study provides method to de-obfuscate the JavaScripts.

3 Proposed Technology

As show in [Fig 3], the technology proposed in this paper is composed by a collection module realized in the form of Local Proxy at host level, and an analysis agent module which decides viciousness and sets the processing policy afterwards. This technology is installed and operated in user PC to collect and analyze HTTP packets generated in web browsers. When decided to be vicious, the technology protects user PC from script-based attacks by deleting the vicious script or by using post-processing methods such as redirecting to a safe page etc.

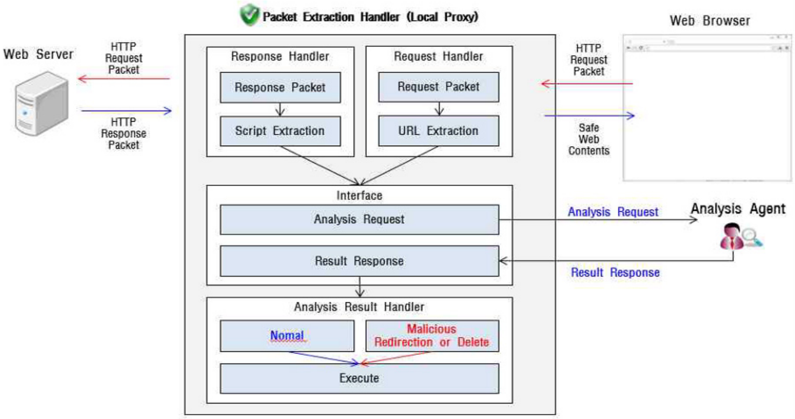


Fig. 3. Propsed Technology architecture

Packet collection module

First, as the collection module for the subjects of analysis, Tinyproxy, a famous light-weight proxy daemon among open-source SWs, was used. [6] To collect HTTP packets using Local-Proxy, registry value related to the proxy setup of Windows was changed as [Table 1]. By doing so, all packets generated from web browsers were received and transmitted through the collection module.

Table 1. Proxy settings related Windows Registry values

Windows Registry Values
HKEY_CURRENT_USER₩Software₩Microsoft₩Windows₩CurrentVersion₩ProxyEnable
HKEY_CURRENT_USER₩Software₩Microsoft₩Windows₩CurrentVersion₩ProxyServer

By collecting HTTP Request packets sent to external web server from the web browser of user, access URL information is extracted. Then, request is made to the analysis agent. If the analysis agent decides that the relevant URL is safe, the packets are sent to the web server. Afterwards, script codes are extracted in tag units from HTTP Response packets sent by the web server. Among the extracted scripts, external scripts where script codes are located outside using src property etc are sent to web browser, so that the browser can request again the relevant script. By doing so, external script codes are extracted from the Response packets. Script codes collected as above are requested to the analysis agent for analysis in tag units. Based on the result of analysis, packets which went through post-processing are sent from the analysis agent to the web browser to protect the web browser of user PC from vicious scripts.

Analysis agent module

As shown in [Fig 4], the analysis agent module mainly performs real-time analysis and precise analysis. Real-time analysis is divided into URL test and static analysis of scripts. Viciousness is decided based on the blacklist of access URL [5] extracted from HTTP Request packets generated in web browser. If decided as vicious URL, information on the access to vicious URL is noticed to user. If the user blocks the access, the module redirects to a safe page and finishes the analysis. If the URL is normal, HTTP Request packets are sent to the relevant web server, and scripts are extracted from HTTP Response packets received from the web server to conduct static analysis of scripts.

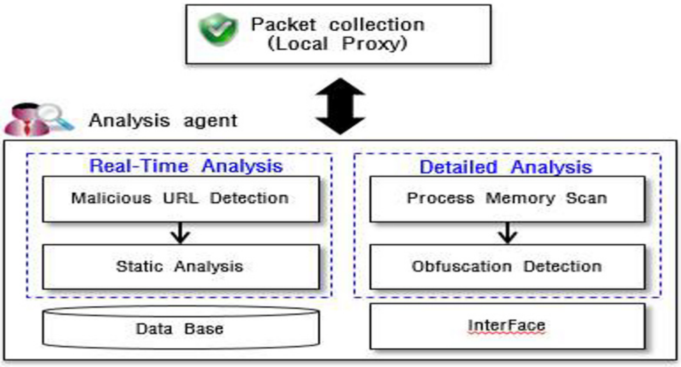


Fig. 4. Analysis Agent Module

Static analysis decides viciousness through signature pattern matching based on YARA[7]. YARA is a technology which searches and classifies vicious codes based on character strings or binary patterns. Its most significant characteristic is the ability to generate Rule Sets which can contain all kinds of expressions. Using YARA Rule, the pattern information of vicious scripts' codes generated before is extracted and used as signature for static analysis. Script codes extracted from HTTP Request packets go through pattern matching based on YARA Rule to decide for the existence of vicious scripts. However, since obfuscated vicious scripts change the patterns themselves into other character strings, it is difficult to detect them using static analysis based on signature pattern matching. To detect such obfuscated vicious scripts which can detour static analysis, the technology of precise analysis has been additionally introduced. Precise analysis is divided into the first step of memory scan which extracts scripts where obfuscation is removed and the second step of obfuscation analysis where obfuscation is removed to extract the original. Memory scan extracts PID of web browser which is currently running, and scans the memory of the relevant browser process to extract scripts currently running in the browser. At this point, scripts where obfuscation is removed can be secured as the first step. Scripts where obfuscation is not removed from the first step are applied with V8 java script interpreter [8] used in Chrome Browser to remove such obfuscation. After running the obfuscated scripts on V8 interpreter, BreakPoint is set at a certain point to extract script codes. By doing so, script codes where obfuscation is removed can be secured. Through signature pattern matching used in static analysis, these script codes are analyzed to see if they are vicious.

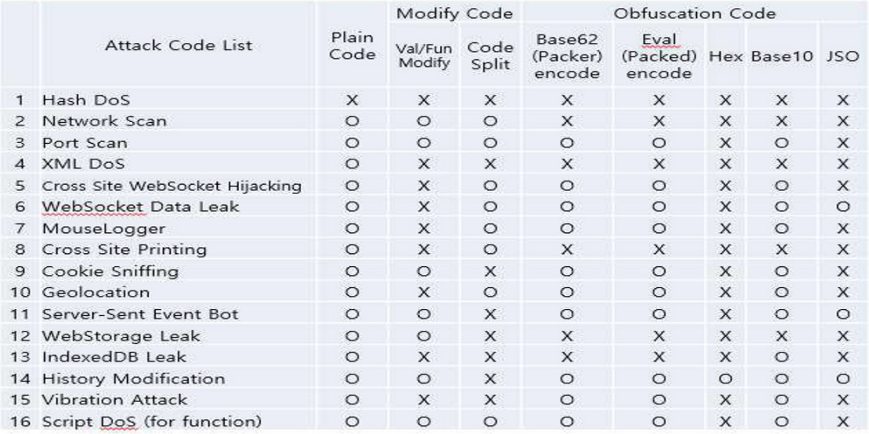
4 Analyzing the result of applying the proposed technology

In order to check the real-time static analysis of script based on YARA Rule, a technology proposed in this paper, as well as the decryption function for obfuscated script using V8 java script interpreter, the detection in 128 samples was checked. These 128 samples were made by applying eight changes such as code division, change of function name, obfuscation etc to 16 types of vicious script attack samples.

Real-time static analysis of script

Static analysis of script determines viciousness by matching major keywords in vicious script code based on YARA Rule. Therefore, detection was checked in 16 original attack samples and 112 samples where detours can be made by changing keywords based on the change of function/variable names. As shown in [Table 2], the detection rate in original samples was more than 95% since the major keywords in vicious scripts were all included in the scripts subjected for analysis. However, in samples which detoured pattern matching through the change of function/variable names, code division, obfuscation etc, the detection rate was less than 50%. In case of changing function/variable names, keywords themselves were changed to detour pattern matching, and in case of code division, detection was difficult since vicious keywords were divided into two or more script tags while the technology of this paper performs analysis in the unit of script tag (<script>…</script>). In case of obfuscation, high detection rate was shown for Base62, Base10, Eval function encoding since there was no change on the keywords themselves. However, for obfuscation samples such as Hex or JSO where the keywords were changed, detection failed.

Table 2. Result of Static Analysis



	Attack Code List	Plain Code	Modify Code		Obfuscation Code				
			Val/Fun Modify	Code Split	Base62 (Packer) encode	Eval (Packed) encode	Hex	Base10	JSO
1	Hash DoS	X	X	X	X	X	X	X	X
2	Network Scan	O	O	O	X	X	X	X	X
3	Port Scan	O	O	O	O	O	X	O	X
4	XML DoS	O	X	X	X	X	X	X	X
5	Cross Site WebSocket Hijacking	O	X	O	O	O	X	O	X
6	WebSocket Data Leak	O	X	O	O	O	X	O	O
7	MouseLogger	O	X	O	O	O	X	O	X
8	Cross Site Printing	O	X	O	X	X	X	X	X
9	Cookie Sniffing	O	O	X	O	O	X	O	X
10	Geolocation	O	X	O	O	O	X	O	X
11	Server-Sent Event Bot	O	O	X	O	O	X	O	O
12	WebStorage Leak	O	O	X	X	X	X	X	X
13	IndexedDB Leak	O	X	X	X	X	X	O	X
14	History Modification	O	O	X	O	O	O	O	O
15	Vibration Attack	O	X	X	O	O	X	O	X
16	Script DoS (for function)	O	O	O	O	O	X	O	X

Decryption of obfuscated script

By decrypting obfuscated scripts using V8 java script interpreter, the extraction of original scripts was checked. Most of the original scripts were extracted for Base62, Base10, Eval function encoding samples which used exclusive tools for obfuscation. However, obfuscation methods such as Hex or JSO were impossible to decrypt. Different from previous methods, Hex or JSO has been confirmed to apply two or more obfuscation methods, making V8 java script interpreter proposed by this paper impossible to extract original codes at BreakPoint.

5 CONCLUSION

Due to the diversity of mobile devices such as laptop PCs, tablet PCs etc, interests have been increased towards HTML5 which pursues cross platform. Along with such increased interests, however, the risk of cyber attacks taking advantage of reinforced java scripts is also increasing. This paper proposed a technology that can deal with cyber attacks taking advantage of java scripts in user PC. The technology can detect cyber attacks based on scripts which are difficult to detect using previous security solutions. The paper also proposed detection methods for obfuscated scripts that detour security technologies based on pattern matching. The technology proposed in this paper can be used to increase detection rate by linking with detection technologies for vicious codes proposed previously. Also, by combining the technology with AP equipments, AP with security functions can be created. Still, the environment of technology proposed in this paper is limited to user PC, and there are also improvements to be made on the analysis technology. First, to prepare for the era of HTML5 which pursues cross platform, technological researches on detecting vicious scripts in mobile environment should be conducted. Also, as the analysis of the result of applying the proposed technology suggests in Paragraph 4, researches should be conducted to use URL instead of script tag as analysis unit so that vicious scripts divided onto two or more tags can be detected. Finally, regarding the analysis of obfuscation, codes using well-known obfuscation tools were decrypted, while codes applied with two or more obfuscation methods were difficult to decrypt. In the future, we are planning to solve the issue of obfuscation by studying algorithms related to decryption.

Acknowledgments. The study has been conducted as a part of ICT R&D project sponsored by Ministry of Science, ICT and Future Planning and Institute for Information & communications Technology Promotion. [B0101-15-0230, Development of preventative and responsive technologies for script based cyberattack]

References

1. W3C, “HTML5 Standard”, April 20 2015, http://www.w3.org/standards
2. Gartner, ″Gartner Identifies Top 10 Mobile Technologies and Capabilities for 2015 and 2016″, February 24 2014.
3. KISA, WebCheck. Available : http://webcheck.kisa.or.kr
4. McAfee. SiteAdviser. Available : http://www.siteadvisor.com
5. Young-wook Lee, Dong-jae Jeong, Sang-hoon Jeon, and Chae-ho Im, ″Design and Implementation of Web-browser based Malicious behavior Detection System(WMDS)″ Journal of Information Security & Cryptology June, 2012
6. Tinyproxy, https://tinyproxy.github.io/
7. YARA Documentation, http://yara.readthedocs.org/en/latest/index.html
8. Chrome V8, https://developers.google.com/v8

A Probabilistic Encounter and Distance-based Routing Protocol for Opportunistic Networks

Sanjay K. Dhurandher, Satya J. Borah, Isaac Woungang, Sahil Gupta, Pragya Kuchal,Makoto Takizawa, Leonard Barolli

Abstract An Opportunistic Network (OppNet) is one of the latest domain of wireless communication where information is transferred from the source to the destination without any infrastructure, internet connectivity and any predefined network topology. The mobile nodes participating in the network contribute in establishing a connectivity between the nodes to transmit the information from the source to the destination using a store-carry and forward mechanism. The nodes store the information in their in-built buffer until a suitable forwarder is available within their wi-fi/bluetooth transmission range. Designing a routing protocol in OppNet is thus a challenging task. This paper proposes a Probabilistic Encounter and Distance-based Routing Protocol (P-EDR) for Opportunistic Networks, which is designed as a combination the ProPHet and the Encounter and Distance-based Routing (EDR) protocols. Simulation results are presented, showing that P-EDR outperforms the EDR, History-based Prediction for Routing (HBPR) and ProPHet routing protocols in terms of message delivery probability, messages overhead ratio and number of messages dropped.

1 Introduction

An Opportunistic Network (OppNet) [1] consists of wireless mobile devices where the probability of existence of an end-to-end path from a source to a destination is almost zero. In OppNets, the initial seed OppNet grows into an expanded OppNet by considering the foreign nodes and neighbouring seed nodes. Since most of the nodes in OppNet are mobile, the connectivity of the network is maintained by nodes only when they come within the transmission ranges of each other. In an OppNet, nodes communicate with each other with the help of their built-in wi-fi/bluetooth and. If a node has a message copy to send but it is not connected to another node, it stores the message in its buffer until an appropriate communication opportunity arises, i.e. it uses the store-carry-forward paradigm to forward the message from the source to the destination. This is why OppNets are often referred to as a subclass of Delay-Tolerant Networks [2].

The routing process in OppNet is completely different from that of traditional networks. In traditional routing protocols, there exists a path for the message from the source to the destination, that is, the communication

Sanjay K. Dhurandher, Satya J. Borah
Division of Information Technology
NSIT, University of Delhi, Delhi, India,
e-mail: dhurandher@gmail.com, satyaborah@yahoo.co.in

Isaac Woungang
Department of Computer Science
Ryerson University, Toronto, ON., Canada
e-mail: iwoungan@scs.ryerson.ca

Sahil Gupta
Department of Computer Science and Engineering,
Indraprastha Institute of Information Technology (IIIT), Delhi, India
e-mail: sahil15057@iiitd.ac.in

Pragya Kuchal
Division of Information Technology
NSIT, University of Delhi, Delhi, India,
e-mail: pragya.kuchhal@gmail.com

Makoto Takizawa
Department of Advanced Sciences
Hosei University, Tokyo, Japan
e-mail: makoto.takizawa@computer.org

Leonard Barolli
Department of Information & Comm. Eng.
Fukuoka Institute of Technology (FIT), Japan
e-mail: barolli@fit.ac.jp

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_48

end-points are always connected. In this case, if a destination is not found, a link failure or network failure will occur. Further effort is also performed to guarantee a future delivery of the message. In OppNets, the message could be transferred between the network devices using the connection opportunities among the nodes which are within the radio range of each other. For efficient and guaranteed communication, the OppNets take into account the contact duration and the inter-contact time between the nodes. In the absence of a transmission opportunity, the data must be stored in the buffer of an intermediate node. If the inter-contact times increase, the storage period of the message in the buffer will increase, which requires a careful buffer management, especially, when the queues are full.

This paper proposes a routing technique called P-EDR that uses our recently proposed Encounter and Distance-based Routing (EDR) [3] as underlying routing protocol. In this newer protocol, the probability of a node to meet its destination is evaluated and then applied over the set of nodes obtained from the EDR protocol in order to select the forwarding node. In a nutshell, the proposed P-EDR is a history-based multi-copy routing protocol.

The rest of the paper is organized as follows. in Section II, representative routing protocols for OppNets are overvieved. In Section III, the proposed P-EDR protocol is described. In Section IV, simulation results are presented. Finally, Section V concludes the paper.

2 Related Work

Many routing algorithms for OppNets have been proposed in the literature. Representative ones are described in [3-10]. In Epidemic [4], considered as a kind of dissemination based routing protocol, the messages are flooded to the nodes that are in the range of the source or intermediate node. In [5], the spray-and-wait routing protocol is introduced, which is divided into two phases: spray and wait. In the spray phase, instead of flooding all the messages, the source node sprays or forwards the message copy to L relay nodes. If the destination is not found during the spraying phase, then the relay nodes carrying the copy of the message perform the direct transmission with the destination. In [7], a hybrid protocol called PRoWait is presented, which uses a simple forwarding strategy. For selecting the neighbour node, the delivery predictability of a node is calculated from ProPHet routing protocol and spraying of the packets to the neighboring nodes is done with the spray and wait protocol. In History based routing protocol HiBOp [9] for opportunistic network utilizes a nodes present context to find a better path for message delivery. The current context of a node is a snapshot of the environment it currently resides in. HiBOp uses and stores all the possible context information available about a node that is generally very hard to find. Similarly, in the genetic algorithm-based energy-efficient routing (GAER) protocol [10] for OppNets, the genetic algorithm (GA) is used to route the message from the source to destination. In this section the brief description of ProPHet, HBPR and EDR protocol have been presented.

2.1 ProPHet

In ProPHet [6], each node, before relaying a message, estimates a probability metric called delivery predictability for each known destination. The calculation is based on the history of encounters between nodes or the history of visits to certain locations. This metric considered the following three factors:
(i) It is updated whenever a node is encountered, so that the nodes that are often encountered have a high delivery predictability, given by;

$$P_A(B)_{new} = P_A(B)_{old} + (1 - P_A(B)_{old}) * P_{init} \tag{1}$$

where $P_{init} \in [0,1]$ is the scaling factor set at a rate at which the predictability increases upon the encounter. (ii) If a pair of nodes do not find each other after a certain time period, they are less likely to be good forwarders of the messages to each other, thus the delivery predictability values must age. This value is reduced according to the following equation:

$$P_A(B)_{new} = P_A(B)_{old} \times \gamma^K \tag{2}$$

where $\gamma \in [0,1]$ is the aging constant, K is the number of time units since the last decay.
(iii) The delivery predictability also has a transitive property, that is, based on the observation that if node A frequently finds node B, and node B frequently encounters node C, then node C is probably a good node to

forward the messages directed to node A, i.e.

$$P_A(C)_{new} = P_A(C)_{old} + (1 - P_A(C)_{old}) \times P_A(B) \times P_B(C) \times \beta \tag{3}$$

where $\beta \in [0,1]$ is a scaling constant that controls how large the impact the transitivity should have over the delivery predictability. When two nodes meet, a message is forwarded to the node whose the delivery predictability of the message to the destination is higher. The first node does not delete the message after transmitting it as long as there is sufficient buffer space available with it. The reason for this being that it might encounter a better node, or even the final destination of the message in the future. ProPHet does not consider any other parameter such as encounter or distance of a node when calculating the delivery probability.

2.2 EDR

The EDR protocol for OppNet [3] has been designed based on the context information of nodes. The scheme initially calculates the Encounter values of each pair of nodes in the network dynamically. Similarly the Euclidean distance of each pair of nodes are calculated. For selecting the next hop node, EDR uses two forwarding parameters namely Encounter forwarding parameter α and Distance forwarding parameter β. Where,
α = Node Encounter/ SumEncounter and β = Node Distance/ SumDistance.

The EDR protocol then tries to maximize the number of encounters with the destination, and minimize the distance from the destination of a neighbouring node for every message. The protocol normalizes the mean of the α and β values and then calculates the best forwarding parameter γ as the ratio of α and β. A threshold value T is calculated as the average of the γ values of neighbouring nodes. The message is then forwarded to all the neighbouring nodes having a γ value greater than or equal to this generated threshold value T. Since EDR is a context based selected copy scheme, hence the protocol does not perform well in terms of message delivery probability.

2.3 HBPR

The History Based Prediction for Routing (HBPR) protocol [8] uses the history of nodes movement to predict their further locations. The forwarding of the messages is depended on the time taken by a node to meet another node and the direction of the nodes movement taken from the history of nodes. The nodes move in the network ares which is further divided into cells with a unique number to identify a particular cell. The cell that a node visits frequently is called its home location. The HBPR protocol design mainly consists of: Home location identification, where nodes are assumed to follow the Human Mobility Model [14]. They visit some of the locations very frequently and some of them rarely. During the course of operation, if a node changes its pattern and a different location has more frequency, it floods the network with this new information and a time stamp to distinguish the new home location information from its old information. All the nodes initially flood their home locations to inform all other nodes in the network. This paragraph provides the information of a node's movement which is used to predict the further locations. Message generation and home location update consists of two parts. First, new messages are generated at some of the nodes, and the destination *ID* is recorded from a newly generated message. Second, during message transmission, if any node changes its location, this information is flooded immediately to the network so that every node can update the network home locations. This helps for adjusting the changing relation and behavior of nodes. For next hop selection the parameter used by the HBPR protocols are a) stability of node's movement b) prediction of the direction of future movement of node c) perpendicular distance of the neighbouring node from the line of sight of source and destination nodes. A utility metric is calculated using these parameters in order to identify the next hop. Any node that has the metric utility value greater than a prescribed threshold (T) is given the message copy. It should be noted that the HBPR has been designed to perform better in human mobility scenarios compared to other mobility models [15]. The HBPR protocol uses two tables (history table and home location table) in the course of Utility Metric calculation. The performance of HBPR is best for Human Walk mobility model.

3 Proposed Scheme: P-EDR

3.1 Motivation

The research literature on OppNets has brought forward a number of routing protocols that have been proposed based on message dissemination in the network and the context-based information. It is also observed from the literature that most of the dissemination based protocols like [4,5,7] do not perform well in terms of network congestion due to message overheads, utilization of network resources due to the use of dissemination of message for next hop selection. Similarly, most of the context-based routing techniques such as [3,7,8,9,10] also reflect the drawback in the performance metrics like message delay and message delivery probability. This concern has lead the researchers to consider the node's context information for selecting the next hop for efficient message transfer towards its destination. Based on this a protocol named EBR in [11] has been proposed, which considers the context information of nodes like number of encounters of a node (i.e. how many times a node encounter with its intermediate node) to select the next hop. Even then, there are some more context information like distance of nodes with respect to corresponding destination (i.e. distance values of a node with corresponding destination) that have not been considered for next hop selection. With this context information (i.e. number of encounter and distance values of node with respect to corresponding destination), a routing technique EDR has been designed. Further, it has been observed that the performance of EDR in terms of message delivery probability, message overhead ratio etc. is very poor. In order to enhance the performance of EDR, in this work a new routing protocol Probability based EDR (P-EDR) has been presented such that it forwards the message to its destination with maximum delivery probability, minimum overhead ratio and minimum number of message drops. In this work it has been assumed that the nodes are cooperative and helpful and have sufficient energy level without presence of any malicious nodes. In P-EDR, the next best forwarder of the message is obtained based on the values of four parameters, i.e. α, β, γ and P_λ. The scheme initially calculates the Encounter values of each pair of nodes in the network dynamically, where encounter value is the number of times they have encountered each other. Likewise, the Euclidean distance between every two set of nodes in the network is dynamically calculated. Then the *TotalEncounter* value for each node in the network is calculated as the total number of encounters with all the other nodes in the network. Similarly, *TotalDistance* value for a node is calculated as the sum of the Euclidean distances of all the nodes in the network with that node.

The source node that wishes to communicate a message towards the destination initially, identifies those neighbour nodes which are within its range. Let us consider there are K number of neighbouring nodes available within the range of the source node. To select the next hop, the so-called forwarding parameter α and β are calculated for each of the K nodes in the network, represented as;

α = *node encounter/ TotalEncounter*.

and

β = *node distance/ TotalDistance*.

Now, to identify the next best forwarding node from the set of K neighbouring nodes, the proposed P-EDR tries to maximize the number of encounters with the destination, and minimize the distance from the destination for each of the neighbouring nodes. To obtain this the P-EDR protocol calculates the forwarding parameter γ, which is the ratio of α and β for each neighbouring node. Hence

$\gamma = \alpha / \beta$.

A threshold value T is calculated as the average of the γ values of the K neighbouring nodes. Next, a set referred as Hashmap is obtained that contains all the neighbouring nodes say N whose γ value is greater than or equal to this generated threshold. In order to obtained the best hop from the Hashmap, the scheme computes the delivery probability of the message for each node in the Hashmap using equation 4.

$$P_\lambda(N) = \gamma \times P_i(j)(S) \times \tau \tag{4}$$

Where $P_\lambda \in [0,1]$ represent the delivery probability of a node in Hashmap and $P_i(j)$ represent the delivery probability of node i from the network and j from the Hashmap as explained in equation 1,2 and 3. Here, the value $\tau \in [0,1]$ is used as a scaling factor.

Finally the scheme P-EDR select the node as a final forwarder of the message from the source to the destination whose delivery probability or parameter P_λ is maximum among the nodes store in the Hashmap.

The algorithm of P-EDR protocol is shown in Algorithm 1.

Algorithm 1

```
//Encounter(source,dest.) returns the number of encounters of source w.r.t. destination.
//TotalEncounter() returns the sum of the encounters of all the neighbouring nodes with destination.
//Distance(source,dest.) returns the distance of source w.r.t. destination.
//ToatlDistance() returns the sum of the distance of all the neighbouring nodes with destination.
//Probability() returns the delivery probability of a node towards the destination.
//$P_\lambda(N)$ is the delivery probability of nodes in Hashmap
//$P_i(j)(S)$ is the the delivery probability of node $i$ from the network and node $j$ from the Hashmap
//P(S) is the delivery probability of the source node S
   Begin
   $K$ = Set of neighbouring nodes of the source
   for each message do
      for each $K$ do
         $\alpha$ = Encounter(n,dest.) / TotalEncounter()
         $\beta$ = Distance(n,dest.) / TotalDistance()
         $\gamma = \alpha / \beta$
      end for
   end for
   for each $K$ do
      if $\gamma \geq$ T then
         Insert node in Hashmap
      end if
   end for
   N = Set of nodes in Hashmap
   S = Source node
   for each $N$ do
      $P_\lambda(N) = \gamma \times P_i(j)(S) \times \tau$
      if $P_\lambda(N) > P(S)$ then
         Transfer message from S to N
      end if
   end for
   End
```

4 Simulation and Results

In this section, the performance of the proposed P-EDR scheme is compared against the EDR, HBPR and ProPHet protocols using the ONE simulator [12]. To evaluate and analyze the performance of the proposed protocol, six groups of mobile nodes have been considered in this simulation, where three groups are pedestrian and three groups comprise of trams and buses. Each pedestrian group has 30 nodes with a walking speed of 0.5-1.5 m/s. The two groups of tram have 2 nodes of each with a speed of 7-10 m/s. A 50 MB buffer size is assigned for each node in the group of trams and a buffer size of 15 MB is assigned for each node in the group of pedestrians. Communication between the nodes is performed using a bluetooth interface that has a transmission range of 20 meters with a transmission speed of 250 Kbps. The high-speed interface transmission speed is 10 Mbps, with a range of 1500 meters. For each message generated a 100 minutes Time-to-Live (TTL) is assigned. A new message of the size of 500KB-1MB at a time interval of 25-35 seconds is generated. A 100000 seconds time is assigned for every simulation. The size of simulation area or world size is considered to be 4500 x 3400 Sq.meters and the shortest path map based movement model [13] is used for nodes movement.

The results of P-EDR are observed and compared with the standard protocols like EDR, HBPR and ProPHet by varying the number of nodes from 66 to 186, the TTL from 100 minutes to 300 minutes, and the message generation interval from 25-35 seconds to 65-75 seconds. The results are shown in Fig.1 to Fig.9.

Fig.1 to Fig.3 depict the effect of the number of nodes, TTL, and message generation interval on the delivery probability of a message. Fig.1 shows that the message delivery probability of P-EDR outperformed among the three protocols such as EDR and HBPR. This is due to fact that P-EDR selects the next hop based on the delivery probability along with the number of encounters and distance of a node with respect to the destination. Whereas in the other protocols delivery probability is not considered and in the ProPHet only delivery probability is considered. In fig.1 it can be observed that the average value of delivery probability for P-EDR is 0.61822, whereas for EDR it is 0.53359, for HBPR it is 0.56052, and for ProPHet it is 0.60262. Thus, the performance of P-EDR is 15% better than EDR, 10% better than HBPR and 3% better than the ProPHet routing protocol. In Fig.2 with the increase in message TTL, the message delivery probability decreases for all the protocols. This is due to the availability of more number of messages in the network. The average value of message delivery probability for P-EDR is 0.53827, which is more compared EDR, HBPR nad ProPHet that

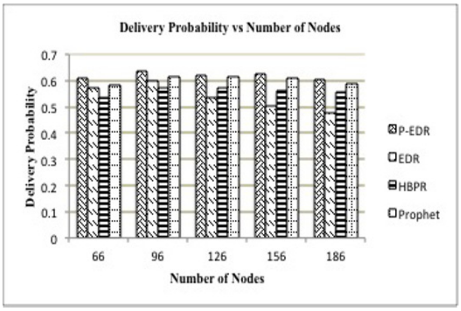


Fig. 1: Delivery Probability vs. No. of Nodes

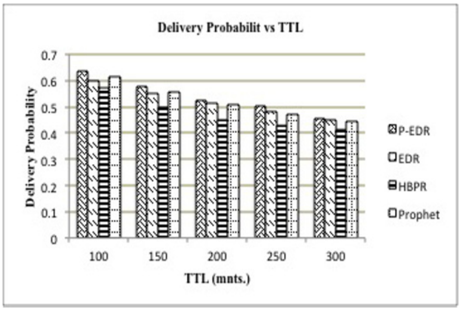


Fig. 2: Delivery Probability vs. TTL

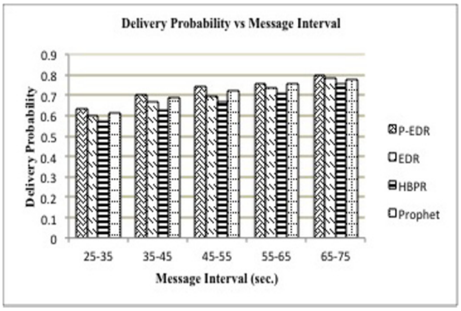


Fig. 3: Delivery Probability vs. Message Interval

have values of 0.51812, 0.4729 and 0.5179 respectively. Further the performance of P-EDR is 3.8% better than EDR, 13.8% better than HBPR and 3.9% better than ProPHet. Similarly, Fig.3 is plotted between message generation interval and message delivery probability for the four protocols. It can be seen that the message delivery probability increases for all the protocol as the message generation interval increases. This is because of the lesser number of messages is being generated in the network and hence their is a decrease in the message dropping rate. It has been observed that the average message delivery probability of P-EDR is 0.72618, which is greater than EDR that has the value of 0.69494, HBPR which has the value of 0.66708 and ProPHet which has value 0.71096 leading to an improved performance of P-EDR. by 4.4%, 8.8% and 2% over the EDR, HBPR and ProPHet routing respectively.

Fig.4 to Fig.6 shows the effect of the number of nodes, TTL, and message generation interval on number of message dropped. Fig.4 shows the effects of the number of nodes versus the messages dropped for the four protocols. It is clear from the graph that the average messages dropped in case of all the four protocols

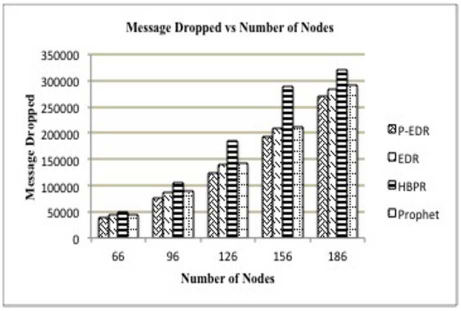


Fig. 4: Message Dropped vs No. of Nodes

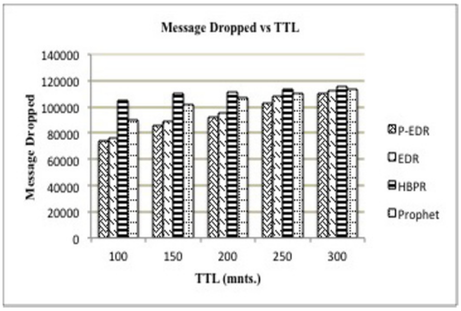


Fig. 5: Message Dropped vs TTL

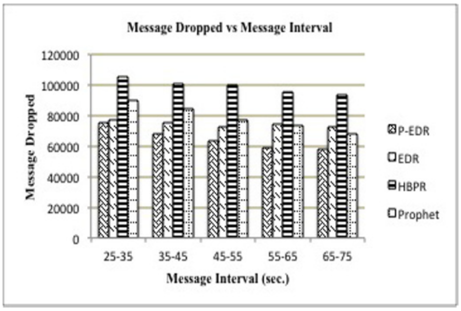


Fig. 6: Message Dropped vs. Message Interval

increases as increased in the number of nodes. The average messages dropped in P-EDR is 140575, which is quite less than EDR, HBPR and ProPHet that have the values 152661, 189743 and 156082 respectively. Further, the P-EDR's performance is 7.9% better than EDR, 25.9% better than HBPR and 9.9% better than ProPHet in terms of average number of messages dropped when the number of nodes is varied. Fig.5 is plotted between TTL and message dropped for the four protocols. It can be seen that the average messages dropped for P-P-EDR is 93025 whereas for EDR, HBPR and ProPHet it is 96241, 111440 and 104470 respectively. Further, the P-EDR's performance is 13%, 34% and 17% better than EDR, HBPR and ProPHet respectively in terms of the number of messages dropped with varying TTL. Fig.6 is drawn between message generation interval and number of messages dropped. It has been observed from this figure that the average messages dropped in P-EDR is 64778, which is less than EDR, HBPR and ProPHet that have the values 74540, 99091 and 78584 respectively. Further, P-EDR's performance is 3% better than EDR, 16% better than HBPR and 10% better than ProPHet.

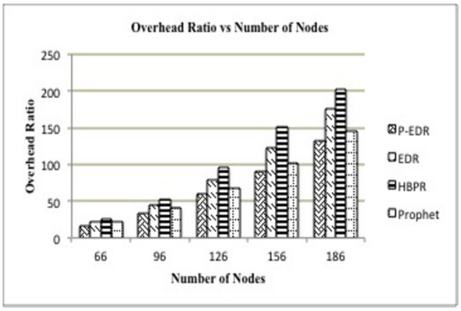


Fig. 7: Overhead Ratio vs. No. of nodes

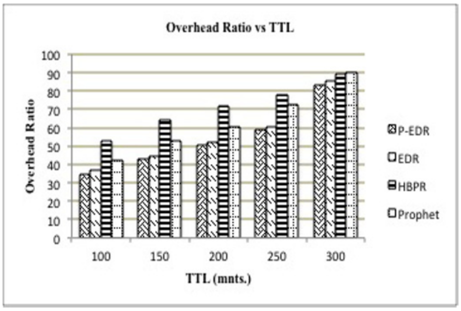


Fig. 8: Overhead Ratio vs. TTL

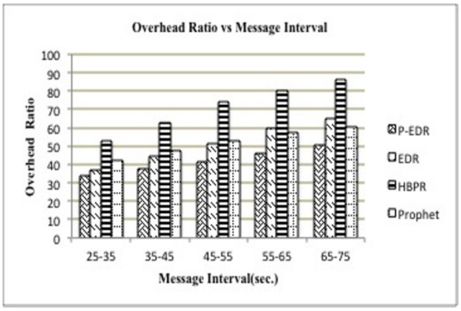


Fig. 9: Overhead Ratio vs. Message Interval

Fig.7 to Fig.9 show the effect of the number of nodes, TTL, and message generation interval on message overhead ratio. Fig.7 is plotted between the number of nodes and message overhead ratio. It is found that the average messages overhead ratio in P-EDR is 66.6635, whereas for EDR it is 88.8751, for HBPR it is 105.5214 and for ProPHet it is 75.5691, which is quite higher than P-EDR. This is due to the increase number of messages as number of node increases in the network. Hence, P-EDR's performance is 25% better than EDR, 36% better than HBPR and 11% better than ProPHet in terms of message overhead ratio. Similarly, Fig.8 is drawn between TTL and messages overhead ratio. It has been observed that the average message overhead ratio of P-EDR is 50.0275 which is quite low as compared to EDR which has the value of 55.8715, HBPR has the value of 71.9420 and the overhead ratio is 63.8078 for ProPHet. Since P-EDR select the next best hop using the context information and delivery probability of nodes for the message, which results a minimum message overhead is generated during message delivery to the destination. Further the performance of P-EDR is 3% better than EDR, 24% better than HBPR and 15% better than ProPHet. Finally, Fig.9 plots the message generation interval versus the messages overhead ratio for the four protocols. From the figure it is inferred that the average messages overhead ratio of P-EDR is 41.9631, for EDR it is 51.58876, for HBPR it is 71.4459 and for ProPHet it is 52.0530. Here P-EDR's performance is 18% better than EDR, 41% better than HBPR and 19% better than ProPHet in terms of messages overhead ratio.

5 Conclusion

This paper proposes a routing protocol called P-EDR, which is based on the delivery probability of the neighbouring node in addition to the number its encounter and its distance with the destination node. The work focusses in the direction of enhancing the performance of EDR by applying probability (delivery predictability of a node) to identify the nodes among the neighbouring nodes that are selected as good forwarders from the EDR protocol. Simulation results have shown that the proposed P-EDR outperforms EDR, HBPR and ProPHet that have been chosen as standard protocol, in terms of average message delivery probability, average number of messages dropped and average message overhead ratio. In future, the energy and security issues with this proposed scheme can be addressed. Further the proposed method can also be explored with different mobility models and scenarios.

References

1. L. Pelusi, A. Passarella, and M. Conti, Opportunistic networking: data forwarding in disconnected mobile ad hoc networks, IEEE Communications Magazine, vol. 44, Issue 11, November 2006, pp. 134-141.
2. K. Fall, A Delay-Tolerant Network Architecture for Challenged Internets, in proceedings of ACM SIGCOMM 2003, Karlsruhe, Germany, 25-29 August, 2003, pp. 27-34.
3. S. K. Dhurandher, S. Borah, I. Woungang, D. K. Sharma, K. Arora, and D. Agar-wal, EDR: An Encounter and Distance Based Routing Protocol for Opportunistic Networks, (Accepted Jan 25, 2016), the 30th IEEE International Conference on Advanced Information Networking and Applications (AINA-2016), Crans-Montana, Switzerland, March 23-25, 2016. To appear.
4. A. Vahdat, and D. Becker, Epidemic routing for partially connected ad hoc networks, Technical Report CS-2000-06, Dept. of Computer Science, Duke University, Durham, NC, 2000.
5. T. Spyropoulos, K. Psounis and C. S. Raghavendra, Spray and Wait: An Efficient Routing Scheme for Intermittently Connected Mobile Networks, in proceedings of SIGCOMM Workshop on Delay-Tolerant Networking, Philadelphia, USA, 22-26 Aug.2005, pp. 252-259.
6. A. Lindgren, A. Doria, and O. Schelen, Probabilistic routing in intermittently connected networks, ACM IGMOBILE, Mobile Computing and Communications, 2003 Review, vol. 7, Issue 3, pp. 1920.
7. Sanjay K. Dhurandher, Satya Jyoti Borah, Mohammad S. Obaidat, Fellow of IEEE, Deepak Kr. Sharma, Sahil Gupta and Bikash Baruah Probability-based Controlled Flooding in Opportunistic Networks WINSYS 2015 International Conference on Wireless Information.
8. S. K. Dhurandher, Deepak Kr. Sharma, I. Woungang, and Shruti Bhati, "HBPR: History Based Prediction for Routing in Infrastructure-less Opportunistic Networks" IEEE 27th International Conference on Advanced Information Networking and Applications(AINA 2013), Barcelona, Spain, pp. 931-936
9. C. Boldrini, M. Conti, I. Iacopini and A. Passarella, HiBOp: A History Based Routing Protocol for Opportunistic Networks, in proceedings Of IEEE International Symposium on World of Wireless, Mobile and Multimedia Networks, 2007 (WoWMoM 2007), Espoo, Finland,18-21 June 2007, pp. 1-12.
10. S. K. Dhurandher , D. K. Sharma, I. Woungang, R. Gupta, S. Gupta,"GAER: Genetic Algorithm based Energy-efficient Routing Protocol for Infrastructure-less Opportunistic Networks", Journal of Supercomputing, Springer, vol. 69, Issue 3, Sept. 2014, pp 1183-1214.
11. Nelson, S.C.; Bakhat, M.; Kravets, R., "Encounter-Based Routing in DTNs," INFOCOM 2009, IEEE, Vol., No., pp.846,854, 19-25 April 2009.
12. A. Keranen. Opportunistic Network Environment Simulator, Special Assignment Report, Helsinki University of Technology, Dept. of Communications and Networking, May 2008.
13. A. Keranen, J.Andott, 2007, Opportunistic increasing reality for DTN protocol simulations, Special Technical Report, Helsinki University of Technology, Networking Laboratory.
14. C. Song, Z. Qu, N. Blumm, and A. Barabasi, Limits of Predictability in Human Mobility", Science, Vol. 327, February 2010, pp. 1018-1021. and 20.
15. S. K. Dhurandher, D. K. Sharma, and I. Woungang, "Mobility Models-Based Performance Evaluation of the History Based Prediction for Routing Protocol for Infrastructure-less Opportunistic Networks", Proc. of 10th Intl. Conference, MOBIQUITOUS 2013, Tokyo, Japan, Dec. 2-4, 2013, pp. 757-767.

Comparison of Biometric and Linguistic Secret Sharing Protocols

Lidia Ogiela, Marek R. Ogiela, Urszula Ogiela
AGH University of Science and Technology
Cryptography and Cognitive Informatics Research Group
30 Mickiewicza Ave., 30-059 Krakow, Poland
logiela@agh.edu.pl, mogiela@agh.edu.pl, ogiela@agh.edu.pl

Abstract. In this paper will be presented comparison and security features of biometric and linguistic threshold schemes. Additionally efficiency evaluation for such protocols will be done. Possible application of presented algorithms will be described with future directions in the area of strategic information management, and security for cloud applications.

1 Introduction

For division of strategic data cryptographic threshold protocols were proposed. The first sharing methods were proposed in late seventies, but till now it have been proposed many complex, efficient, and secure algorithm. All such techniques define two different classes i.e. secret sharing techniques and secret splitting. Data sharing algorithms were presented manly in [1], [2], [3], and the main idea of such methods is to secure information by split them between particular groups of participants. All secret splitting methods allow to generate a particular number of secret parts (called shadows), than distribute them among participant of protocol. But to restore the original information it is necessary to compile all the secret parts. In secret sharing approaches shadow generation is very similar, but to restore the original information it is enough to compile a less number of secret parts. Secret sharing is more universal and allows to restore the previous information also in case of losing any secret parts.

For similar tasks we propose two new types of threshold procedures called biometric threshold schemes and linguistic threshold schemes. These algorithms allow involving some personal information into the encryption process [4], [5], [6], [7]. In following section will be presented these procedures with theirs features evaluation and comparison.

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_49

2 An Idea of Linguistic Threshold Schemes

The first proposed technique for information sharing is linguistic threshold procedure [1]. The main idea of such methods lays in using mathematical linguistic formalisms for representation of shared data and encoding procedure. In this technique it is necessary to define special type of formal grammars which enable encoding bit sequences with different length. It only depends on the defined formal grammar as well as some features, which may be additionally encoded in one of generated secret parts. The way of information encoding using linguistic procedures is more general encoding scheme use in DNA cryptography [8], [9]. However in classic DNA cryptography can use only four nitrogen bases, to encode particular bits of information or two bits block in particular nitrogen bonds.

In linguistic threshold schemes it is possible to create more general encoding structure, which allows encoding in one step, more than two bits of information e.g. 5, 6 or more.

3 Information Division Using Biometric Threshold Schemes

Second approach is connected with using some personal features in sharing protocol. Such technique is called biometric threshold schemes and was proposed by authors in [9]. In biometric threshold schemes each shadow is generated using biometric features. In biometric threshold schemes is possible to use the single biometric feature or several different patterns [10], [11]. The most popular biometric patterns appropriate for this purpose are:

- fingerprint patterns,
- handwriting features,
- retina patterns,
- facial features,
- hand vein layouts,
- voice parameters.

Sometimes we can also consider different non-standard personal features obtained from different sources like medical records, personal habits or behavioral feature [12], [13].

The biometric data encryption is realized in two separated steps. The first one, is after splitting the information, and contains indexing procedure for each shadow by biometric features. The second one, is realized while combining the strategic information.

Such techniques allow to perform secure data sharing processes, because each participant gives only shadow marked by his or her personal features. It isn’t possible to give shadow to non-trusted participants.

4 Comparison of Linguistic and Biometric Sharing Protocols

Both described classes of proposed threshold procedures i.e. linguistic threshold procedures and biometric threshold protocols are not only very interesting from scientific point of view, but also extend features of classic threshold procedure. Biometric and linguistic threshold procedures are extensions for classic threshold procedure, and remain all security features, which characterize classic protocols. Both of them have also some important additional features, which are not present in classic threshold algorithms. Among such additional features in linguistic threshold procedures we can find:

1. Application of formal grammars and languages to split strategic information.
2. Possibility to encode block of information with different bit length.
3. Polynomial complexity which depends on applied formal grammar.
4. Possible application for strategic data sharing in different management structures like layered as well as hierarchical structure [14], [15].
5. Possibility to generate personalized shadows, which determine the way of information encoding.
6. Application in secure information management tasks for different structures.
7. Possibility to generate different number of secret parts considering personal accessing grant to original information.

Most important additional features in biometric threshold procedures are following:

1 Possibilities of creating personalized shadows. Such shadows allow not only restoring original information but also determining the owner of secret part [16].
2 Applicability with cognitive information systems at the stage of personal feature extraction [17], [18].
3 Standard and non-standard biometrics may be use in shadow generation.
4 Unlimited number of shadows can be generated.

Mentioned features, make these systems very universal with many possibilities of different application.

5 Conclusions

Described in this paper sharing protocols have many important features, which make them applicable in personalized cryptography or secure information management tasks. These protocols seem to be very efficient and secure because security features are guaranteed by basic threshold procedure, which may be use in the whole sharing protocol. Additionally both of these procedures have some special properties, which extend its functionality.

In biometric threshold procedures it is possible to use some personal characteristics, which finally allow creating personalizes parts of divided information. Such feature allows determining who is the owner of secret part, what also prevent the information leakage, when such protocol may be violated.

In linguistic threshold schemes it is possible to divide information in different manners considering the numbers of trusted persons and also theirs accessing grants to restore original information. Both of these protocols may be applied in general secret sharing application, but also in professional strategic data sharing and management, and trusted communication infrastructures [19]. They may also be applied in secured data distribution in the cloud environment, and information or services management in ubiquitous computing or ambient world.

Acknowledgments. This work has been supported by the National Science Centre, Republic of Poland, under project number DEC-2013/09/B/HS4/00501.

References

1. Ogiela, M.R., Ogiela, U.: Linguistic Approach to Cryptographic Data Sharing, FGCN 2008 – The 2nd International Conference on Future Generation Communication and Networking, December 13-15, 2008, Hainan Island, China, 1 (2008) 377–380
2. Ogiela, M.R., Ogiela, U.: Grammar Encoding in DNA-Like Secret Sharing Infrastructure. 2nd International Conference on Advanced Science and Technology (AST), Miyazaki, Japan, Jun 23-25, 2010. LNCS 6059, (2010) 175-182
3. Ogiela, M.R., Ogiela, U.: Shadow Generation Protocol in Linguistic Threshold Schemes. in: D. Slezak, T.H. Kim., W.C. Tang et all., Security Technology, Communications in Computer and Information Science 58 (2009) 35-42
4. Ogiela, L., Ogiela, M.R.: Cognitive systems for intelligent business information management in cognitive economy. International Journal of Information Management, 34 (2014) 751-760
5. Shi, J., Lam, K-Y.: VitaCode: Electrocardiogram Representation for Biometric Cryptography in Body Area Networks. 1st International Conference on Ubiquitous and Future Networks, Hong Kong, China, Jun 07-09 (2009) 112-115
6. Kumar, A., Kumar, A.: Adaptive management of multimodal biometrics fusion using ant colony optimization. Information Fusion 32 (2016) 49–63
7. Shi, J., Lam, K-Y.: VitaCode: Electrocardiogram Representation for Biometric Cryptography in Body Area Networks. 1st International Conference on Ubiquitous and Future Networks, Hong Kong, China, Jun 07-09 (2009) 112-115
8. Ogiela, L.: Semantic analysis and biological modelling in selected classes of cognitive information systems. Mathematical and Computer Modelling, 58 (2013) 1405-1414
9. Ogiela, L., Ogiela, M.R.: Cognitive systems and bio-inspired computing in homeland security. Journal of Network and Computer Applications, 38 (2014) 34-42
10. Bajwa, G., Dantu, R.: Neurokey: Towards a new paradigm of cancelable biometrics-based key generation using electroencephalograms. Computers & Security 62 (2016) 95-113
11. Hani, M.K., Marsono, M.N., Bakhteri, R.: Biometric encryption based on a fuzzy vault scheme with a fast chaff generation algorithm. Future Generation Comp. Syst. (2013) 800–810
12. Hachaj, T., Ogiela, M.R.: CAD system for automatic analysis of CT perfusion maps. Opto-Electronic Review, 19 (2011) 95-103
13. Nandakumar, A.K.J.K., Pankanti, S.: Fingerprint-based fuzzy vault: Implementation and performance. In: IEEE Transactions on Information Forensics and Security. 2 (2007) 744–757
14. Ogiela, L.: Towards cognitive economy. Soft Computing 18 (2014) 1675-1683
15. Ogiela, L., Ogiela, M.R.: Management Information Systems. in: J.J. Park, Y. Pan, H.C. Chao, et all., 2nd FTRA International Conference on Ubiquitous Computing Application

and Wireless Sensor Network (UCAWSN), South Korea, 07-10 July 2014, Ubiquitous Computing Application and Wireless Sensor, Lecture Notes in Electrical Engineering 331 (2015) 449-456

16. Ogiela, L.: Cognitive informatics in image semantics description, identification and automatic pattern understanding. Neurocomputing 122 (2013) 58-69
17. Ogiela, L.: Cognitive Computational Intelligence in Medical Pattern Semantic Understanding. in: M. Guo, L. Zhao, L. Wang (Eds.), Fourth International Conference on Natural Computation, ICNC 2008, Jinan, Shandong, China, 18-20 October, 2008, 245-247
18. Ogiela, L.: Computational Intelligence in Cognitive Healthcare Information Systems. in: I. Bichindaritz, S. Vaidya, A. Jain et all., Computational Intelligence in Healthcare 4: Advanced Methodologies, Studies in Computational Intelligence 309 (2010) 347-369
19. Ogiela, L.: Data management in cognitive financial systems. International Journal of Information Management 33 (2013) 263-270

Concealing Additional Secrets Using Sharing Approach in Steganography

Marek R. Ogiela, Katarzyna Koptyra
AGH University of Science and Technology
Faculty of Electrical Engineering, Automatics, Computer Science
and Biomedical Engineering
30 Mickiewicza Ave., 30-059 Krakow, Poland
mogiela@agh.edu.pl, kkoptyra@agh.edu.pl

Abstract. This paper describes a method of concealing additional secret data in fuzzy vault cryptosystem. The hidden information is placed on the second level of the system, what means that it is impossible to reveal higher level secret before decoding all related data from lower level. This property gives an opportunity of using presented technique as a secret sharing system in which information from previous step is used in the next part of the algorithm (no additional keys are required). As the existence of second-level secrets is not obvious for external observer, this idea may be applied for steganography purposes. The format of concealed secret is two-dimensional point (x, y), thus the method presented in this paper is suitable for protecting all data that can be presented as a pair of numbers.

1 Introduction

Secret sharing techniques allow to divide some data into pieces that can be joined later when participants agree to cooperate with each other. In many cases it ought to be done confidentially to prevent unauthorized people from discovering the existence of the secret information, as it could create diverse threats to users being in conspiracy. Sometimes even the mere fact of storing keys or shares may be dangerous and contributes to compromise oneself. Therefore for really important data it is worth to look for solutions based on steganography, which do not require storing any key related to shared secret.

In modern world to conceal the message from malicious third parties one can use various information systems. But the problem is that such non-standard functionalities are very rarely implemented out-of-box. On the other way, creating a new system from scratch is not always an option. So we need to either find a way of using some parts of existing system for our own purposes or modify the system to support our new, hidden function. It is difficult especially in multi-user systems, in which every modification should be transparent and not affecting to participants.

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_50

This paper is a continuation of previous work [1] that describes a method of concealing many independent secrets in a fuzzy vault scheme (which can serve as multi-user system). Current idea is an extension of that concept giving some participants the possibility of sharing additional secrets with assumption that other functions remain intact for rest of users. Therefore it refers to multi-level steganography [2] as second secret is hidden in such a way that lower-level information is used to reconstruct higher-level data.

2 Multi-secret Fuzzy Vault

This section describes multi-secret fuzzy vault, which is a basis for the presented idea. At the beginning the underlying conception of fuzzy vault scheme is discussed.

Fuzzy vault [3] is a cryptosystem that relies on polynomial reconstruction. It uses a key in form of unordered set for locking and retrieving a secret. The entire vault is consisted of a great number of points, some of which are significant and remaining are chaff. The creation process begins with choosing a polynomial that encodes the secret (e.g. as a free term). Then this formula is evaluated on all elements of the key. As a result, genuine points are obtained. To hide the secret, we need also a number of false points which are placed more or less randomly, but with two constraints. Firstly, they cannot lie on polynomial and secondly, their x coordinates may not be members of the key. After producing both groups of points, the vault is ready and the polynomial can be erased. The recovering stage requires a key that should be identical or very similar to the key used in encoding process. With this set, the user can properly identify genuine points and reconstruct the polynomial. If the discrepancies between these keys are too big, some chaff points will be selected, what cases obtaining a wrong formula and an incorrect secret. Two important properties of fuzzy vault scheme are error tolerance (provided by error correction) and order invariance (provided by the form of the key). As a consequence, this cryptosystem is used willingly in biometrics systems with particular focus on fingerprint-based, like [4].

It turns out that fuzzy vault scheme is suitable for locking many secrets at the same time, as described in [1]. To do this, we need more keys, which are still in form of unordered sets, but all of them have to be disjunctive. Each secret information is encoded in individual polynomial, which is then used together with related key to obtain genuine points. After that chaff points are generated. This time we have two requirements: x coordinate of every false point cannot be a member of any key and also every that point may not be placed on any polynomial. There are many examples of such algorithms that can be easily modified to meet these conditions, for instance [5][6]. Thereafter, the multi-secret fuzzy vault is formed of chaff points and all groups of genuine points. Each secret may be reconstructed with use of the corresponding key or with not exactly the same, but very similar one. The recovering process is identical as in original scheme. It should be noted that important properties of fuzzy vault mentioned earlier – error tolerance and order invariance – are also preserved in multi-secret version of this cryptosystem. What is more, above solution can serve either for single user with many secrets or as multi-user system in which each participant has own secret and key.

3 Concealing Additional Shared Secrets

The construction of multi-secret fuzzy vault gives an opportunity of hiding additional, shared secrets. The idea is based on the fact that one vault can have inside many secrets encoded in various polynomials. The points in which the polynomials intersect are places where additional information can possibly be concealed. So the format of shared data is a pair (x, y). The number of secrets we can embed is dependent on degree of polynomials in the vault. To be more precise, for degree *n* we can hide up to *n* secrets (the assumption is that *n* is equal for all formulas). Because second level secrets are embedded in intersection points, the polynomials used in this process should not be selected totally randomly, but with specific algorithm. Such method not only has to generate formulas encoding lower level secrets, but also requires that selected points belong to both polynomials. Algorithm 1 depicts it in more detailed way.

Algorithm 1. Polynomial generation.

```
Input: $n$ – degree of polynomials (a number), $(x_{s1}, y_{s1})$, $(x_{s2}, y_{s2})$, …, $(x_{sn}, y_{sn})$ – second
level secrets (2D points), $S$ – first level secret (a number)
Output: $(a_n, a_{n-1}, \ldots, a_1, a_0)$ – coefficients of polynomial
    1. Create general formula with unknown coefficients
       $w(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + S$
    2. Make a system of equations ($n$ equations of degree $n$)
       $a_n x_{s1}^n + a_{n-1} x_{s1}^{n-1} + \ldots + a_1 x_{s1} + S = y_{s1}$
       $a_n x_{s2}^n + a_{n-1} x_{s2}^{n-1} + \ldots + a_1 x_{s2} + S = y_{s2}$
       …
       $a_n x_{sn}^n + a_{n-1} x_{sn}^{n-1} + \ldots + a_1 x_{sn} + S = y_{sn}$
    3. Solve the system from point 2.
    4. return $(a_n, a_{n-1}, \ldots, a_1, S)$
```

The explanation of Algorithm 1 is as follows.

First we create a general formula which encode first level secret. Then we have to fit its remaining coefficients to conceal shared information from second level. To do this we create the system of *n* equations to find missing *n* unknowns. After solving this system we obtain all coefficients needed to form a polynomial of degree *n* that embeds both secrets.

Next stages of hiding phase (genuine and chaff points generation) are the same as described in [1]. Later in this paper is presented an example showing how to conceal and reveal secrets for Alice and Bob.

Below is explained how to restore a secret shared between two users. It requires reconstruction of both polynomials first. During this operation two first level secrets are also decoded as they are stored in one of the coefficients. The whole process is shown in Algorithm 2.

Algorithm 2. Recovering 2nd level secret.

```
Input: $w_A$, $w_B$ – polynomials
Output: $(x_{s1}, y_{s1})$, $(x_{s2}, y_{s2})$, …, $(x_{sn}, y_{sn})$ – shared secrets (2D points)
```

1. Create an equation
   $w_A(x) = w_B(x)$
   $a_nx^n + a_{n-1}x^{n-1} + \ldots + a_1x + a_0 = b_nx^n + b_{n-1}x^{n-1} + \ldots + b_1x + b_0$
2. Solve the equation from point 1.
3. return $(x_{s1}, y_{s1}), (x_{s2}, y_{s2}), \ldots, (x_{sn}, y_{sn})$

It should be underlined that reconstruction of shared secret requires cooperation between participants. With only one polynomial it is impossible to compute points of intersection. The users have to decode their first level secrets first, as they take part in higher level information reconstruction. It means that revealing process does not need any additional key as it uses only data from lower level. This is an important aspect in security and will be discussed wider in Conclusions section.

4 Example

This section depicts an easy example which shows how presented idea can works in practice. Suppose that we use multi-secret fuzzy vault cryptosystem with all polynomials of degree 2. It means that it is possible to hide two additional shared secrets in form of 2D points. So we select them as follows: (1, 8) and (-1, 4). These will be shared between Alice and Bob on 2nd level. The participants have also their 1st level secrets, which are: 5 for Alice and 10 for Bob.

Hiding phase. This stage starts from choosing the polynomials for users. Because degree is equal to two, the formula is:

$$w(x) = ax^2 + bx + c \tag{1}$$

Now it is time to find *a*, *b* and *c* from formula (1) in such a way that the polynomial encodes 1st level secret and also can be considered as a share (used for recovering 2nd level secrets) at the same time. For Alice we can write:

$$\begin{gathered} a + b + 5 = 8 \\ a - b + 5 = 4 \end{gathered} \tag{2}$$

The solution is a = 1 and b = 2. Thus we receive Alice's polynomial

$$w_A(x) = x^2 + 2x + 5 \tag{3}$$

For Bob we have:

$$\begin{gathered} a + b + 10 = 8 \\ a - b + 10 = 4 \end{gathered} \tag{4}$$

The solution is a = -4 and b = 2. Therefore Bob's polynomial is

$$w_B(x) = -4x^2 + 2x + 10 \tag{5}$$

Next steps (like evaluating the formula on key elements) are identical as in [1] and are not presented here.

Recovering phase. The strategy for decoding 1st level secret is identical as described in [3] and for this reason is omitted in our example. Below we present how to reveal information from 2nd level. If Alice and Bob want to recover their shared secret, they should reconstruct their polynomials first. Then they have to cooperate and find all points of interception, as shown below in (6).

$$\begin{gathered} w_A(x) = w_B(x) \\ x^2 + 2x + 5 = -4x^2 + 2x + 10 \end{gathered} \tag{6}$$

The solution is x = 1 or x = -1. To find y values, we have to evaluate any of the polynomials (w_A or w_B) on computed x values. So y = 8 or y = 4. Finally our points are (1, 8) and (-1, 4).

In order to demonstrate the situation visually, we provide a graph containing all elements from above example (Fig. 1).

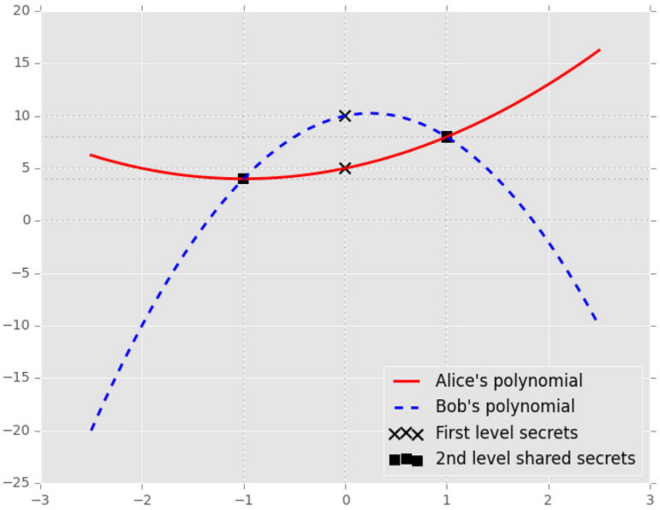


Fig. 1. Generated polynomials with marked 1st and 2nd level secrets.

5 Conclusions

This paper extends the concept of multi-secret fuzzy vault by adding the possibility of concealing additional shared secrets. To recover this data, the participants have to cooperate and reconstruct their polynomials to find points of intersections. In extended method the polynomial generation algorithm is different – the coefficients are not random, but are selected on the basis of both secrets. Due to this fact all information needed for recovering shared secrets are obtained from reconstructed polynomials. If we consider security, it is an important feature, because no additional keys are required and the users do not have to store suspicious data which may compromise them. In fact, only the key for 1st level secret is necessary, what can be considered as normal situation in multi-secret fuzzy vault cryptosystem.

It should be noted that in presented technique it is impossible to reveal shared information with only one formula. This fact has two consequences. If a participant is not able to reconstruct the polynomial, there is no way to recover the secret. However, it is also true in case of leakage. When an enemy intercepts one formula, he will not be able to reveal shared secret without the second polynomial (guessing identity of the other conspirator and stealing the key). Of course, in general, the existence of second level secret is unknown for unintended third parties and with honest users there is no trace of shared data. First level secrets may serve to deceive an adversary as they are not important for participants (they are visible during joining shares).

Finally, the shared secret is in form of 2D points. It gives an opportunity of hiding specific types of data which are presented as a pair of numbers, like day and month, hour and minutes or geographic coordinates (latitude and longitude). For every pair is known that these two values are related, what cannot be done in systems in which a set of one-dimensional numbers is stored. Therefore the secret may contain, for example, a place where something is hidden for both users and time when they can start.

To sum up, presented construction is able to conceal shared information for two participants in situations in which high level of conspiracy is required.

Acknowledgments. This work was supported by the AGH University of Science and Technology research Grant No 15.11.120.868.

References

1. Koptyra, K., Ogiela, M.R.: Fuzzy vault schemes in multi-secret digital steganography. In: 10th International Conference on Broadband and Wireless Computing, Communication and Applications, BWCCA 2015, Krakow, Poland, November 4-6, 2015. (2015) 183–186
2. Ogiela, M.R., Koptyra, K.: False and multi-secret steganography in digital images. Soft Comput. 19(11) (2015) 3331–3339
3. Juels, A., Sudan, M.: A fuzzy vault scheme. Des. Codes Cryptography 38(2) (2006) 237–257
4. Nandakumar, A.K.J.K., Pankanti, S.: Fingerprint-based fuzzy vault: Implementation and performance. In: IEEE Transactions on Information Forensics and Security. Volume 2. (December 2007) 744–757
5. Hani, M.K., Marsono, M.N., Bakhteri, R.: Biometric encryption based on a fuzzy vault scheme with a fast chaff generation algorithm. Future Generation Comp. Syst. (2013) 800–810
6. Nguyen, T.H., Wang, Y., Nguyen, T.N., Li, R.: A fingerprint fuzzy vault scheme using a fast chaff point generation algorithm. In: Signal Processing, Communication and Computing (ICSPCC), 2013 IEEE International Conference on, IEEE (2013) 1–6

Enhance Robotics ability in Hand Gesture Recognition by Using Leap Motion Controller

Alaa Ahmed Almarzuqi[1], Seyed Mohamed Buhari[2]

[1] King Abdulaziz University, Jeddah, Saudi Arabia
a.almarzoki@gmail.com
[2] King Abdulaziz University, Jeddah, Saudi Arabia
mibuhari@gmail.com

Abstract. Usage of intelligence in the interactions between human and robot is one of the significant topics in research. As an added assistance to human, robots are used to facilitate and assist human in many ways. Robots could be made to understand and recognize human gestures. Therefore, robots should be programmed to deal with certain gestures, that, the robots can identify and act accordingly. Our research aims to enhance the robot gesture recognition ability by using supported hand gesture detection device known as Leap Motion Controller (LMC). This research aims to expose the accuracy of hand gesture recognition using Leap Motion depth sensor to enhance intelligent system where human hand gestures are used to interact with robot in learning or gaming system.

1 Introduction

Gesturing is one of the natural forms of communication. Human always use gestures throughout the day and sometimes they use it instead of words as a common way of communication. Meanwhile, intelligent machines are manufactured and continuously developed to assist the human life in several areas. Robots, as intelligent machines, have significant economic and scientific importance in various walks of human life. These two aspects along with some significant research on human-machine-interaction motivated us two think how to enhance the relationship between human and machine using gestures.

Human–Robot Interaction (HRI), a robot research area, focuses on the interaction between human and robot and studying related aspects [1]. Development of HRI methodologies and systems is a successive and continuous process, since this interaction does not depend on any specific method.

The interaction between human and robot takes numerous forms, which can be divided into two categories: (i) direct interaction with a robot through speech recognition or body gestures (ii) indirect interaction through intermediate devices such as sensors, touch screens and sensing gloves.

As an added assistance to human, robots have started to facilitate and assist human in many ways, thus, the need for enhancing the interaction with robot is significant.

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_51

The gestures recognition involves using one camera or more or using sensor that captures the exact gestures. The presence of low resolution pictures hinders the effectiveness of image processing. Zhou et al. [2] proposed Finger-Earth Mover's Distance (FEMD) to increase the accuracy of recognizing the gestures taken using Kinect sensor. The purpose of this research is to develop a gesture recognition system that enables human robot interaction in an enhanced manner. The concept of this research is intended to enhance any robotic system where human use gestures to as a mean of interaction, while researcher will conduct their experiment on NAO (which is a humanoid robot manufactured for research and education purposes). It aims to improve the human-robot relation and assist in children education process, such that children will communicate with NAO via our system to get knowledge about the environment or learn new things. Our assistive system aims to make machine (Nao robot) accommodate with human environment and be able to recognize the human gestures then perform the intended action that is equivalent to the detected instruction with satisfactory performance and speed. Nao is a personal friendly robotic with humanoid appearance that takes great interest among researchers [3].

To test and evaluate our hand gesture recognition system, we prepare a real home environment to conduct the experiments of performing the programmed instructions that are given to the robot in form of hand gestures and measure the robot ability to understand these gestures and act appropriately.

We organize this proposal as follows: in the next section we identify the objectives of our research. After that, we will discuss some state-of-the-art gesture recognition techniques. Then, a summarized methodology section shows the proposed procedure involved achieving our objectives. Finally, the plan of our work is discussed.

2 Problem Statement

The evolution in the field of human interaction with the computer resulted in the use of robots, as in most aspects of life and educational entertainment for all age groups. The employment of the machine in some respects require greater accuracy in the results, and this requires the use of additional tools and software to overcome the weaknesses of the robot and improve its capabilities even further.

It is also known that the camera attached to the robot NAO head color camera is usually normal and that its proportion of accuracy in distinguishing human hand gestures is much lower than the validity of the results obtained using depth cameras or special hand gesture sensors, as studies have shown.

3 Objectives

The objectives of this research are:

1. To build a robust gesture–based recognition system to assist in children educational through enhanced human-robot interaction.
2. To develop multiple systems interaction interfaces that satisfies different user needs through providing more accessibility options.

3. To test the applicability of recognition system in a real HRI environment in different scenarios and applications

4 Related Work

The main goal for using the proposed system is to enrich intelligent learning system for children. Due to this purpose this section will be arranged as following, first, the importance of using hand gesture in children educational aspects will be reviewed. Then, some of intelligent learning system for children will be discussed. Finally, authors will describe Leap Motion device as sensitive sensor for recognizing hand gesture and show some researches that using it for learning purposes.

4.1 Hand gestures recognition and robotic

Firstly, in this section we are going to review some surveys that discuss hand recognition technologies and a number of works in this field. Then we will identify the importance of educational robotics and the significant role that gesture plays in the learning process. Finally, due to limited the number of gesture-based educational robotic works in the literature, we will represent similar works that focus on children education, entertainment or edutainment systems that use robotic.

Two surveys [4, 5] discuss hand gesture recognition techniques and review several gesture recognition systems. Authors have mentioned how interaction with machines through gesture recognition is an important field of study and they argue the gesture recognition adequately. There are a large number of applications that rely on the recognition of gesture, where they are used in many areas to serve different purposes; examples of these uses are sign language recognition, entertainment applications and children learning etc.

Gesture recognition involves using some tracking devices like gloves and various types of camera. Mitra and Acharya survey [1] generally discusses two main things; first it discusses some gesture recognition techniques. Secondly, under each discussed tool, authors review and discuss some examples of gesture recognition applications. By identifying disadvantages and limitations of these applications, this survey addressed the existing challenges that guide the researcher for the future needs.

Technologies of hand gestures recognition are divided into three categories as shown in *Figure 1,* they are instrumented (data) glove, vision based, and colored marker approaches [2].

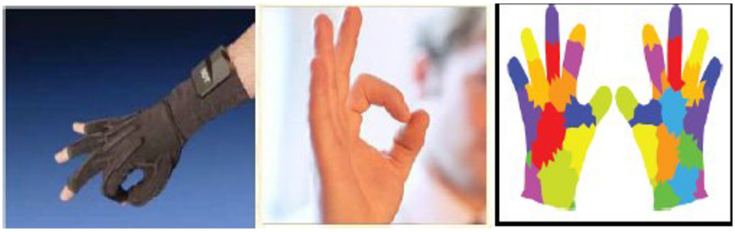


Fig.1 Categories of hand gestures recognition technologies [2]

To be specific, in our research, we are looking for using vision-based approach in our research. Vision-based approaches only require a camera to capture the images and there is no need to use additional devices. Actually, the types of camera are disparate in some metrics such as cost, range of view, accuracy, ease of use, and latency, standalone or equipped in a machine such as computer or robots.

As one of the significant studies about gesture, Ruth Breckinridge and et al. conducted a study about the importance in education through gesture and shows how it can enhance learning[6]. The children were the focus of this study while it discusses different roles of gesture in education. The gesture plays significant role in nonverbal communication, and used to convey educational concepts such as in [7] that will be discussed in details next. In addition, it was used as redundant with speech as it provides information not found in speech. Their experiment was conducted on group of children in bilingual classroom. The children speaking English or Spanish were divided into groups: first is exposed to instruction with gesture and the second group was exposed to instruction without gesture. The result showed that 92% of English-speaking and 50% of Spanish-speaking of children who were exposed to instruction with gestures benefited significantly more than the other children are, and they improved in their understanding of the conservation concept.

Actually not all of gesture roles identified previously were conducted or implemented on robots. However, still there are some countable systems in assistive educational operation field that using gesture. One of these system built by Akihiro Yorita et al. [7] developed an interactive Human-friendly assistive Partner robot with two main objectives, first objective is to discuss the applicability of robots in the remote education. While the second one is to enhance children education using partner robotic by assisting teacher in education and students monitoring. They developed three MOBiMac robots, which is a mobile PC robot that consists of two CPUs, camera, microphone, and ultrasonic sensors. The three robots have to be with students in the classroom while the teacher is in a remote controlling room. Each one of the three MOBiMac robots has a specific mode and role to do: the first one speaks according to the content of a lecture, second plays the role of a teacher (Dynamic expressions) such as moving head and hand, and the last one performs conversation and interaction with students. All of MOBiMac robots are connected with SAYA through wireless connection. SAYA is robot teacher (face robot) that can express the basic emotions but not replace the teacher. They also can control it according to teacher request from the remote room. In addition, they can understand the facial expressions of students and send them remotely as a feedback to the teacher. This kind of system is actually called edutainment, which means education plus entertainment.

As one of the new significant contributions in mutual learning using Nao, Pierre Rouanet and Pierre-Yves Oudeyer designed an interaction system between user and robot. This system allows human to teach the robot new elements that makes the robot interact with the surrounding environment [3]. The system supporta four different interfaces; one of them is based on natural human gestures while the other three are based on using mediator artifacts such as Wiimote and Laser. All of them have two main objectives, first to train and make the robot learn new objects in the surrounding environment as mentioned before while the second objective is to have an efficient interaction system between human and robot. Through gesture-based interface user can guide the robot by making hand or arm gestures. User is not restricted about the kinds of gestures that he/she chooses to make.

Then, Nao should recognize and identify that object. The gesture capturing depends on using NAO camera, which has a limited range of view. This makes the possibility of seeing the objects more restricted. Also this system may not work efficiently with non-trained user, it should be used by professional users who can identify the scope and full range of the camera view, and therefore they ensure that Nao vision of the object is clear and complete.

Kose H. et al [8] exploit the communication between hearing impaired children and a humanoid robot to assist teaching Sign Language (SL). The communication depends on hand movement, body, and face gestures. Authors conducted their study on NAO H25 humanoid robot with 106 preschool children to teach Turkish Sign Language. The main purpose of the study is to evaluate the efficiency of using the robots in the sign language learning abilities of the children. In this study, the robot tells a story that includes chosen sign language words and performs the signs of the words. Then, the children fill the color flashcards that match the signs based on what they have seen from the robot. These cards are used to evaluate the learning abilities of the children of the sign language words.

Henrik Hautop et al. present another project considered as an edutainment system [9]. It assists children with dyslexia who have difficulty in the scholastic learning, in learning linguistic structures and supports them in the performance of logic and grammatical abstraction tasks using I-BLOCKS. It is a special kind of device (Intelligent Blocks) developed by authors for ambient intelligence solutions in edutainment environments. It consists of a number of 'intelligent' building blocks where block contains the following: microprocessor, communication channels, sensors and two microphones. These blocks can be manipulated to create both physical functional and conceptual structures. There are 15 different sentences to learn each can be constructed with blocks in the following steps:

1. Child has to read the sentence from a card.
2. Choose blocks that indicate the verbs and subjects in the sentence.
3. Construct the sentences using the intelligent blocks.

As aforementioned, most of the existing robotic gesture based systems provided to children are targeted to the children with disabilities. All these works are not feasible to normal children and cannot be applicable for them, because it is more specialized and focus on enhancing children disabilities to make them act as normal children, thus normal children cannot get substantial benefits from these systems. In our proposed educational system, we would like to take advantage of the audio and visual senses in addition to children gestures to mutual interaction with robot, which is not suitable for children with disabilities. The rarity of this kind of systems encourages us to provide a robust gesture-based system for children using educational robotics. Our system aims to improve the learning abilities of children, such that the child will communicate with NAO via our system to get knowledge about the environment or learn letters, numbers, or words. We also aim to increase the knowledge base and gesture types that the robot can handle.

All system and researches above indicate the importance of using hand gestures in children learning and education which supports authors aim to go further in this research. While researchers are intended to enhance these system by using third party device that enhance the operation of gesture recognition and increase its accuracy and reliability. The next section describes Leap Motion Device.

4.2 Leap Motion Controller

Nowadays, two famous device come into researchers' minds when they start their projects about hand gesture recognition learning systems, Microsoft's Kinect and Leap Motion Controller. The major factor between the both is availability and ease of use. In this study, researchers decide to use Leap Motion Controller due to two main reasons, first of them Leap Motion is specialized for hand and identify every single bone in human hands. Second, the small size of it and the single USB wire plug where there is no need to connect an electrical jack facilitate the movement of the devise from one place to another. Table 1 shows a brief comparison between the two devices.

Table 1. Hand and fingers information extracted by Leap Motion

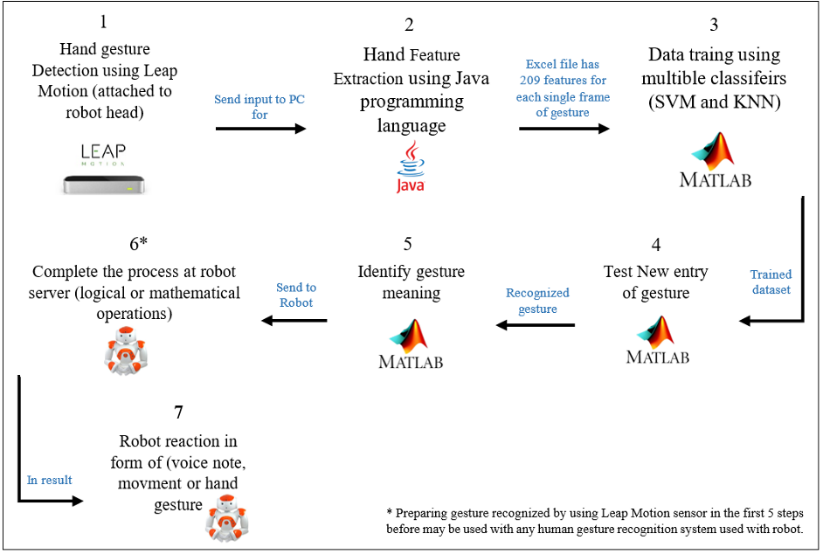
	Microsoft's Kinect sensor	Leap Motion Controller
Picture		
Depth sensor	yes	yes
Interaction Area	Range between 1.2–3.5 m from the front side	2 feet (60 cm) above the device (150°)
Wires	HD plug Electrical plug	USB plug only

Price	\$149.99 - \$199.99	Around \$20
Features	RGB camera, depth sensor and multi-array microphone	two monochromatic IR cameras and three infrared LEDs
Detection	Full body recognition	Special for hand recognition
Data	All tips and hand bones position in 3 dimension and more in Table 2.	Full body bones position in 3 dimension

Many studies and experiment was conduct on robotic environment to test the abilities of LMC. In [10] Frank Weichert et al. analyze its accuracy and robustness of the data provided by the LMC with industrial robot using a reference pen. While Jože Guna et al. intend in their study to analyze precision and reliability of Leap Motion sensor [11] by using Industrial hand on the shape of the human hand. Nowadays, many LMP system were developed to control different things such as TV [12], motors [13], robots [14, 15] etc.

5 Methodology

This research aims to expose the accuracy of hand gesture recognition using Leap Motion depth sensor to enhance intelligent system where human use hand gestures to interact with robots in learning or gaming system. The complete proposed methodology is shown in *Figure 2.*


1
Hand gesture Detection using Leap Motion (attached to robot head)
LEAP MOTION
Send input to PC for
2
Hand Feature Extraction using Java programming language
Java
Excel file has 209 features for each single frame of gesture
3
Data traing using multible classifeirs (SVM and KNN)
MATLAB
Trained dataset
4
Test New entry of gesture
MATLAB
Recognized gesture
5
Identify gesture meaning
MATLAB
Send to Robot
6*
Complete the process at robot server (logical or mathematical operations)
In result
7
Robot reaction in form of (voice note, movment or hand gesture

* Preparing gesture recognized by using Leap Motion sensor in the first 5 steps before may be used with any human gesture recognition system used with robot.

Fig. 2. Complete proposed Methodology

5.1 Gesture recognition Methodology

Hand gestures are detected using Leap motion sensor. Then, researchers used Java programing language to extract the features of gestures that they captured using Leap motion device. The extracted features include general numerical data that represent the hand such as (palm width, basis, hand direction and palm position) and details of fingers and joint position. Table 2. Shows the extracted feature from hand gesture capture by Leap motion.

Table 2. Hand and fingers information extracted by Leap Motion

Hand Information					
Frame ID	Timestamp	Hand (R or L)	Hand ID	No. of fingers	Palm width
Palm position (x, y, z)	Wrist position (x, y, z)	Hand Direction (x, y, z)	Pitch	Roll	Yaw

Fingers Information				
Finger	**Bones**			
	Metacarpal	Proximal	Intermediate	Distal
Thumb	Start (x, y, z) Always equal to Proximal Start	Start (x, y, z)	Start (x, y, z)	Start (x, y, z)
	End (x, y, z) Always equal to Proximal Start	End (x, y, z)	End (x, y, z)	End (x, y, z)
	Direction (x, y, z)	Direction (x, y, z)	Direction (x, y, z)	Direction (x, y, z)
Index	Start (x, y, z)	Start (x, y, z)	Start (x, y, z)	Start (x, y, z)
	End (x, y, z)	End (x, y, z)	End (x, y, z)	End (x, y, z)
	Direction (x, y, z)	Direction (x, y, z)	Direction (x, y, z)	Direction (x, y, z)
Middle	Start (x, y, z)	Start (x, y, z)	Start (x, y, z)	Start (x, y, z)
	End (x, y, z)	End (x, y, z)	End (x, y, z)	End (x, y, z)
	Direction (x, y, z)	Direction (x, y, z)	Direction (x, y, z)	Direction (x, y, z)
Ring	Start (x, y, z)	Start (x, y, z)	Start (x, y, z)	Start (x, y, z)
	End (x, y, z)	End (x, y, z)	End (x, y, z)	End (x, y, z)
	Direction (x, y, z)	Direction (x, y, z)	Direction (x, y, z)	Direction (x, y, z)

Pinky	Start (x, y, z)	Start (x, y, z)	Start (x, y, z)	Start (x, y, z)
	End (x, y, z)	End (x, y, z)	End (x, y, z)	End (x, y, z)
	Direction (x, y, z)	Direction (x, y, z)	Direction (x, y, z)	Direction (x, y, z)

Figure 3. As shown in the table above the values of start and end metacarpal bone of thumb are always equal to proximal bone starting point as it does not exist.

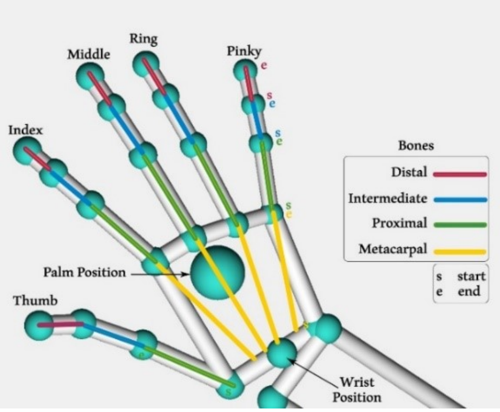


Fig. 3. Fingers bones positions

The extracted data are saved as integer values in an excel sheet file arranged as following.

1. All data collected for the training set gestures are saved in one excel sheet file.
2. It is divided into 290 labeled columns with unlimited rows.
3. Each column represent one feature and each row represent all features of one gesture frame.
4. The first column show what that gesture means (number, math sign, letter, action, word et. all).
5. The number of rows represent the number of gestures (one gesture may have multiple captured frames, one raw for each frame).

5.2 Testing Methodology

The extracted data should be trained by means of Matlab using three well known classifier (SVM, KNN, and HMM). Each gesture are trained by at least 1000 frames taken from 5-7 children. Then, new entries of frame were inserted to test the accuracy of gesture classification using the three previously mentioned classifier. Hand gesture recognition test should be conducted on 20 child who are normal or with special needs (Deaf) their ages around (4-12).

Initially, researches prepared an interface for children to teach them simple mathematical operations such as addition and subtraction. Therefore, at least 13 gestures should be recognized, 10 Gestures represent numbers from 0 to 9 and three gesture represent math signs (-, +, =).

Recognized Gestures which collected using Leap Motion and recognized correctly are send to Nao server in order manipulate them and present the result through by one of the following means or all of them:

1. Voice note (Robot should say if the answer is true or false)
2. Hand gesture that represent if the answer is right or wrong.
3. Send the result to window or mobile application.

6 Conclusion

In this paper, research aims to enhance intelligent system where human hand gestures are used to interact with robot. Improving the way of gestures recognition to be more valid and accurate needs to use sensitive hand sensor which provide depth data such as Leap Motion. Using this device beside to the robot increase the validity of hand gesture recognition where it provides detailed data about hand position and orientation in three dimensions which is not provided by embedded camera with android itself. Researchers are working now in testing stage where new children hand gestures are collected to be trained and recognized. It is highly recommended to go further in this research which enriches the intelligent computing environment and the way of HRI.

References

1. Gonzalez-Sanchez, T. and D. Puig, *Real-time body gesture recognition using depth camera.* Electronics Letters, 2011. **47**(12): p. 697.
2. Zhou, R., et al., *Robust Part-Based Hand Gesture Recognition Using Kinect Sensor.* Multimedia, IEEE Transactions on, 2013. **15**(5): p. 1110-1120.
3. .Rouanet, P., et al., *The Impact of Human&#x2013;Robot Interfaces on the Learning of Visual Objects.* Robotics, IEEE Transactions on, 2013. **29**(2): p. 525-541.
4. Mitra, S. and T. Acharya, *Gesture recognition: A survey.* Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 2007. **37**(3): p. 311-324.
5. Ibraheem, N.A. and R. Khan, Survey on various gesture recognition technologies and techniques. International Journal, 2011. **50**.
6. Church, R.B., S. Ayman-Nolley, and S. Mahootian, *The role of gesture in bilingual education: Does gesture enhance learning?* International Journal of Bilingual Education and Bilingualism, 2004. **7**(4): p. 303-319.
7. Yorita, A., et al., Remote education based on robot edutainment, in Progress in Robotics. 2009, Springer. p. 204-213.
8. Kose, H. and R. Yorganci. Tale of a robot: humanoid robot assisted sign language tutoring. in Humanoid Robots (Humanoids), 2011 11th IEEE-RAS International Conference on. 2011. IEEE.

9. Lund, H.H., P. Marti, and V. Palma. Educational robotics: manipulative technologies for cognitive rehabilitation. in Ninth international symposium On Artificial life and robotics (AROB 9th'04), Oita, JAPAN. 2004.
10. F. Weichert, D. Bachmann, B. Rudak, and D. Fisseler, "Analysis of the Accuracy and Robustness of the Leap Motion Controller," *Sensors*, vol. 13, no. 5, pp. 6380–6393, May 2013.
11. J. Guna, G. Jakus, M. Pogačnik, S. Tomažič, and J. Sodnik, "An Analysis of the Precision and Reliability of the Leap Motion Sensor and Its Suitability for Static and Dynamic Tracking," *Sensors*, vol. 14, no. 2, pp. 3702–3720, Feb. 2014.
12. I.-A. Zaiţi, Ş.-G. Pentiuc, and R.-D. Vatavu, "On free-hand TV control: experimental results on user-elicited gestures with Leap Motion," *Pers. Ubiquitous Comput.*, vol. 19, no. 5–6, pp. 821–838, Jun. 2015.
13. J. J. Kim, D. A. Gonzalez, A. Mintz, E. A. Roy, and J. Y. Tung, "Motor Control Assessment using Leap Motion: Filtering Methods and Performance in Indoor and Outdoor Environments," in *World Congress on Medical Physics and Biomedical Engineering, June 7-12, 2015, Toronto, Canada*, D. A. Jaffray, Ed. Springer International Publishing, 2015, pp. 1150–1154.
14. S. Chen, H. Ma, C. Yang, and M. Fu, "Hand Gesture Based Robot Control System Using Leap Motion," in *Intelligent Robotics and Applications*, H. Liu, N. Kubota, X. Zhu, R. Dillmann, and D. Zhou, Eds. Springer International Publishing, 2015, pp. 581–591.
15. T. V. S. N. Venna and S. Patel, "Real-Time Robot Control Using Leap Motion A Concept of Human-Robot Interaction," Apr. 2015.

Fast Signature Verification with Shared Implicit Certificates for Vehicular Communication

Hee-Yong Kwon, Mun-Kyu Lee
Department of Computer and Information Engineering,
Inha University, Incheon 22212, Korea
heeyong.kr@gmail.com, mklee@inha.ac.kr

Abstract. Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communications are considered as building blocks in many promising applications related to traffic safety and management. However, network security issues related to V2V and V2I communications should be considered together for more robust service. For example, message recipients (vehicles or road side units) should be able to verify the origin and integrity of a message by verifying the signature corresponding to the message. Because in some cases a recipient has to verify a large number of signatures in a short time period, faster signature verification is required. In this paper, we show that the signature verification process can be accelerated by sharing public key extraction processes between messages and we propose an efficient algorithm for fast signature verification. Our proposal can be applied to both separate verification and batch verification scenarios.

1 Introduction

Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communications enable vehicles to communicate with each other and road side units (RSUs) to enhance traffic safety and traffic management. Since a message may be forged by malicious adversaries in a vehicular communications environment, the integrity and authenticity of a message should be guaranteed in communication protocols. To achieve this goal, in many communication protocols, a message sender sends the message with its own digital signature and a message recipient (a vehicle or road side unit) should be able to verify whether the signature from the sender is valid or not. For example, the IEEE 1609.2 standard for Wireless Access in Vehicular Environments (WAVE) [2] uses ECDSA [3] as the underlying signature scheme. For signature verification, the recipient should know the sender's public key. In general a public key (or a chain of public keys) is provided using public key certificates. In [2], both explicit and implicit certificates are supported and an Elliptic Curve Qu-Vanstone (ECQV) certificate [1] is considered as an implicit certificate.

Because in some cases, a recipient has only limited time to verify a large number of signatures, fast signature verification is required. In our previous research [9], we

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_52

showed that by batch-processing multiple signatures, signature verification time can be shortened compared to separate verification of each signature. In this paper, we propose an efficient method to extract a public key corresponding to each signature by sharing the computation results of duplicated implicit certificates. We show that by applying the proposed method, signature verification can be significantly accelerated in both separate verification and batch verification scenarios.

2 Preliminaries

2.1 Notations

Notations used throughout this paper are summarized as follows:

b: Security parameter
E: Elliptic curve
n: Group order of elliptic curve
G: Generator of an elliptic curve group
N: Batch size
d_U: U's private key
Q_U: U's public key
$Cert_U$: U's certificate
l: Length of implicit certificate chain
e_U, P_U: U's reconstruction value
e_{CA_n}, P_{CA_n}: n-th CA's reconstruction value

2.2 Implicit Certificate

A traditional (explicit) public key certificate contains its owner's public key as well as an issuer's signature to guarantee the validity of this public key. On the other hand, an implicit certificate is a digital certificate which does not have an explicit signature or a public key. Instead, the owner's public key is reconstructed, i.e., *extracted,* by processing the certificate itself. Because an implicit certificate includes neither a signature nor a public key, the size of a certificate is smaller than that of an explicit certificate. Therefore, implicit certificates are particularly well suited for application environments where resources such as bandwidth, computing power and storage are limited, providing a more efficient alternative to traditional certificates [1].

In this paper, we consider the ECQV implicit certificate which is included in representative standards on vehicular communication such as [2]. To issue an ECQV certificate to a user U, a CA performs a protocol presented in Figure 1 with U [1]. Receiving an ephemeral key pair (k_U, R_U), the CA generates a certificate $Cert_U$ using reconstruction value P_U and user's identity U. It then sends the user a certificate as well as an additional value r which will be used for private key generation. Finally, the user generates its key pair (d_U, Q_U) using its ephemeral key and certificate.

U		CA
$k_U \in_R [1, \ \dots, n-1]$		
$R_U = k_U G$		
	$U, R_U \rightarrow$	
		$k \in_R \ [1, \ \dots, n-1]$
		$P_U = R_U + kG$
		$Cert_U = Encode(P_U, U, \ *)$
		$e = H(Cert_U)$
		$r = ek + d_{CA}(mod\, n)$
	$\leftarrow r, \ Cert_U$	
$e = H(Cert_U)$		
$d_U = ek_U + r(mod\, n)$		
$Q_U = eP_U + Q_{CA}$		

Fig. 1. ECQV certificate issuance protocol

It should be noted that the final step where the user extracts its public key Q_U can be done by anyone who possesses the certificate $Cert_U$ and the CA's public key Q_{CA}, because e is the hash of $Cert_U$ and P_U is encoded in $Cert_U$. If s/he does not have Q_{CA}, but has the CA's implicit certificate $Cert_{CA}$, then the above public key extraction procedure can be applied recursively. Let us assume that the length of an implicit certificate chain is l and we have $Cert_U$, $Cert_{CA_1}, \dots, Cert_{CA_{l-1}}$ as well as the public key Q_{CA} of a CA. Then, the user's public key can be extracted as follows:

$$Q_U = e_U P_U + e_{CA_1} P_{CA_1} + \cdots + e_{CA_{l-1}} P_{CA_{l-1}} + Q_{CA} \ . \quad \textbf{(1)}$$

2.3 Modified ECDSA

A signature in standard ECDSA [3] is a pair of (r, s) where r is the x-coordinate of some point R which is temporarily generated through a signature generation procedure. To verify a signature, one reconstructs R' which should be the same as R in a legitimate signature and compares r with the x-coordinate of R' instead of directly comparing R' and R [3]. In this paper, however, we consider a slightly modified version of ECDSA to apply batch verification according to the convention in the literature [4-6]. The security level of the modified ECDSA is equivalent to that of the standard ECDSA [4-6]. A signature generated by the modified ECDSA is (R, s) instead of (r, s). In signature verification, R' and R are directly compared. To be precise, $R' = H(m)s^{-1}G + rs^{-1}Q_U$, where r is the x-coordinate of R, $H(m)$ is the hash of message m, G is the generator of the elliptic curve group, and Q_U is U's public key.

Then, the signature verification can be actually done by verifying if

$$R = H(m)s^{-1}G + rs^{-1}Q_U \quad \textbf{(2)}$$

holds, given a message m, a signature (R, s) for m, and a public key Q_U.

2.4 Batch Verification

Batch verification is a method to verify a number of signatures at once instead of verifying each signature separately. In [7], Bellare et al. proposed three batch verification methods called Random Subset Test, Small Exponents Test, and Bucket Test. In this paper, we adopt the Small Exponents Test (SET) [7] which is the most frequently used in the literature. In [7], SET to verify multiple exponentiations over a multiplicative group is given as a building block for SET for signature verification. Algorithm 1 below describes it in our context of elliptic curve group.

Algorithm 1. Small Exponents Test (SET) to verify multiple point multiplications (modified version of Fig. 2 in [7])

```
Input: $l$ (security parameter), $G$ (generator point), and $(x_1, R_1), (x_2, R_2), \dots, (x_N, R_N)$ with
    $x_i \in Z_n$ and $R_i \in E$
Check: That $\forall i \in \{1, \dots, N\} : R_i = x_i G$
    1. Pick $z_1, \dots, z_N \in \{0,1\}^l$ at random.
    2. Compute $x = \sum_{i=1}^{N} z_i x_i \bmod n$, and $R = \sum_{i=1}^{N} z_i R_i$.
    3. If $xG = R$ then accept, else reject.
```

It is obvious that $xG = R$ if $R_i = x_i G$ for all $i = 1, \dots, N$. However, there could be the case that $xG = R$ even though $R_i \neq x_i G$ for some i. It is proved that this probability is very small, i.e., it is at most 2^{-l} [7].

A batch verification equation for modified ECDSA is derived as

$$\sum_{i=1}^{N} z_i R_i = \sum_{i=1}^{N} \{z_i H(m_i) s_i^{-1} G + z_i r_i s_i^{-1} Q_i\} \tag{3}$$

by applying SET to (2). This equation can be easily transformed to a more efficiently computable form, i.e., to verify if

$$\sum_{i=1}^{N} (z_i R_i - z_i H(m_i) s_i^{-1} G - z_i r_i s_i^{-1} Q_i) \tag{4}$$

is 0. Let V be the computed result in (4). If all of input signatures are valid, V will always be zero. On the other hand, $V = 0$ guarantees that all signatures are valid with probability $\geq 1 - 2^{-l}$ [7]. If V is not zero, it means that there is at least one invalid signature among input signatures.

3 Proposed Method

To verify a signature, a user's public key should be extracted from an implicit certificate, which is shown in (1). Our proposal is based on the observation that some part in (1) can be shared and reused between different users. To explain this, we consider an example scenario for vehicular communication, where a vehicle can be regarded as a user U. That is, a vehicle has its own public key. When this vehicle sends a message and the corresponding signature to another vehicle or RSU, the verification of this signature will be possible after extracting the public key of this vehicle. The key extraction procedure may involve a chain of certificates. For example, we may consider a certification hierarchy depicted in Figure 2. In this figure, the root CA certifies the identity of Korean root CA by issuing a certificate, the Korean root CA certifies Hyundai CA, and Hyundai CA issues certificates for vehicles it manufactures. A vehicle should have all certificates on this certification chain from the root CA to itself. In this case, it is expected that there should be an internal node in this tree which has many branches, which implies that many vehicles may share certificates, e.g., they may have identical manufacturer or country certificates. For example, in Figure 2, vehicles manufactured by Hyundai share a significant part of the certificate chain, that is, all certificates from Hyundai to root CA are identical. If many Hyundai vehicles are sending messages with signatures to an identical recipient, this recipient can do an operation to extract the public key of Hyundai only once and reuse the result for the other Hyundai vehicles.

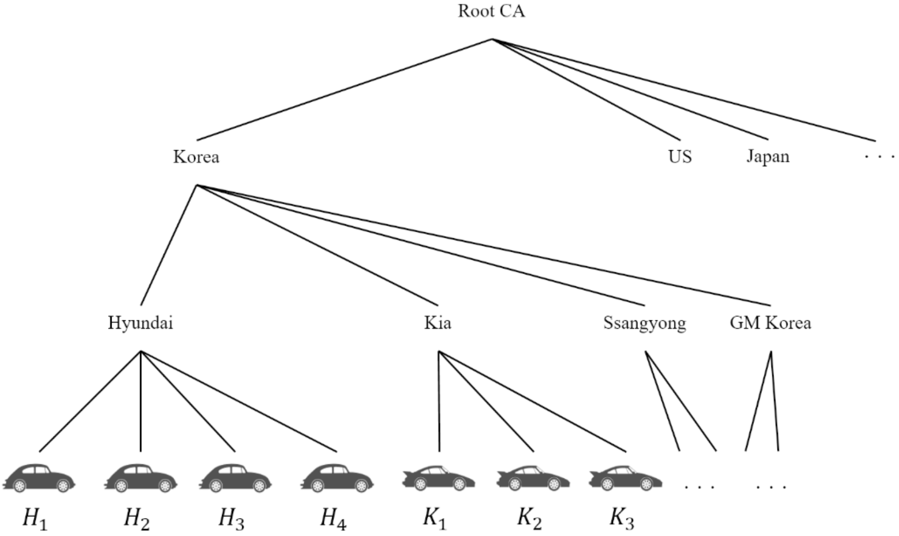


Fig. 2. Example tree of implicit certificate chains

3.1 Sharing Public Key Extraction Processes

In this subsection, we present a novel method to speed up the key extraction process by sharing some part of a certificate chain. First, we see that (1) is represented by a multi-scalar multiplication

$$Q_U' = e_U P_U + e_{CA_1} P_{CA_1} + \cdots + e_{CA_{l-1}} P_{CA_{l-1}} \ , \qquad \textbf{(1)}'$$

and a point addition by Q_{CA}. The most well-known algorithm for efficient multi-scalar multiplication is a variant of Straus's algorithm [8]. In a multi-scalar multiplication, this algorithm computes $\mathrm{e}_1 P_1 + e_2 P_2 + e_3 P_3 + \cdots + e_N P_N$ by initializing an intermediate result $R = O$ and repeating

1) w doublings on R, and
2) accumulation of each w-bit portion $e_i' P_i$ to R for $i = 1, \ldots, N$,

from MSB to LSB, where w is a window size. For example, in the last step in the computation of (1)', the accumulated value to R is

$$T = (e_{CA_{l-1}}\, mod\ 2^w)P_{CA_{l-1}} + (e_{CA_{l-2}}\, mod\ 2^w)P_{CA_{l-2}} + \cdots + (e_{CA_1}\, mod\ 2^w)P_{CA_1} + (e_U\, mod\ 2^w)P_U\ .$$

The actual computation of T is done in the order of

$$\begin{aligned} &(e_{CA_{l-1}}\, mod\ 2^w)P_{CA_{l-1}} \\ &\rightarrow (e_{CA_{l-1}}\, mod\ 2^w)P_{CA_{l-1}} + (e_{CA_{l-2}}\, mod\ 2^w)P_{CA_{l-2}} \\ &\rightarrow (e_{CA_{l-1}}\, mod\ 2^w)P_{CA_{l-1}} + (e_{CA_{l-2}}\, mod\ 2^w)P_{CA_{l-2}} + (e_{CA_{l-3}}\, mod\ 2^w)P_{CA_{l-3}} \\ &\qquad\vdots \\ &\rightarrow (e_{CA_{l-1}}\, mod\ 2^w)P_{CA_{l-1}} + (e_{CA_{l-2}}\, mod\ 2^w)P_{CA_{l-2}} + \cdots + (e_{CA_1}\, mod\ 2^w)P_{CA_1} \\ &\qquad + (e_U\, mod\ 2^w)P_U\ . \end{aligned}$$

Note that some of these results can be stored in memory if required and reused to extract another public key. In the example in Figure 2, l is 3 for vehicle H_1, and the last step of public key extraction for H_1 will be the accumulation of

$$T_1 = (e_{CA_{Korea}}\, mod\ 2^w)P_{CA_{Korea}} + \left(e_{CA_{Hyundai}}\, mod\ 2^w\right)P_{CA_{Hyundai}} + (e_{H_1}\, mod\ 2^w)P_{H_1},$$

and that for vehicle H_2 will be the accumulation of

$$T_2 = (e_{CA_{Korea}}\, mod\ 2^w)P_{CA_{Korea}} + \left(e_{CA_{Hyundai}}\, mod\ 2^w\right)P_{CA_{Hyundai}} + (e_{H_2}\, mod\ 2^w)P_{H_2}.$$

Because T_1 and T_2 share the following part,

$$(e_{CA_{Korea}}\, mod\ 2^w)P_{CA_{Korea}} + \left(e_{CA_{Hyundai}}\, mod\ 2^w\right)P_{CA_{Hyundai}},$$

it can be computed only once when T_1 is computed and reused for computation of T_2. Even though we explained the last iteration of the computation of (1)' for notational simplicity, the above idea can be applied to every iteration. Let k be the number of identical implicit certificates between two public key extraction processes. Then, the cost for each iteration drops down from $w + l$ (w doublings and l additions) to $w + l - (k-1)$ for the second public key. In other words, the cost for k terms is decreased to that of a single term.

3.2 Performance Analysis for Public Key Extraction

We analyze how much speed-up we can get when the proposed idea is applied to public key extraction. We assume that we use the w −NAF method to construct w −bit windows. According to [9], the cost for extracting a single public key is

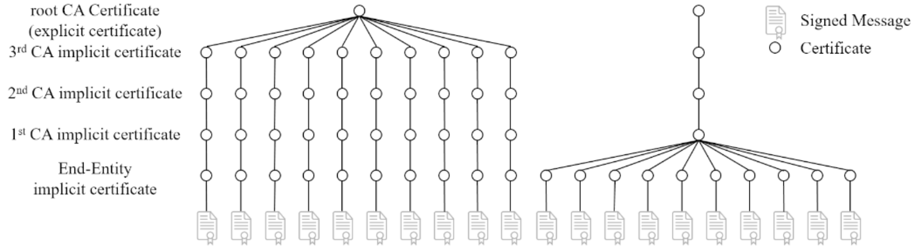
$$(l + 2b)\ D + \left\{\frac{2bl}{w+1} + (2^{w-2} - 1)l + 1\right\} A\ , \qquad \textbf{(5)}$$

and the cost for N public keys is N times (5), where A is a point addition, and D is a point doubling.

If we apply the proposed method, the saved cost for extracting a public key is $kD + \left\{\frac{2b(k-1)}{w+1} + (2^{w-2} - 1)k\right\} A$, where $k\ (k \geq 1)$ is the number of shared certificates. As a result, the cost for extracting N public keys is reduced to

$$\left[(l+2b)\,D+\left\{\frac{2bl}{w+1}+(2^{w-2}-1)l+1\right\}A\right]N$$
$$-\left[kD+\left\{\frac{2b(k-1)}{w+1}+(2^{w-2}-1)k\right\}A\right](N-1)\ . \qquad \textbf{(6)}$$

Figure 3 presents the worst and best cases when there are 10 signatures. The worst case is when there is no shared implicit certificate between signatures. The total cost for this case is 10 times (5). In the best case, all messages share the implicit certificates from first to last CAs. But because the root CA certificate is an explicit certificate, we cannot apply the above saving technique to the root CA. Consequently, the total public key extraction costs for 10 signatures are $2{,}573\,D + 1{,}040\,A$ and $2{,}600\,D + 1{,}997\,A$, respectively, for the best and worst cases, when $w = 5, b = 128$.



(a) Worst case (a) Best case

Fig. 3. The worst and best cases of sharing implicit certificates when the implicit certificate chain length is 4 (The certificate of root CA is an explicit certificate. We do not count this for chain length.).

4 Performance Analysis for Signature Verification

In this section, we analyze the cost of signature verification, considering two factors. The first factor is whether signatures are verified separately or as a batch. The second factor is whether the intermediate results in public key extraction processes are shared or not. As a result, we consider the following four combinations:

a) Separate verification, no sharing (basic method)
b) Batch verification, no sharing [9]
c) Separate verification, with sharing
d) Batch verification, with sharing

4.1 Cost Analysis

The overall cost is composed of the two costs; public key extraction cost and signature verification cost. We examine the latter. According to [9], the verification costs for separate and batch verification are

$$\left[(1+2b)\,D+\left\{\frac{4b}{w+1}+(2^{w-2}-1)\right\}A\right]N\ , \tag{7}$$

and

$$(2N+2b)\,D+\left\{\frac{b(3N+2)}{w+1}+2N(2^{w-2}-1)\right\}A\ , \tag{8}$$

respectively. The costs for public key extraction were already analyzed in (5) and (6). Then, the overall costs for the above four combinations are as follows:

a) (N times (5)) + (7) [9]
b) (N times (5)) + (8) [9]
c) (6) + (7)
d) (6) + (8)

4.2 Comparison using practical values

Table 1. Comparison of Four Combinations (Number of Point Additions and Doublings)

N		16		32		64	
l		Sep	Batch	Sep	Batch	Sep	Batch
2	No Sharing	11,323	7,312	22,645	14,325	45,291	28,352
	Sharing	11,203	7,192	22,397	14,077	44,787	27,848
	Saving	1.06%	1.64%	1.10%	1.73%	1.11%	1.78%
4	No Sharing	12,944	8,933	25,888	17,568	51,776	34,837
	Sharing	11,304	7,293	22,499	14,179	44,888	27,949
	Saving	12.67%	18.36%	13.09%	19.29%	13.30%	19.77%
6	No Sharing	14,565	10,555	29,131	20,811	58,261	41,323
	Sharing	11,405	7,395	22,600	14,280	44,989	28,051
	Saving	21.70%	29.94%	22.42%	31.38%	22.78%	32.12%

Let window size w be 5 and security parameter b be 128. The costs of four algorithms with different number of signatures (N) and length of implicit certificate chains (l) are presented in Table 1. Even though the complexities of a point addition and a point doubling are slightly different, we assume that their complexities are the same, which is a frequently used assumption in the literature for simpler analysis. In Table 1, there are three rows for each l, which represent the two options whether to share intermediate results or not, and the savings by sharing. The values in the "Sharing" row present the best cases. On the other hand, as explained in section III, the worst case in Figure 3 consumes essentially the same cost as that of an algorithm which does not consider sharing intermediate results, i.e., the values in the "No Sharing" rows. Therefore, the actual number of point operations when intermediate results are shared may fall somewhere between the two extreme cases. The values in the "Saving" rows are the maximum improvement corresponding to the best cases. In the table, there are two columns for each N to represent separate and batch verifications. For example, for $N = 16, l = 2$, the costs of combination *a), b), c)*, and *d)* are 11,323, 7,312, 11,203, and 7,192 point operations, respectively. According to the table, by sharing intermediate results in public key extraction, we can save up to 32.12% of the computational costs. We also see that *a)* is the worst combination and *d)* is the best combination. On average, the cost of *d)* is about 55.58% of that of *a)*. It means that we can accelerate signature verification by about 1.82 times if both batch verification technique and sharing technique are applied.

5 Conclusion

We showed that if we share intermediate results in public key extraction, we can save up to 32.12% of the computational costs in the best case. By combining this technique with the batch verification technique in the previous work, about 44.42% of the costs can be saved. It would be an interesting future work to verify our theoretical analysis result by implementing the proposed method.

Acknowledgments. This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2016-H8501-16-1008) supervised by the IITP (Institute for Information & communications Technology Promotion).

References

1. Certicom Research: Standard for Efficient Cryptography 4: Elliptic Curve Qu-Vanstone Implicit Certificate Scheme (ECQV). January 24 (2013)
2. IEEE: Standard for Wireless Access in Vehicular Environments-Security Services for Applications and Management Messages. IEEE Std 1609.2-2013 (2013)
3. NIST FIPS 186-4. Digital Signature Standard (DSS). July 2013 (2013) 19–26
4. Jung Hee Cheon, Mun-Kyu Lee: Improved batch verification of signatures using generalized sparse exponents. Computer Standards & Interfaces 40 (2015) 42–52
5. A. Antipa, D. Brown, R. Gallant, R. Lambert, R. Struik, S. Vanstone: Accelerated verification of ECDSA signatures. SAC 2005 LNCS, Vol. 38972006 307–318
6. C. Boyd, C. Pavlovski: Attacking and repairing batch verification schemes. Proc. of Asiacrypt 2000, LNCS, Vol. 1976, Springer-Verlag (2000) 58–71
7. Mihir Bellare, Juan A. Garay, Tal Rabin: Fast batch verification for modular exponentiation and digital signatures. Eurocrypt '98 (1998) 236–250
8. Ernst G. Straus: Addition chains of vectors (problem 5125). American Mathematical Monthly 70 (1964) 806–808
9. HeeYong Kwon, Mun-Kyu Lee: Fast Batch Verification of ECQV Certificate-based Signatures. KIISE Summer Workshop on Computer Communications (2016). (in Korean)

Design of an Adhoc Testbed for IoT and WSAN Applications using Raspberry Pi

Hiroya Oda, Elis Kulla, Ryo Ozaki, Noritaka Nishihara

Abstract With recent advancement in wireless sensor technologies, IoT and WSAN applications have emerged and new algorithms and protocols have been proposed. Many simulations have been conducted to test these new protocols and algorithms. However in order to verify the new protocols and algorithms in real experiments, we designed an adhoc testbed, where we plan to conduct experiments for IoT and WSAN applications. In this paper we describe our testbed and show some results, while investigating the performance of the network for different packet sizes.

1 Introduction

In the next generation of wireless communication systems, such as 5G [1, 2], the need for rapid deployment of independent mobile users will increase [3]. Significant examples include establishing networks which are survivable, efficient and able to communicate dynamically for emergency/rescue operations, sensing in Smart Cities [4], Internet of Things (IoT) [5] and so on. Such networks cannot rely on centralized management, but rather distributed and autonomous operation. In general they can be considered as applications of Adhoc Networks, or if nodes are mobile, known as Mobile Adhoc Networks (MANETs).

A MANET is a collection of wireless mobile hosts that can dynamically establish a temporary network without any aid from fixed infrastructure. The mobile hosts act

Hiroya Oda
Okayama University of Science, Department of Information ad Computer Engineering, 1-1 Ridai-cho, Kita-ku, 700-0005 Okayama, Japan
e-mail: t16jm01oh@ous.jp

Elis Kulla, Ryo Ozaki, Noritaka Nisihara
Okayama University of Science, Department of Information ad Computer Engineering, 1-1 Ridai-cho, Kita-ku, 700-0005 Okayama, Japan
e-mail: kulla@ice.ous.ac.jp, ozaki@ice.ous.ac.jp, nisihara@ice.ous.ac.jp

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_53

as routers for each other and they are connected via wireless links. Mobility and the absence of any fixed infrastructure make MANET very attractive for rescue operations and time-critical applications. Wireless Sensor Networks (WSN), Wireless Sensor Actor Networks (WSAN), Vehicular Ad-hoc Networks (VANET) and other new and existing network technologies are based on MANETs. Thus, in order to expand the usage of MANET, extensive research and verification are required.

A lot of research for MANETs is going on, usually in simulations, because in general, a simulator can give a quick and inexpensive evaluation of protocols and algorithms. However, experimentations in the real world are very important to verify the simulation results and to revise the models implemented in the simulator.

In our paper, we show the design and implementation of a testbed for MANET, where we can test and verify different algorithms. By adding Raspberry Pi devices to the testbed, the network becomes heterogeneous. Raspberry Pi can be converted in a sensor node by connecting sensors to its 40-pin GPIO. On the other hand, it can be used to control different robots (we have implemented for RAPIRO [6]), and because Raspberry Pi is portable, our testbed network has mobile sensors and mobile actors, which makes it suitable to implement WSN, IoT and WSAN applications in it.

Moreover, we conduct some experiments in order to analyze the performance of BATMAN routing protocol in a mixed outdoor and indoor environment. We implement simple scenarios in our adhoc testbed, which consists of 10 mobile machines (Note PCs and Raspberry Pi). We investigate the network performance considering throughput and delay, by sending data in different format.

The structure of the paper is as follows. In Section 2, some related works are discussed. An overview of BATMAN routing protocol is given in Section 3. In Section 4, we describe the design and implementation of our adhoc testbed. We discuss the experimental results and evaluate the performance of the testbed, in Section 5. Finally, we draw conclusions in Section 6.

2 Related Work

The authors in [7, 8, 9, 10] conducted many experiments with their MANET testbed. They carried out the experiments with different routing protocols such as OLSR and Better Approach to Mobile Ad-hoc Networks (BATMAN) [11] and found that throughput of TCP was improved by reducing Link Quality Window Size (LQWS), but there were packet loss because of experimental environment and traffic interference. Moreover, they found that the node join and leave operations affect more the TCP throughput and Round Trip Time (RTT) than UDP [9]. In [10], they showed that BATMAN buffering feature showed a better performance than Ad-hoc On-demand Distance Vector (AODV), by handling the communication better when routes changed dynamically.

In [12], the authors implemented multi-hop mesh network called Massachusetts Institute of Technology (MIT) Roofnet, which consists of about 50 nodes. They consider the impact of node density and connectivity in the network performance. The authors show that the multi-hop link is better than single-hop link in terms of throughput and connectivity. In [13], the authors analyze the performance of an outdoor ad-hoc network, but their study is limited to reactive protocols such as AODV [14] and Dynamic Source Routing (DSR) [15].

The authors of [16] compare the performance of two typical routing protocols, AODV and DSR, in real multi-hop environment. Apart from testing the end-to-end packet loss, delay and routing path parameters, they also assess the performance of AODV and DSR in terms of some applications based on IOT, such as Radio Frequency Identification (RFID) service, voice service and temperature monitoring service.

The paper in [17] proposes an original solution to integrate and exploit MANET overlays, collaboratively formed over WSNs, to boost urban data harvesting in IoT.

3 Overview of BATMAN Routing Protocol

In the well-known OLSR, there was a serious synchronization problem between the topology messages and the routing information stored inside every node. In other words, a mismatch between what is currently stored in the routing tables and the actual topology of the network may arise. This is due to the propagation time of the topology messages. Routing loops are the main effect of such problem. To solve this problem, BATMAN has been introduced. In BATMAN, there is no topology message dissemination. Every node executes the following operations.

1. Sending of periodic advertisement messages, called OriGinator Message (OGM). The size of these messages is just 52 bytes, containing: the IP address of the originator, the IP address of the forwarding node, a Time To Live (TTL) value and an increasing Sequence Number (SQ).
2. Checking of the best one-hop neighbor for every destination in the network by means of a ranking procedure.
3. Re-broadcasting of OGMs received via best one-hop neighbor.

The BATMAN uses a timer for sending OGMs, which are used by BATMAN nodes to create routes for all the nodes in the network. In few words, every node ranks its neighboring nodes by means of a simple counting of total received OGMs from them. The ranking procedure is performed on OriGinator (OG) basis, i.e. for every originator. Initially, for every OG, every node stores a variable called Neighbor Ranking Sequence Frame (NBRF), which is upper bounded by a particular value called ranking sequence number range. We suppose that there is a rank table in every node which stores all the information contained in the OGMs. Whenever a new OGM is being received, the receiving node executes the following steps.

1. If the sequence number of the OGM (SQ(OGM)) is less than the corresponding NBRF, then drop the packet.
2. Otherwise, update the NBRF=SQ (OGM) in the ranking table.

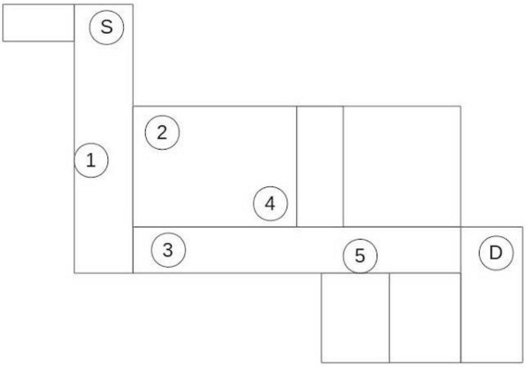


Fig. 1 Experimental environment.

3. If SQ (OGM) is received for the first time, store OGM in a new row of the rank table.
4. Otherwise, increment by one the OGM count or make ranking for this OGM.

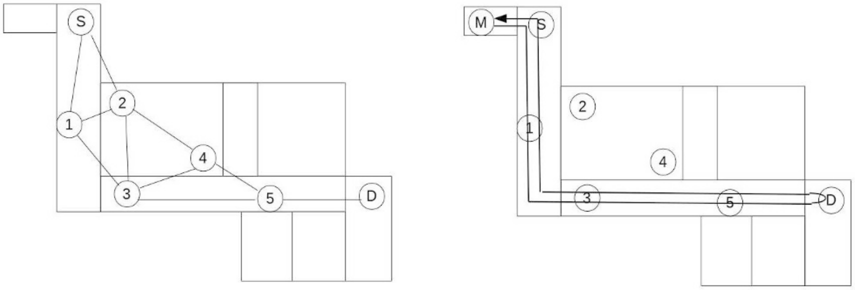
Finally, the ranking procedures select the best one-hop neighbor, the neighbor which has the highest rank in the ranking table. Let us note that the same OGM packet is used for: link sensing, neighbor discovery, bi-directional link validation and flooding mechanism. Other details on BATMAN can be found in [11].

4 Design and Implementation of MANET Testbed

4.1 Testbed Description

We implemented our testbed in our academic environment around our five-floor academic building, as shown in Fig. 1. The experiments were conducted in a mixed environment, where nodes are put inside rooms or in the outside compartments (stairs etc.) of Building 18 of Okayama University of Science campus. We used six Note PCs (Panasonic CF-T7 Let's Note model) equipped with external USB wireless cards (BUFFALO WLI-UC-GNM LAN Adapter) [18] and two Raspberry Pi 2 model B devices equipped with the same wireless cards.

The Note PC machines operate on UBUNTU 14.04 LTS OS [19], with kernel 3.16.0-30, while Raspbery Pi machines operate on the native Raspbian Jessie OS [20]. We set up their wireless cards to operate with transmitting power of 16+/-1dBm and receiving sensitivity of -80dBm.



(a) Static

(b) Mobile

Fig. 2 Static and Mobile Scenarios.

The traffic in the network is sent by Distributed Internet Traffic Generator (D-ITG) software, version 2.7.0 Beta2, which is an Open Source Traffic Generator [21]. With D-ITG, we can inject different type of data flows in the network. After finishing the transmission, D-ITG offers decoding tools to get information about different metrics along the whole transmission duration.

Our testbed provides an experimental platform for evaluating protocols and algorithms using realistic parameters. In this testbed, we can implement different topology scenarios and analyze different routing protocols considering different metrics.

4.2 Experimental Scenario Settings

In order to evaluate our testbed, we implemented two physical scenarios, based on node position and movement, and different cases, based on the type of data sent in the network.

Two scenarios can be seen in Fig. 2. In Static scenario, all nodes are static. S and D are Raspberry Pi devices, and we think of them as sensor and actor nodes, respectively. When testing the settings of the testbed, we noticed that some packets were sent directly from S to D or by using only two hops. But we wanted our network to use multihop routes. Therefore, we setup MAC filtering in order to limit the connectivity distance, as shown in Fig. 2(a). The results of Static scenario will be used as basis when compared to Mobile scenario. In Mobile scenario, we introduce a mobile node M, which does not use filtering (it can connect to every node in the network) and moves from the location of node S to location of node D and back to node S in 3 minutes, as shown in Table 1. Mobility of node M, brings topology changes during movement, so it makes the network more dynamic.

Table 1 Node M Mobility pattern

Time	Action	Description
$0s - 30s$	Warm-Up	Node M needs some time to enter the network by sending and receiving OGMs.
$30s - 90s$	Move1	Moving from node S to D.
$90s$	Turn	Turn at the location of node D.
$90s - 150s$	Move2	Moving from node D to node S.
$150s - 180s$	Cool-Down	In order to match the Warm-up period.

4.3 Considerations

An experimental environment gives us realistic results and can be used to test and verify real aspects of MANETs and its applications. However, our experiments are based on some simple considerations, in order to make the results consistent and usable by other researchers.

- We analyze the data based on two metrics: Throughput and Round-trip delay.
- Moving nodes are moved by human force. We carry laptops, while walking in the experimental area. When reaching turning points, we stop for about 3 seconds before resuming movement.
- We run the experiments 10 times for every setting. This is important, so we can get an average measure of network performance.
- We analyze the effect of multihop communication and mobility in real networks, and we try to implement scenarios with a high degree of similarity with realistic applications.

In our testbed, we have a systematic traffic sources we could not eliminate. There are other wireless Access Points (APs) interspersed within the campus. This brings interferences occupying the available bandwidth, which is typical in an academic scenario.

Table 2 Data Types

Parameter	Case1	Case2	Case3	Case4
Packet Size (*Bytes*)	600	900	1200	1500
Packet Rate (*pps*)	104	69	52	42
Sent Throughput (*kbps*)	500	500	500	500

5 Experimental Results

We sent data through the network, from node S to node D, in four different patterns and evaluated the performance of the transmission based on two metrics: Throughput and Delay. We wanted to investigate the effect of packetsize in multihop communications, so we differ the packetsize and packetrate of the data, and keep the throughput constant at 500kbps ((See Table. 2)). The results are shown in average values, in Fig. 3 and in time-domain representation in Figs. 4 and 5.

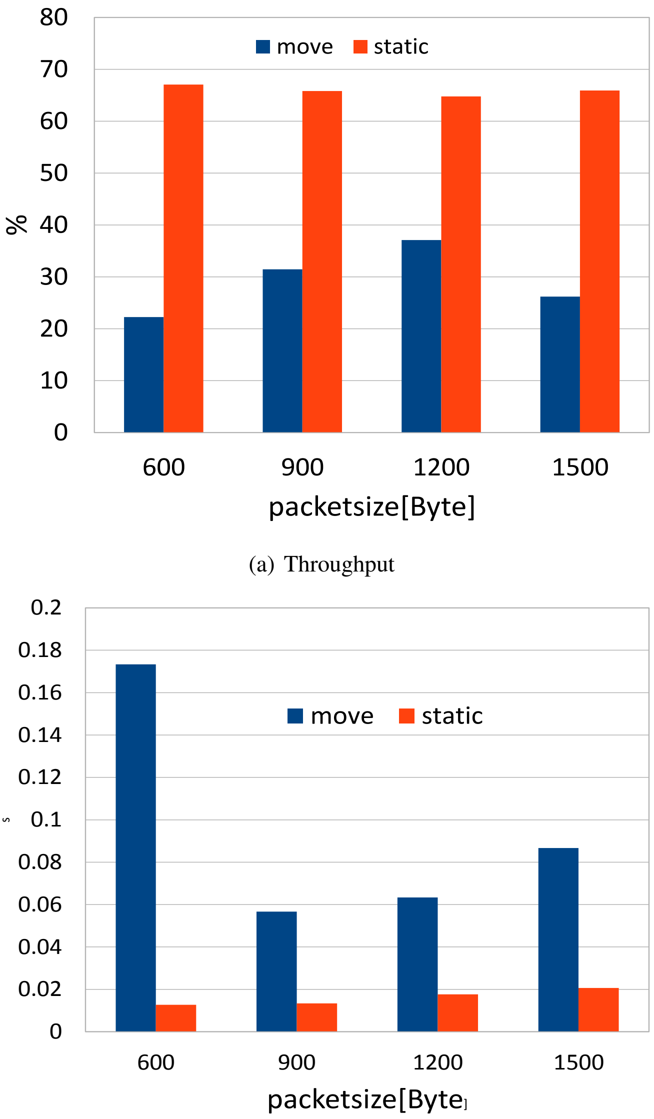
We can see from the average values of throughput (Fig. 3(a)), that for Static scenario, throughput is more than 60% and similar in all cases, which is an acceptable value for wireless multihop communication. However, in Mobile scenario, the throughput rate drops to less than 40%. In this scenario, we also notice the effect of packetsize in the performace. Greater packetsizes have smaller packetrate, so the network is not congested and vice-versa. In cases with frequent route changes, congestion lowers the performance, because it means more possible lost packets. But in Case4, where packetsize is 1500Bytes (MTU value), the throughput rate drops, because a big packet means a lot of operation time before forwarding. Big packets keep the nodes busy for longer periods of time. The same applies for the delay metric.

We also investigate the performance of the network in different times during the experiment. In all the cases, throughput and delay are almost constant in the Static scenario. The interesting part is the Mobile scenario. In general, As node M moves from node S to node D, we notice a drop in performance (decreased throughput and increased delay). At the beginning, the routes are stabilized (Warm-up period), so the packets arrive correctly at node D. While node M starts to move, the topology becomes dynamic, and because the routes change, there are more lost packets, due to occurring of unavailable links and routes. Throughput becomes 0% in most cases, at the period of time 130 – 150, when node M is almost back to node S.

We would like to investigate on that more, in out future works, but it seems that routes are disturbed by the moving node, and there might be a false-positive effect because of mobility. Node D is the destination, so node S and other relay nodes are interested in OGM packets coming from node D. Those packets transverse the network from node D to node S, the same direction node M is moving. In this period of time, it might happen that OGM packets are stored inside node M's memory and forwarded at a later time (remember that BATMAN creates a buffer in each node for this purpose). When forwarded they may create the idea that node M is the best ranked neighbour, but that may be a false positive, because node M may not even have a valid route to node D anymore.

6 Conclusions and Future Works

In this paper, we showed the design and implementation of a testbed for MANET, where we can test and verify different algorithms. It is suitable to implement WSN,



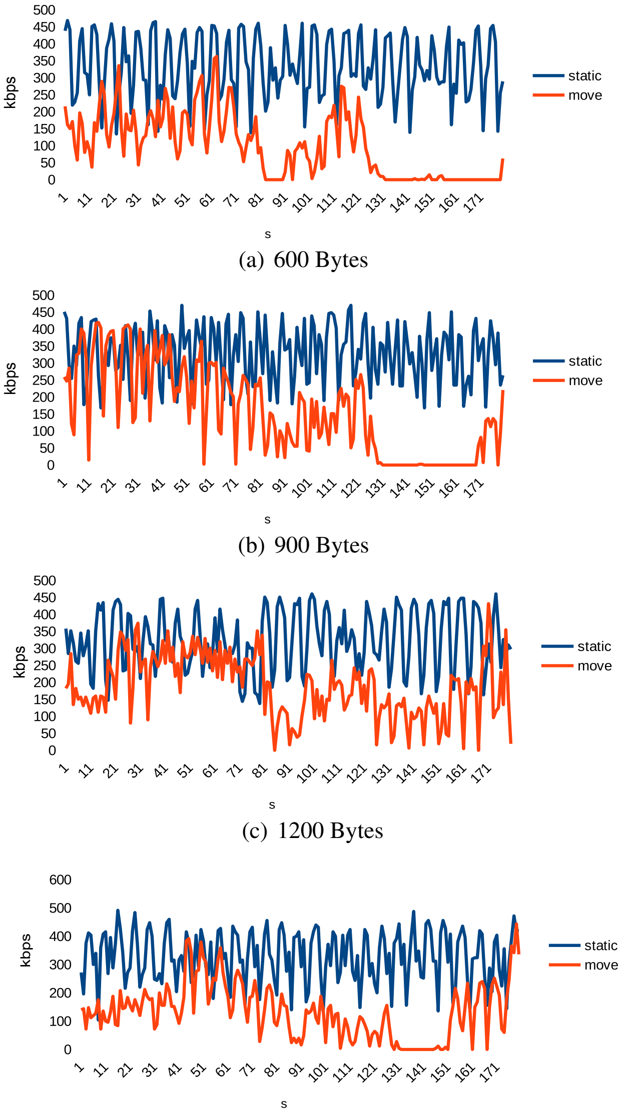
(a) Throughput

(b) Delay

Fig. 3 Average Results for Throughput and Delay.

IoT and WSAN applications in it. We conducted some experiments in order to analyze the performance of BATMAN routing protocol based on throughput and delay. From simulation results we concluded the following.

- For small packet sizes and very big packet sizes, the performance drops, due to congestion of network and nodes, respectively.
- Mobility of nodes, brings topology changes, route changes and decreased performance.



(a) 600 Bytes

(b) 900 Bytes

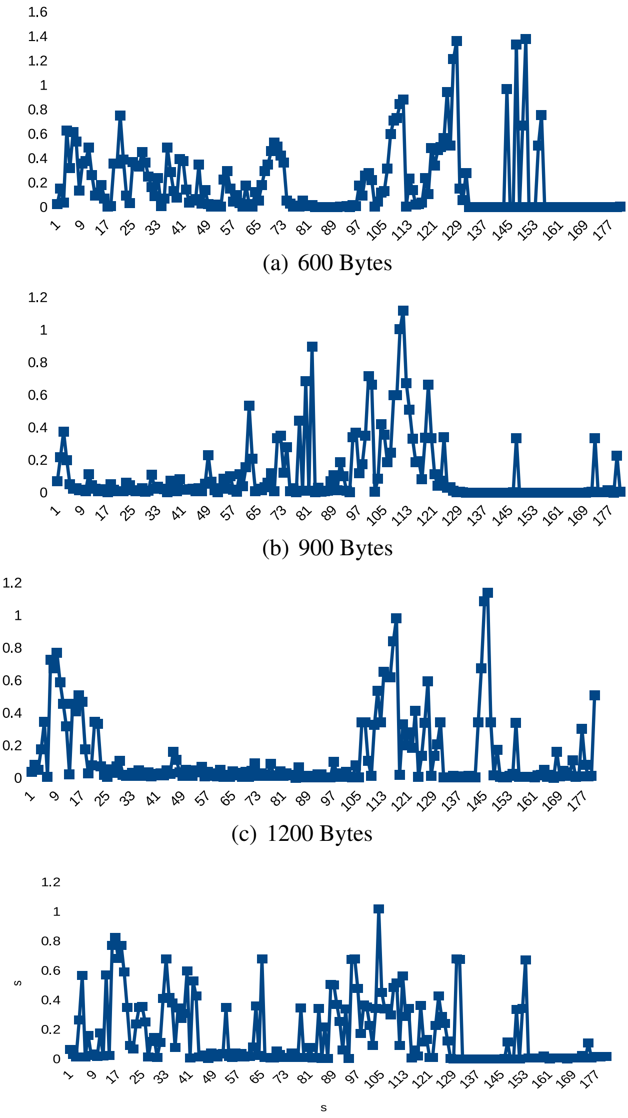
(c) 1200 Bytes

(d) 1500 Bytes

Fig. 4 Time-Domain Results for Throughput.

- Mobility of nodes, brings topology changes, route changes and decreased performance.
- The direction of movement may have an effect on Ranking Procedure of BATMAN.

In out future works, we would like to continue our implementations of sensor node and actor node in the network. We would also like to develop new algorithms



(a) 600 Bytes

(b) 900 Bytes

(c) 1200 Bytes

(d) 1500 Bytes

Fig. 5 Time-Domain Results for Delay.

that will be suitable for WSAN applications. Moreover, we plan to implement a smart Laboratory, where home automation and IoT will be researched and improved.

References

1. C. Felita and M. Suryanegara, “5G Key Technologies: Identifying Innovation Opportunity”, in International Conference QiR (Quality in Research), pp. 235-238, June 2013.
2. A. Gohil, H. Modi and S. Patel, “5G Technology of Mobile Communications: A Survey”, International COnference on INtelligent Systems and Signal Processing (ISSP-2013), pp. 288-292, March 2013.
3. X.Y. Li, “Wireless Ad Hoc and Sensor Networks: Theory and Applications”, Cambridge University Press, 2008, Hardcover Print.
4. S. Pellicer, G. Santa, A.L. Bleda, R. Maestre, A.J. Jara, A. Gomez Skarmeta, ”A Global Perspective of Smart Cities: A Survey”, in Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS-2013) pp.439,444, 3-5 July 2013. doi: 10.1109/IMIS.2013.79
5. L. Atzoria, A. Ierab, G. Morabitoc, “The Internet of Things: A survey”, Computer Networks, Elsevier, Vol. 54, Issue 15, pp. 2787-2805, October 2010.
6. RAPIRO, The programmable DIY Robot with Endless Possibilities, http://www.rapiro.com/, Last Accessed on August 2016.
7. G. De Marco, M. Ikeda, T. Yang and L. Barolli, “Experimental Performance Evaluation of a Pro-active Ad-hoc Routing Protocol in Outdoor and Indoor Scenarios”, Proc. of International Conference on Advanced Information Networking and Applications (AINA-2007), pp. 7-14, May 2007.
8. L. Barolli, M. Ikeda, G. De Marco, A. Durresi and F. Xhafa, “Performance Analysis of OLSR and BATMAN Protocols Considering Link Quality Parameter”, Proc. of International Conference on Advanced Information Networking and Applications (AINA-2009), pp. 307-314, May 2009.
9. M. Ikeda, L. Barolli, M. Hiyama, T. Yang, G. De Marco and A. Durresi, “Performance Evaluation of a MANET Testbed for Differenet Topologies,” Proc. of International Conference on Network-Based Information Systems (NBiS-2009), Indianapolis, pp. 327-334, August 2009
10. E. Kulla, M. Ikeda, L. Barolli and R. Miho, “Impact of Source and Destination Movement on MANET Performance Considering BATMAN and AODV Protocols“, Proc. of International Conference on Broadband, Wireless Computing, Communication and Applications (BWCCA-2010), pp. 94-101, 2010.
11. Open Mesh Webpage for a Collection of Tools to Build Free and Open Mesh Networks, Available online at http://www.open-mesh.org/
12. J. Bicket, D. Aguayo, S. Biswas, and R. Morris, “Architecture and Evaluation of an Unplanned 802.11b Mesh Network”, Proc. of International Conference on Mobile Computing and Networking (MOBICOM-2005), pp. 31-42 2005.
13. D. A. Maltz, J. Broch, and D. B. Johnson, “Lessons from a Full-scale Multihop Wireless Ad-Hoc Network Testbed”, Journal on Personal Communications, Vol. 8, No. 1, pp. 8-15, February 2001.
14. C. Perkins, E. Belding-Royer, and S. Das, “Ad-hoc On-Demand Distance Vector (AODV) Routing”, RFC 3561 (Experimental), Jul. 2003.
15. D. B. Johnson, D. A. Maltz, and J. Broch, “DSR: The Dynamic Source Routing Protocol for Multi-Hop Wireless Ad-Hoc Networks”, Journal on Ad-Hoc Networking, Chap. 5, pp. 139-172, 2001.
16. S. Hou, M. Wu, W. Liao and D. Wang, “Performance Comparison of AODV and DSR in MANET Test-Bed Based on Internet of Things”, IEEE 82nd Vehicular Technology Conference (VTC Fall), pp. 1-5, Boston, MA, 2015. doi: 10.1109/VTCFall.2015.7391074
17. P. Bellavista, G. Cardone, A. Corradi and L. Foschini, “Convergence of MANET and WSN in IoT Urban Scenarios” IEEE Sensors Journal, vol. 13, no. 10, pp. 3558-3567, October 2013. doi: 10.1109/JSEN.2013.2272099
18. Buffalo Products, WLI-UC-GNM: AirStation N-Technology 150Mbps USB 2.0 Client, http://www.buffalo-technology.com/, Last Accessed on February 2015.

19. Ubuntu 14.04.2 LTS (Trusty Tahr), http://releases.ubuntu.com/14.04/, Last Accessed on February 2015.
20. Raspbian Jessie Operating System for Raspberry Pi, https://www.raspberrypi.org/downloads/raspbian/, Last Accessed on August 2016.
21. A. Dainotti, A. Botta, and A. Pescap, "Do you know what you are generating?" Proc. of International Conference on Emerging Networking Experiments and Technologies (CoNEXT-2007), New York, USA, pp. 1-2, 2007.

Part II
The 18-th International Symposium on Multimedia Network Systems and Applications (MNSA-2016)

Performance Evaluation of an IoT-Based E-Learning Testbed Using Mean Shift Clustering Approach Considering Electroencephalogram Data

Masafumi Yamada, Tetsuya Oda, Yi Liu, Keita Matsuo and Leonard Barolli

Abstract Due to the opportunities provided by the Internet, people are taking advantage of e-learning courses and enormous research efforts have been dedicated to the development of e-learning systems. So far, many e-learning systems are proposed and used practically. However, in these systems the e-learning completion rate is low. One of the reasons is the low study desire and motivation. In this work, we design and implement an IoT-Based E-Learning testbed using Raspberry Pi mounted on Raspbian. We analyze the performance of mean shift clustering algorithm considering electroencephalogram data. For evaluation we considered attention value. The evaluation results show that by the mean shift clustering algorithm the learner concentation is increased.

1 Introduction

The Internet is growing every day and the performance of computers is significantly increased [1]. Also, with appearance of new technologies such as ad-hoc networks, sensor networks, body networks, home networking, new network devices and application are appearing. Therefore, it is very important to monitor and control the network devices via communication channels and exploit their capabilities for the everyday real life activities. However, in large scale networks such as Internet, it is very difficult to control the network devices.

Masafumi Yamada and Yi Liu
Graduate School of Engineering, Fukuoka Institute of Technology (FIT), 3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811–0295, Japan, e-mail: masafumi008355563@gmail.com, ryuui1010@gmail.com

Tetsuya Oda, Keita Matsuo and Leonard Barolli
Department of Information and Communication Engineering, Fukuoka Institute of Technology (FIT), 3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811–0295, Japan e-mail: oda.tetsuya.fit@gmail.com, kt-matsuo@fit.ac.jpandbarolli@fit.ac.jp

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_54

So for many e-learning systems are proposed and used practically. In [2], the authors presents a work-in-progress intending to enhance the learning experience for distance university students enrolled at the Open University of Catalonia (UOC). The UOC virtual campus has an integrated e-learning environment that allows students to pursue their studies completely online with the exception of taking final exams. By integrating the technologies of the IoT, they want to expand the learning environment and add a new learning place to the one existing on the computer. The authors hope to combine both the virtual and the physical environments in order to provide a better learning experience to their students. The authors consider two applications types: one related to the learning process and learning materials, the other related to creating a university community as well as fighting dropout and loneliness.

In [3], the authors present the context-aware and culture-oriented aspects of an adaptability approach called Adapt-SUR. Adapt-SUR is an international joint project between Argentina and Brazil. The approach is designed to be integrated into two distinct E-learning environments (ELEs): the AdaptWeb (Adaptive Web based learning Environment) system [4] and the eTeacher+SAVER (Software de Asistencia Virtual para Educacion Remota) environment [5]. This study describes the main features of the context-aware and culture-oriented aspects of a student profile and shows how to organize this contextual information in a multidimensional space where each dimension is represented by a different ontology, which may be handled separately or jointly. Finally the authors use some examples to discuss and illustrate how to use cultural information to provide context-based e-learning personalization.

In this work, we implement an IoT-based e-learning testbed using Raspberry Pi and investigate the performance of mean shift clustering apploach for electroencephalogram data. For evaluation, we considered the attention value.

The structure of the paper is as follows. In Section 2, we explain the overview of IoT and ULE. In Section 3, we present an overview of mean shift clustering algorithm. In Section 4, we show the description and design of the testbed. In Section 5, we discuss the simulation results. Finally, conclusions and future work are given in Section 6.

2 Overview of IoT and ULE

2.1 Internet of Things (IoT)

The Internet of Things (IoT) is a recent communication paradigm that envisions a near future, in which the objects of everyday life will be equipped with microcontrollers, transceivers for digital communication, and suitable protocol stacks that will make them able to communicate with one another and with the users, becoming an integral part of the Internet [6, 7]. The IoT concept aims at making the Internet even more immersive and pervasive. Furthermore, by enabling easy access and interaction with a wide variety of devices such as, for instance, home appliances, surveillance cameras, monitoring sensors, actuators, displays, vehicles, and so on, the IoT will foster the development of a number of applications that make use of the potentially enormous amount and variety of data generated by such objects to provide new services to citizens, companies, and public administrations. This paradigm indeed finds application in many different domains, such as home automation, industrial automation, medical aids, mobile healthcare, elderly assistance, intelligent energy management and smart grids, automotive, traffic management, and many others [8].

2.2 Ubiquitous Learning Environment (ULE)

Ubiquitous learning is a seamless learning whenever it is in information space or in physics space, through ubiquitous computing information space and physics space are converged. In ULE Learning, learning demands and learning resources are everywhere; study, life and work are connected each other. When learners meet any practice problem ubiquitous computing help them to resolve it at anytime, anywhere. In the future, school, library, classroom, meeting room, museum, and the circulation fields send their information and knowledge to the learner through all kinds of technology, every learner immerse into information ecology surroundings that the real world and digital world intermingle. The learners can easily perception and obtaining learning objects detailed information and content through situational perception of mobile devices. Using dialogue, living community, cooperation studies, social process of internalization, participate in joint activity to realize social learning. An effective ubiquitous learning depends on founding of learning environment.

2.3 Role of IoT in ULE

According to learning environment classification, ubiquitous learning environment belong to a kind of learning environment that are deeper, and the highest flexibility. While the basic elements of constructing the learning environment mainly include three parts: ubiquitous communication network, learning terminal device, learning resources. The traditional single point centralized resource storage mode is unable to meet with the ubiquitous learning requirements whether the resource storage or the promptness of obtaining resources. IoT make not only real world are connected, but also the real world (physical narrow room) and virtual worlds (digital information space) are all interconnected, and it support effectively M2M interaction. IoT make every things of learning environment digital, intelligence and networking, make learning seamless integration, learner study what they need at any time,

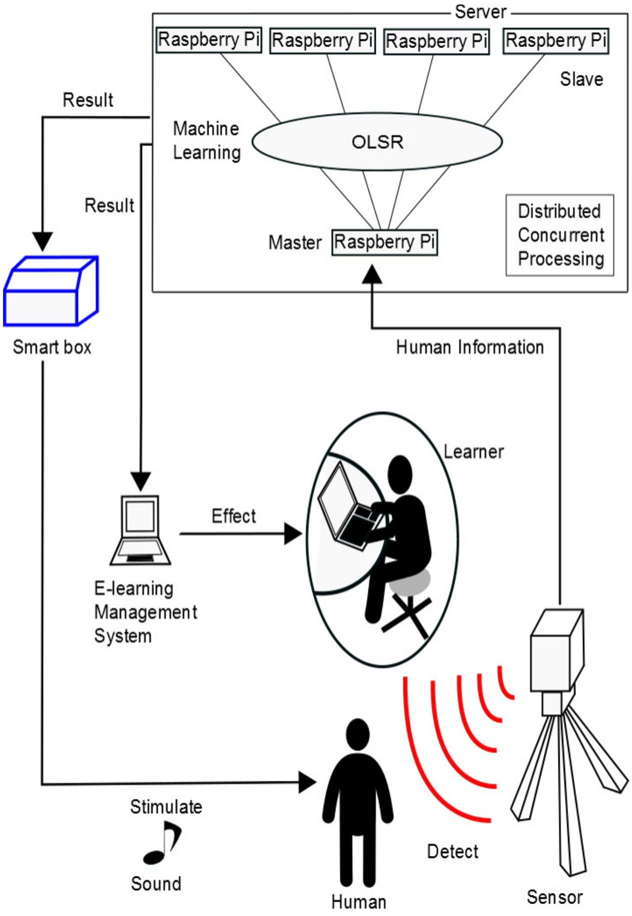


Fig. 1 Structure of IoT-Based E-Learning testbed.

at anyplace, and adjust corresponding learning content, and make learning environment intelligence. For example, monitor and control light brightness by sensor; learn outdoor things by RFID, and so on.

3 The Mean-shift Clustering Algorithm

Mean shift represents a general non-parametric mode finding/clustering procedure [9, 10]. In contrast to the classic K-means clustering approach, there are no embedded assumptions on the shape of the distribution nor the number of modes/clusters. Mean shift was first proposed by Fukunaga and Hostetler, later adapted by Cheng for the purpose of image analysis and more recently extended by Comaniciu, Meer and Ramesh to low level vision problems, including, segmentation adaptive smoothing and tracking.

The main idea behind mean shift is to treat the points in the d-dimensional feature space as an empirical probability density function where dense regions in the feature space correspond to the local maxima or modes of the underlying distribution. For each data point in the feature space, one performs a gradient ascent procedure on the local estimated density until convergence. The stationary points of this procedure represent the modes of the distribution. Furthermore, the data points associated (at least approximately) with the same stationary point are considered members of the same cluster.

Here is briefly described the variable bandwidth mean shift procedure [11, 12]. Given n data points x_i on a d-dimensional space R^d and the associated bandwidths $h_i = h(x_i)$, $i = 1, \ldots, n$, the sample point density estimator obtained with profile $k(x)$ is given by:

$$f(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_i^d} k\Big(\| \frac{x - x_i}{h_i} \|^2\Big). \tag{1}$$

It is utilized multivariate normal profile:

$$k(x) = e^{-\frac{1}{2}x}, x \geq 0. \tag{2}$$

Taking the gradient of Eq. (1), the stationary points of the density function satisfy:

$$\frac{2}{n}\sum_{i=1}^{n} \frac{1}{h_i^{d+2}}(x_i - x)g\Big(\| \frac{x - x_i}{h_i} \|^2\Big) = 0, \tag{3}$$

where $g(x) = -k'(x)$. The solution can be found iteratively via the fixed point algorithm

$$\bar{x} = \frac{\sum_{i=1}^{n} \frac{x_i}{h_i^{d+2}} g\Big(\| \frac{x-x_i}{h_i} \|^2\Big)}{\sum_{i=1}^{n} \frac{1}{h_i^{d+2}} g\Big(\| \frac{x-x_i}{h_i} \|^2\Big)}, \tag{4}$$

which is called mean shift procedure. Comaniciu and Meer [13] show that the convergence to a local mode of the distribution is guaranteed when the mean shift iterations are started at a data point. The mean shift procedure for a given point x_i is as follows.

Algorithm 1 The process of mean shift algorithm.

```
1: Compute the mean shift vector $m(x_i^t)$.
2: Translate density estimation window: $x_i^{t+1} = x_i^t + m(x_i^t)$.
3: Iterate steps 1. and 2. until convergence, i.e., $\nabla f(x_i) = 0$.
```

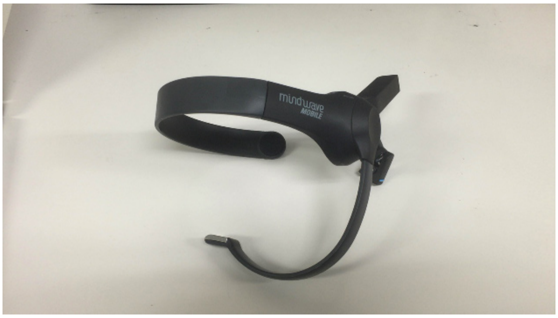


Fig. 2 A snapshot of MindWave Mobile.

Table 1 Descriptions of eSense meter values.

Values	Description
1-20	Strongly lowered levels
20-40	Reduced levels
40-60	Neutral / Baseline levels
60-80	Slightly elevated / higher than normal levels
80-100	Elevated / heightened levels

4 Testbed Description

In Fig. 1 is shown the structure of IoT-based e-learning testbed [14]. Our testbed is composed of five Raspberry Pi B+ [15, 16]. The Raspberry Pi is a credit card-sized single-board computer developed by the Raspberry Pi Foundation. We use MindWave Mobile (MWM) to get the human brain waves.

4.1 MindWave Mobile

A snapshot of the MWM is shown in Fig. 2. MWM is a device capable of acquiring the human brain waves [17]. The device measures the raw signal, power spectrum (alpha, beta, delta, gamma, theta), attention level, mediation level and blink detection. The raw EEG data are received at a rate of 512 Hz. Other measured values are made every second. Therefore, raw EEG data is a main source of information on EEG signals using MWM [18]. By MWM can be determined how effectively the user is engaging Atteention (similar to concentration) by decoding the electrical signals and applying algorithms to provide readings on a scale of 0 to 100. These values are described in Table 1.

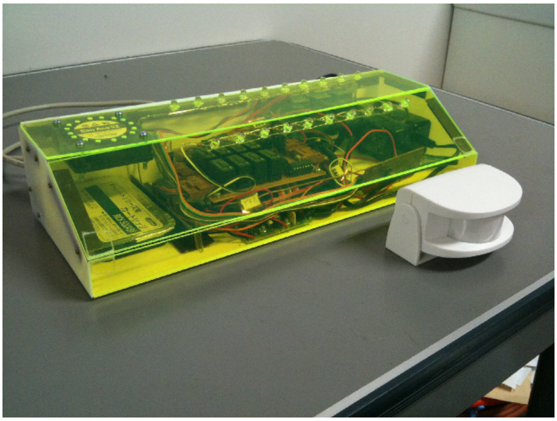


Fig. 3 A snapshot of SmartBox.

4.2 SmartBox Description

We implemented a SmartBox device [1, 19]. The size of the SmartBox is 35×7×12 [*cm*]. The SmartBox is equipped with different sensors (for sensing learner situation) and devices (used for stimulating learner's motivation). The SmartBox has the following sensors and functions.

- Body Sensor: for detecting the learner's body movement.
- Chair Vibrator Control: for vibrating the learner's chair.
- Light Control: for adjusting the room light for study.
- Smell Control: for controlling the room smell.
- Sound Control: to emit relaxing sounds.
- Remote Control Socket: for controlling AC 100 [*V*] socket (on-off control).

A snapshot of the SmartBox is shown in Fig. 3.

5 Simulation Results

The simulation parameters are shown in Table 2. We have collected the attention value data by MWM. These data are collected by using the scikit-learn, which is a general purpose machine learning library for the Python. The scikit-learn provides efficient implementations of state-of-the-art algorithms, accessible to non-machine learning experts, and reusable across scientific disciplines and application fields. The scikit-learn also takes advantage of Python interactivity and modularity to supply fast and easy prototyping.

In Fig. 4, we present the result by mean-shift clustering algorithm. Since the first 100 seconds is the time until the values should be stabilized, we collected the data after 100 seconds. We can see 3 regions of clustering. Based on this data, the system

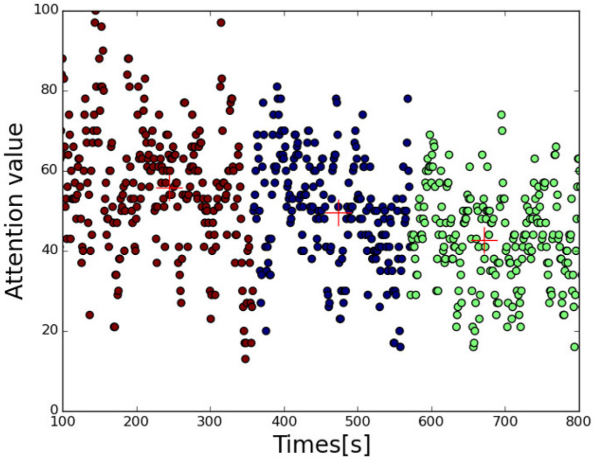


Fig. 4 Mean-Shift clustering algorithm results.

Table 2 Simulation parameters.

Parameters	Values
Number of clusters	3
Initial centroids	designation
Precompute distance	true

is able to judge concentration of learner. The + mark represents the center of gravity of each color. The center of gravity values are: red = 55.7927, blue = 49.3902 and green = 42.5305, respectively. All these values are classified as Neutral / Baseline levels.

6 Conclusions

In this paper, we presented the an IoT-based e-learning testbed and evaluation results of mean-shift clustering algorithm. We clustered sensed data by mean-shift clustering algorithm. From evaluation results, we conclude as following.

- The mean shift clustering algorithm can cluster sensed data.
- Using our testbed, the concentration of learner can be improved.

In the future, we will carry out many experiments using the implemented testbed.

References

1. K. Matsuo, L. Barolli, F. Xhafa, V. Kolici, A. Koyama, A. Durresi, R. Miho, "Implementation of an E-Learning System Using P2P, Web and Sensor Technologies", Proc. of AINA-2009, pp. 800-807, 2009.
2. M. G. Domingo, J. A. M. Forner, "Expanding the Learning Environment: Combining Physicality and Virtuality - The Internet of Things for eLearning", Proc. of IEEE 10th International Conference on Advanced Learning Technologies (ICALT), pp. 730-731, 2010.
3. I. Gasparini, V. Eyharabide, S. Schiaffino, M. S. Pimenta, A. Amandi, J. P. M. de Oliveira, "Improving User Profiling for a Richer Personalization: Modeling Context in E-Learning", Intelligent and Adaptive Learning Systems: Technology Enhanced Support for Learners and Teachers, Chapter 12, pp. 182-197, 2012.
4. V. de Freitas, V. P. Marcal, I. Gasparini, M. A. Amaral, M. L. Proenca Jr., M. A. C. Brunetto, M. S. Pimenta, C. H. F. P. Ribeiro, J. V. de Lima, J. P. M. de Oliveira, "AdaptWeb: an adaptive web-based courseware", Proc. of ICTE-2002, pp. 131-134, 2002.
5. S. Schiaffino, P. Garcia, A. Amandi, "eTeacher: Providing personalized assistance to e-learning students", Computers & Education, Vol. 51, pp. 1744-1754, 2008.
6. A. Zanella, N. Bui, A. Castellani, L. Vangelista, "Internet of Things for Smart Cities", IEEE Internet of Things Journal, Vol. 1, No. 1, pp. 22-32, 2014.
7. L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey", Comput. Netw., Vol. 54, No. 15, pp. 2787-2805, 2010.
8. P. Bellavista, G. Cardone, A. Corradi, and L. Foschini, "Convergence of MANET and WSN in IoT urban scenarios", IEEE Sens. J., Vol. 13, No. 10, pp. 3558-3567, Oct. 2013.
9. R. Obukata, T. Oda, D. Elmazi, L. Barolli, K. Matsuo, I. Woungang, "Performance Evaluation of an Ambient Intelligence Testbed for Improving Quality of Life: Evaluation Using Clustering Approach", The 9-th International Workshop on Intelligent Informatics and Natural Inspired Computing (IINIC-2016) , Fukuoka Institute of Technology, Fukuoka, Japan, July 6-8, 2016.
10. K. G. Derpanis, "Mean shift clustering", See `http://www.cse.yorku.ca/~kosta/CompVis_Notes/mean_shift.pdf` accessed on 14 September 2016.
11. O. Tuzel , F. Porikli , P. Meer, "Kernel methods for weakly supervised mean shift clustering", IEEE 12th International Conference on Computer Vision, pp. 48-55, 2009.
12. D. Comaniciu, "Variable bandwidth density-based fusion", In Proc. IEEE Conf. on Comp. Vis. and Pat. Recog., Madison, WI, Vol. 1, pp. 56-66, 2003.
13. D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis", IEEE Trans. Pat. Anal. Mach. Intell., 24:603-619, 2002.
14. M. Yamada, T. Oda, K. Matsuo, L. Barolli, "Design of an IoT-Based E-Learning Testbed", The 9-th International Symposium on Mining and Web (MAW-2016), pp. 720-724, 2016.
15. "Raspberry Pi Foundation.", `http://www.raspberrypi.org/`.
16. T. Oda, A. Barolli, S. Sakamoto, L. Barolli, M. Ikeda, K. Uchida, "Implementation and Experimental Results of a WMN Testbed in Indoor Environment Considering LoS Scenario", The 29-th IEEE International Conference on Advanced Information Networking and Applications (AINA-2015), pp. 37-42, 2015.
17. "NeuroSky to Release MindWave mobile", `http://mindwavemobile.neurosky.com`.
18. W. Salabun, "Processing and spectral analysis of the raw EEG signal from the MindWave", Przeglad Elektrotechniczny, Vol. 90, No. 2, pp. 169-174, February 2014.
19. K. Matsuo, L. Barolli, J. Arnedo-Moreno, F. Xhafa, A. Koyama, A. Durresi, "Experimental Results and Evaluation of SmartBox Stimulation Device in a P2P E-learning System", Proc. of NBiS-2009, pp. 37-44, 2009.

A Testbed for Admission Control in WLAN: A Fuzzy Approach and Its Performance Evaluation

Takaaki Inaba, Shinji Sakamoto, Tetsuya Oda, Makoto Ikeda, Leonard Barolli

Abstract With the popularization of mobile devices such as smart phones, many device communicate over Wireless Local Area Networks (WLANs). The IEEE 802.11e standard is an important extension of the IEEE 802.11 standard focusing on QoS that works with any PHY implementation. The IEEE 802.11e standard introduces EDCF and HCCA. Both these schemes are useful for QoS provisioning to support delay-sensitive voice and video applications. EDCF uses the contention window to differentiate between high priority and low priority services. However, it does not consider the priority of users. In this paper, in order to deal with this problem, we propose and implement a testbed for Admission Control in WLANs based on Fuzzy Logic. We evaluate by experiment the performance of implemented testbed. The experimental results show that the time for connection swap becomes longer, if the user priority is higher. Also, the user request succsess ratio and connected time ratio is increased when the user priority is higher.

1 Introduction

With the development of wireless technology and Internet, there is an increasing need towards portable and mobile computers such as smart phones. The wireless networks need to provide communications between mobile terminals. The Wire-

Takaaki Inaba, Shinji Sakamoto
Graduate School of Engineering,
Fukuoka Institute of Technology (FIT),
3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811-0295, Japan,
e-mail: g.takaaki.inaba@gmail.com, shinji.t.sakamoto@gmail.com

Tetsuya Oda, Makoto Ikeda, Leonard Barolli
Department of Information and Communication Engineering,
Fukuoka Institute of Technology (FIT)
3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811-0295, Japan,
e-mail: oda.tetsuya.fit@gmail.com, makoto.ikd@acm.org, barolli@fit.ac.jp

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_55

less Local Area Networks (WLANs) provide high bandwidth access for users in a limited geographical area. With the popularization of mobile devices, many device communicate together over WLANs [24]. WLANs have a lot of restriction on communication resources comparing wired LANs. Therefore, it is more difficult to guarantee the Quality of Service (QoS).

The IEEE 802.11 standard [24] defines the Distributed Coordination Function (DCF) which provides best-effort service at the Medium Access Control (MAC) layer of the WLANs. The IEEE 802.11e standard [29] specifies the Hybrid Coordination Function (HCF) which enables prioritized Quality-of-Service (QoS) services at the MAC layer, on top of DCF. The HCF combines a distributed contention-based channel access mechanism called Enhanced Distributed Channel Access (EDCA), and a centralized polling-based channel access mechanism called HCF Controlled Channel Access (HCCA) [29].

The EDCA scheme uses Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) and slotted Binary Exponential Backoff (BEB) mechanism as the basic access method. The EDCA defines multiple Access Categories (AC) with AC-specific Contention Window (CW) sizes, Arbitration Interframe Space (AIFS) values, and Transmit Opportunity (TXOP) limits to support MAC-level QoS and prioritization [29].

In this paper, we present the implementation of a testbed for admission control in WLANs. The admission decision is done by a Fuzzy-based Admission Control System (FACS) [9].

The paper is organized as follows. In Section 2, we introduce IEEE 802.11, EDCA and HCCA. In Section 3, we give some related works. In Section 4, we discuss the application of Fuzzy Logic (FL) for control. In Section 5, we present the implemented testbed. In Section 6, we show the experimental results. Finally, in Section 7, we conclude the paper.

2 IEEE 802.11

IEEE802.11 uses CSMA/CA as shown is Fig. 1. The nodes check whether others device are communicating or not before starting communication [7]. If other nodes are communicating with the AP, the node waits for a period of time, called Distributed Inter Frame Space (DIFS). After that, it waits an additionally random time called back-off time. After the back-off time, if other devices are not communicating, the node starts to send data [23]. AP which received the data waits a constant time, called Short Inter Frame Space (SIFS) and send ACK to the node which sent the data. Because ACK frame should be sent soon, it is shorter than DIFS.

The IEEE 802.11e standard is an important extension of the IEEE 802.11 standard focusing on QoS [16] that works with any PHY implementation. Wireless nodes equipped with IEEE 802.11e features are now known as QoS stations (QSTAs) and they are associated with a QoS access point (QAP) to form a QoS basic service set (QBSS). The main feature of the IEEE 802.11e standard is that it im-

Fig. 1 CSMA/CA.

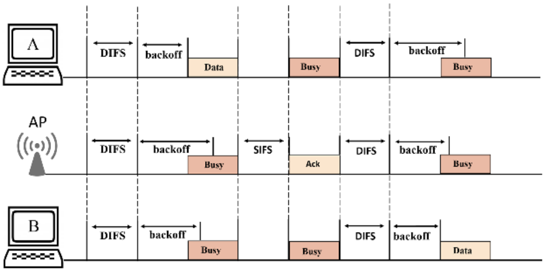
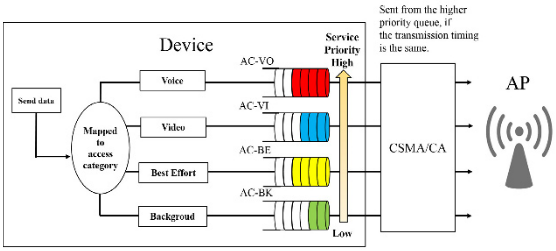


Fig. 2 EDCA.


Sent from the higher priority queue, if the transmission timing is the same.

proves the MAC layer for QoS provisioning by providing support for: segregation of data packets based on priority requirements; negotiation of QoS parameters through a central coordinator or AP; and admission control.

The IEEE 802.11e standard introduces EDCF and HCCA. Both these schemes are useful for QoS provisioning to support delay-sensitive voice and video applications [6].

2.1 EDCF

In the DCF configuration, a contention window is set after a frame is transmitted. This is done to avoid any collisions. The window defines the contention time of various stations who contend with each other for access to channel. However, each of the stations cannot seize the channel immediately, rather the MAC protocol uses a randomly chosen time period for each station after that channel has undergone transmission [27]. As shown in Fig. 2, the EDCF uses this contention window to differentiate between high priority and low priority services [22]. The central coordinator assigns a contention window of shorter length to the stations with higher priority that helps them to transmit before the lower priority ones [13, 21]. To differentiate further, Inter Frame Spacing (IFS) can be varied according to different traffic categories. Instead of using a DIFS as for the DCF traffic, a new inter-frame spacing called Arbitration Inter Frame Spacing (AIFS) is used. The AIFS used for traffic has a duration of a few time slots longer than the DIFS duration. Therefore, a traffic category having smaller AIFS gets higher priority.

2.2 HCCA

The HCF controlled channel access (HCCA) is IEEE802.11e specific, and it makes use of a Hybrid Coordinator (HC) to manage the bandwidth allocation of wireless medium [11]. The HC can obtain a transmission opportunity (TXOP) and initiate data deliveries to provide transmission opportunities to a station with a higher priority without any backoff; that is to say, the HC can access the channels after a PIFS amount of time rather than a DIFS amount of time as for the other stations [16]. As PIFS is smaller than DIFS and AIFS, the HC has a priority over the DCF traffic, and also over the ECF traffic that uses AIFS.

IEEE802.11e provides EDCA and Hybrid coordination function controlled channel access (HCCA) as a priority control method [6, 15, 29]. Mainly, EDCA is used because of easy implementation and compatibility of CSMA/CA.

3 Related Work

The enhancements at the MAC layer provides service differentiation among different traffic flows, but it can ensure QoS only when network load is reasonable. If the load increases beyond a certain limit, the QoS guarantees are not ensured even to high priority traffic [15, 26]. This is where the admission control mechanism helps in preventing the network from becoming congested, by allowing or disallowing flows depending on whether the conditions are favorable to meet QoS requirements. More specifically, the purpose of admission control is to limit the amount of newly admitted traffic such that the QoS performance of existing flows is not degraded [7]. Admission control is a key component to adapt to the traffic variations according to the changing environment of IEEE 802.11-based wireless networks [1, 8]. Admission control can be categorized into three different methodologies [3].

3.1 Measurement-based Admission Control

In this scheme, the decisions are made through continuous monitoring of network status, such as throughput and delay. A certain threshold is maintained according to the network status for admission of new traffic flows. Nor et al. in [19] proposed a metric called network utilization characteristic (NUC) as a means for admission of traffic flows into network. NUC defines the amount of channel utilized to transmit the flow over the network. This scheme guarantees QoS to high priority flows under loaded channel environments. Another scheme presented by Wu et al. in [25] is that each traffic class is assigned a certain portion of available resources, and these resources are then remaining reserved for that particular class. In this regard, only the traffic with higher priority compared to the existing traffic is admitted.

3.2 Model-based Admission Control

In model-based schemes, the network status is measured based on some models. The Markov chain models are quite popular in attempts at modeling IEEE 802.11 although other approaches are also being explored due to some limitations of Markovian models [5]. In [4], an analytical model is used to estimate the minimum bandwidth requirement of all flows. When a newly admitted flow need to be activated, the algorithm checks if it is going to result in preservation of QoS requirements of existing flows.

3.3 Measurement-aided, Model-based Admission Control

It is a hybrid of measurement-based and model-based schemes. The algorithm in [14] takes network measurements in a loaded environment and also the data rate requirements of the flow that is requesting for admission. Furthermore, a channel model is applied to predict the network conditions and provides QoS enhancements accordingly. Another solution is the threshold-based approach proposed in [2] in which the channel conditions are continuously monitored and the contention probability is measured. When any new flows request for admission, the admission control checks for the competing flows. The absolute bandwidth and the expected delay of the new flow are measured. If this satisfies the threshold conditions, then this flow is admitted.

4 Application of Fuzzy Logic for Control

The ability of fuzzy sets and possibility theory to model gradual properties or soft constraints whose satisfaction is matter of degree, as well as information pervaded with imprecision and uncertainty, makes them useful in a great variety of applications.

The most popular area of application is Fuzzy Control (FC), since the appearance, especially in Japan, of industrial applications in domestic appliances, process control, and automotive systems, among many other fields.

4.1 FC

In the FC systems, expert knowledge is encoded in the form of fuzzy rules, which describe recommended actions for different classes of situations represented by fuzzy sets.

In fact, any kind of control law can be modeled by the FC methodology, provided that this law is expressible in terms of “if ... then ...” rules, just like in the case of expert systems. However, FL diverges from the standard expert system approach by providing an interpolation mechanism from several rules. In the contents of complex processes, it may turn out to be more practical to get knowledge from an expert operator than to calculate an optimal control, due to modeling costs or because a model is out of reach.

4.2 Linguistic Variables

A concept that plays a central role in the application of FL is that of a linguistic variable. The linguistic variables may be viewed as a form of data compression. One linguistic variable may represent many numerical variables. It is suggestive to refer to this form of data compression as granulation [10].

The same effect can be achieved by conventional quantization, but in the case of quantization, the values are intervals, whereas in the case of granulation the values are overlapping fuzzy sets. The advantages of granulation over quantization are as follows:

- it is more general;
- it mimics the way in which humans interpret linguistic values;
- the transition from one linguistic value to a contiguous linguistic value is gradual rather than abrupt, resulting in continuity and robustness.

4.3 FC Rules

FC describes the algorithm for process control as a fuzzy relation between information about the conditions of the process to be controlled, x and y, and the output for the process z. The control algorithm is given in “if ... then ...” expression, such as:

If x is small and y is big, then z is medium;
If x is big and y is medium, then z is big.

These rules are called *FC rules*. The “if” clause of the rules is called the antecedent and the “then” clause is called consequent. In general, variables x and y are called the input and z the output. The “small” and “big” are fuzzy values for x and y, and they are expressed by fuzzy sets.

Fuzzy controllers are constructed of groups of these FC rules, and when an actual input is given, the output is calculated by means of fuzzy inference.

4.4 Control Knowledge Base

There are two main tasks in designing the control knowledge base. First, a set of linguistic variables must be selected which describe the values of the main control parameters of the process. Both the input and output parameters must be linguistically defined in this stage using proper term sets. The selection of the level of granularity of a term set for an input variable or an output variable plays an important role in the smoothness of control. Second, a control knowledge base must be developed which uses the above linguistic description of the input and output parameters. Four methods [17, 20, 28, 30] have been suggested for doing this:

- expert's experience and knowledge;
- modelling the operator's control action;
- modelling a process;
- self organization.

Among the above methods, the first one is the most widely used. In the modeling of the human expert operator's knowledge, fuzzy rules of the form "If Error is *small* and Change-in-error is *small* then the Force is *small*" have been used in several studies [12, 18]. This method is effective when expert human operators can express the heuristics or the knowledge that they use in controlling a process in terms of rules of the above form.

4.5 Defuzzification Methods

The defuzzification operation produces a non-FC action that best represent the membership function of an inferred FC action. Several defuzzification methods have been suggested in literature. Among them, four methods which have been applied most often are:

- Tsukamoto's Defuzzification Method;
- The Center of Area (COA) Method;
- The Mean of Maximum (MOM) Method;
- Defuzzification when Output of Rules are Function of Their Inputs.

5 Implemented Testbed

We show the testbed structure in Fig. 3. The testbed is constructed by one AP and 11 users. The admission decision is done by FACS [9]. Fuzzy Logic Controller (FLC) is a main part of FACS and its components are shown in Fig. 4. The Fuzzy Rule Base (FRB) and the membership functions are shown in Table 1 and in Fig. 5, respectively. In FACS, we consider 3 input parameters: User Priority (UP), Received

Fig. 3 Proposed system.

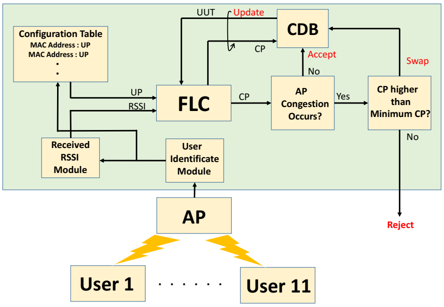
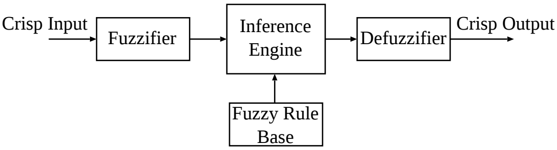
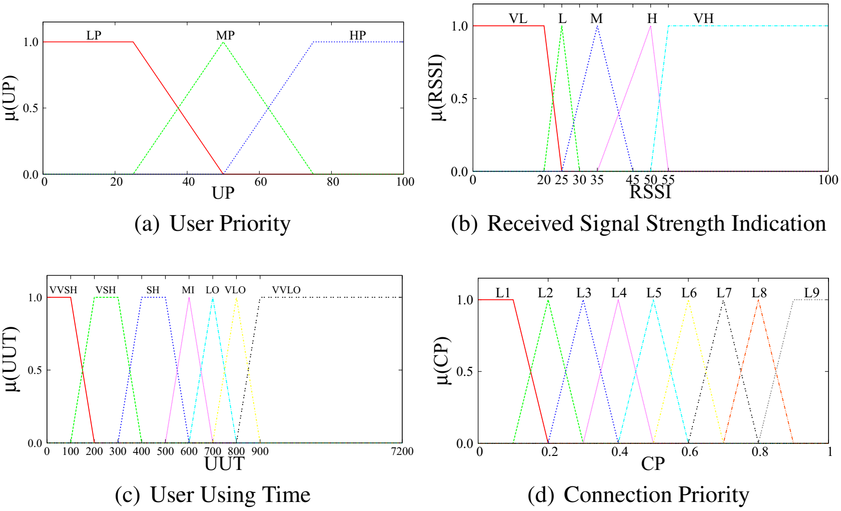


Fig. 4 Fuzzy logic controller.





(a) User Priority

(b) Received Signal Strength Indication

(c) User Using Time

(d) Connection Priority

Fig. 5 Membership functions of FACS.

Signal Strength Indication (RSSI) and User Using Time (UUT). The output is Connection Priority (CP). The term sets of *UP*, *RSSI* and *UUT* are defined respectively as:

$$UP = \begin{pmatrix} Low\ priority \\ Middle\ priority \\ High\ priority \end{pmatrix} = \begin{pmatrix} Lp \\ Mp \\ Hp \end{pmatrix};$$

Table 1 Fuzzy Rule Base

Rule	UP	RSSI	UUT	CP	Rule	UP	RSSI	UUT	CP	Rule	UP	RSSI	UUT	CP
1	LP	VL	VVSH	L4	36	MP	VL	VVSH	L5	71	HP	VL	VVSH	L5
2	LP	VL	VSH	L3	37	MP	VL	VSH	L4	72	HP	VL	VSH	L4
3	LP	VL	SH	L2	38	MP	VL	SH	L3	73	HP	VL	SH	L3
4	LP	VL	MI	L1	39	MP	VL	MI	L3	74	HP	VL	MI	L3
5	LP	VL	LO	L1	40	MP	VL	LO	L2	75	HP	VL	LO	L2
6	LP	VL	VLO	L1	41	MP	VL	VLO	L1	76	HP	VL	VLO	L4
7	LP	VL	VVLO	L1	42	MP	VL	VVLO	L1	77	HP	VL	VVLO	L4
8	LP	L	VVSH	L4	43	MP	L	VVSH	L6	78	HP	L	VVSH	L8
9	LP	L	VSH	L3	44	MP	L	VSH	L5	79	HP	L	VSH	L7
10	LP	L	SH	L2	45	MP	L	SH	L4	80	HP	L	SH	L7
11	LP	L	MI	L2	46	MP	L	MI	L3	81	HP	L	MI	L6
12	LP	L	LO	L1	47	MP	L	LO	L3	82	HP	L	LO	L6
13	LP	L	VLO	L1	48	MP	L	VLO	L2	83	HP	L	VLO	L5
14	LP	L	VVLO	L1	49	MP	L	VVLO	L1	84	HP	L	VVLO	L4
15	LP	M	VVSH	L5	50	MP	M	VVSH	L7	85	HP	M	VVSH	L8
16	LP	M	VSH	L4	51	MP	M	VSH	L6	86	HP	M	VSH	L8
17	LP	M	SH	L3	52	MP	M	SH	L5	87	HP	M	SH	L7
18	LP	M	MI	L3	53	MP	M	MI	L4	88	HP	M	MI	L7
19	LP	M	LO	L2	54	MP	M	LO	L4	89	HP	M	LO	L7
20	LP	M	VLO	L1	55	MP	M	VLO	L3	90	HP	M	VLO	L6
21	LP	M	VVLO	L1	56	MP	M	VVLO	L2	91	HP	M	VVLO	L5
22	LP	H	VVSH	L7	57	MP	H	VVSH	L8	92	HP	H	VVSH	L9
23	LP	H	VSH	L6	58	MP	H	VSH	L7	93	HP	H	VSH	L9
24	LP	H	SH	L5	59	MP	H	SH	L6	94	HP	H	SH	L8
25	LP	H	MI	L4	60	MP	H	MI	L6	95	HP	H	MI	L8
26	LP	H	LO	L4	61	MP	H	LO	L5	96	HP	H	LO	L8
27	LP	H	VLO	L3	62	MP	H	VLO	L4	97	HP	H	VLO	L7
28	LP	H	VVLO	L2	63	MP	H	VVLO	L4	98	HP	H	VVLO	L7
29	LP	VH	VVSH	L8	64	MP	VH	VVSH	L9	99	HP	VH	VVSH	L9
30	LP	VH	VSH	L7	65	MP	VH	VSH	L8	100	HP	VH	VSH	L9
31	LP	VH	SH	L7	66	MP	VH	SH	L8	101	HP	VH	SH	L9
32	LP	VH	MI	L6	67	MP	VH	MI	L7	102	HP	VH	MI	L9
33	LP	VH	LO	L6	68	MP	VH	LO	L7	103	HP	VH	LO	L9
34	LP	VH	VLO	L5	69	MP	VH	VLO	L6	104	HP	VH	VLO	L8
35	LP	VH	VVLO	L4	70	MP	VH	VVLO	L6	105	HP	VH	VVLO	L8

$$RSSI = \begin{pmatrix} Very\ Low \\ Low \\ Middle \\ High \\ Very\ High \end{pmatrix} = \begin{pmatrix} VL \\ L \\ M \\ H \\ VH \end{pmatrix};$$

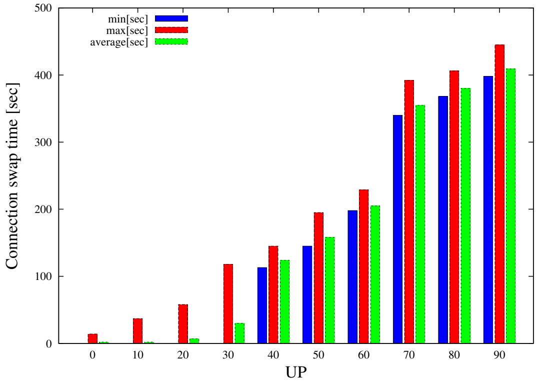


Fig. 6 Experimental results for connection swap time vs. user priority.

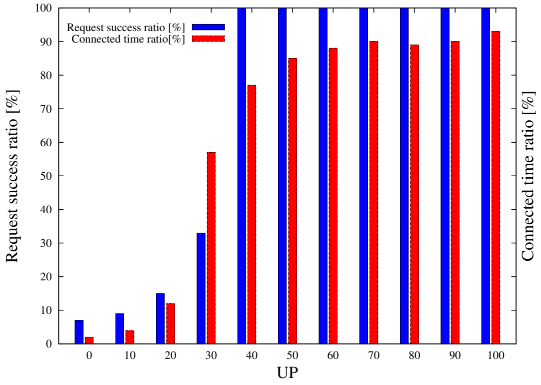


Fig. 7 Experimental results for request success ratio and connected time ratio vs. user priority.

$$UUT = \begin{pmatrix} Very\ Very\ Short \\ Very\ Short \\ Short \\ Middle \\ Long \\ Very\ Long \\ Very\ Very\ Long \end{pmatrix} = \begin{pmatrix} VVSH \\ VSH \\ SH \\ MI \\ LO \\ VLO \\ VVLO \end{pmatrix}.$$

The term set for the output CP is defined as:

$$CP = \begin{pmatrix} Level\ 1 \\ Level\ 2 \\ Level\ 3 \\ Level\ 4 \\ Level\ 5 \\ Level\ 6 \\ Level\ 7 \\ Level\ 8 \\ Level\ 9 \end{pmatrix} = \begin{pmatrix} L1 \\ L2 \\ L3 \\ L4 \\ L5 \\ L6 \\ L7 \\ L8 \\ L9 \end{pmatrix}.$$

Table 2 Experimental Setting and Environment.

Parameters	Values
Total number of user devices	11
Maximum number of connected users	7
Size of download file [MB]	1, 5, 10, 20, 50, 100, 200, 300, 500, 1000
Clients period for reconnecting to AP	5 seconds

Our testbed works as follows (see Fig. 3). An user sends a connection request to AP. At this moment, the UUT for this user is zero. Then, input UUT, RSSI and UP to FLC in order to get CP. The user will be accepted if congestion does not occur in AP. When the AP is congested, the output CP value of FLC is compared with the minimum value of CP in Connection DataBase (CDB). A request will be rejected if the user's CP is lower than the minimum value of CP in CDB. The CDB updates frequently each UUT of connected users.

6 Experimental Results

We show the parameter settings for the experiment in Table 2. We prepared 11 user devices and 1 AP. The AP can accept 7 users, simultaneously. Every 5 seconds, each user device tries to connect to AP. We conducted the experiment for 2 hours. The user device download a file when it is connected to AP. The download file size is decided randomly from 1MB to 1000MB.

The experimental results are shown in Fig. 6 and Fig. 7. In Fig. 6, we see that the time for connection swap becomes longer, if the user priority is higher. Fig. 7 shows that the user request succsess ratio and connected time ratio is increased when the user priority is higher.

7 Conclusions

In this paper, we proposed and implemented a testbed for Admission Control in WLANs based on Fuzzy Logic. We evaluated by experiment the implemented performance of the testbed. The experimental results show that the time for connection swap becomes longer, if the user priority is higher. Also, the user request succsess ratio and connected time ratio is increased when the user priority is higher.

In our future work, we would like to evaluate the performance by considering not only user priority but also RSSI. Moreover, we would like to compare the implemented testbed with other systems.

References

[1] Andreadis A, Zambon R (2012) Techniques for Preserving QoS Performance in Contention-based IEEE802.11e Networks. INTECH Open Access Publisher

[2] Bensaou B, Kong ZN, Tsang DH (2009) A Measurement-assisted, Model-based Admission Control Algorithm for IEEE802.11e. Journal of Interconnection Networks 10(04):303–320

[3] Brewer OT, Ayyagari A (2010) Comparison and Analysis of Measurement and Parameter based Admission Control Methods for Quality of Service (QoS) Provisioning. Military Communications Conference (MilCom-2010) pp 184–188

[4] Cano C, Bellalta B, Oliver M (2007) Adaptive Admission Control Mechanism for IEEE802.11e WLANs. The 18th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC-2007) pp 1–5

[5] Chen X, Zhai H, Tian X, Fang Y (2006) Supporting QoS in IEEE802.11e Wireless LANs. Transactions on Wireless Communications 5(8):2217–2227

[6] Choi S, Del Prado J, Mangold S, et al (2003) IEEE 802.11e Contention-based Channel Access (EDCF) Performance Evaluation. International Conference on Communications (ICC-2003) 2:1151–1156

[7] Gao D, Cai J, Ngan KN (2005) Admission Control in IEEE802.11e Wireless LANs. IEEE Network 19(4):6–13

[8] Hanzo I, Tafazolli R, et al (2009) Admission Control Schemes for 802.11-based Multi-hop Mobile Ad hoc Networks: A Survey. Communications Surveys & Tutorials, IEEE 11(4):78–108

[9] Inaba T, Sakamoto S, Oda T, Ikeda M, Barolli L (2016) A QoS-Aware Admission Control System for WLAN Using Fuzzy Logic. The 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA-2016) pp 499–505

[10] Kandel A (1991) Fuzzy Expert Systems. CRC press

[11] Khan M, Khan TA, Beg M (2013) Optimization of Wireless Network MAC Layer Parameters. International Journal of Innovative Technology and Exploring Engineering (IJITEE)

[12] Klir GJ, Folger TA (1988) Fuzzy Sets, Uncertainty, and Information. Prentice Hall

[13] Krithika P, Pushpavalli M (2012) Quality of Service Optimization in IEEE802.11e Networks Using Enhanced Distributed Channel Access Techniques. International Journal of Computer Networks and Wireless Communications (IJCNWC)

[14] Ksentini A, Nafaa A, Gueroui A, Naimi M (2007) ETXOP: A Resource Allocation Protocol for QoS-sensitive Services Provisioning in 802.11 Networks. Performance Evaluation 64(5):419–443

[15] Mangold S, Choi S, May P, Klein O, Hiertz G, Stibor L (2002) IEEE802.11e Wireless LAN for Quality of Service. Proc European Wireless 2:32–39

[16] Mangold S, Choi S, Hiertz GR, Klein O, Walke B (2003) Analysis of IEEE 802.11e for QoS Support in Wireless LANs. Wireless Communications, IEEE 10(6):40–50

[17] McNeill FM, Thro E (1994) Fuzzy Logic: A Practical Approach. Academic Press

[18] Munakata T, Jani Y (1994) Fuzzy Systems: An Overview. Communications of the ACM 37(3):68–76

[19] Nor S, Mohd A, Cheow C (2006) An Admission Control Method for IEEE802.11e. Network Theory and Applications pp 105–122

[20] Procyk TJ, Mamdani EH (1979) A Linguistic Self-organizing Process Controller. Automatica 15(1):15–30

[21] Qashi R, Bogdan M, Hänssgen K (2011) Evaluating the QoS of WLANs for the IEEE802.11 EDCF in Real-time Applications. International Conference on Communications and Information Technology (ICCIT-2011) pp 32–35

[22] Romdhani L, Ni Q, Turletti T (2003) Adaptive EDCF: Enhanced Service Differentiation for IEEE802.11 Wireless Ad-hoc Networks. Wireless Communications and Networking (WCNC-2003) 2:1373–1378

[23] Song NO, Kwak BJ, Song J, Miller LE (2003) Enhancement of IEEE802.11 Distributed Coordination Function with Exponential Increase Exponential Decrease Backoff Algorithm. The 57th IEEE Semiannual Vehicular Technology Conference 4:2775–2778

[24] Wu H, Peng Y, Long K, Cheng S, Ma J (2002) Performance of Reliable Transport Protocol over IEEE802.11 Wireless LAN: Analysis and Enhancement. The 21st Annual Joint Conference of the IEEE Computer and Communications Societies 2:599–607

[25] Wu HT, Yang MH, Ke KW (2010) The Design of QoS Provisioning Mechanisms for Wireless Networks. IEEE International Conference on Communications Workshops (PERCOM Workshops 2010) pp 756–759

[26] Xiao Y, Li H (2004) Local Data Control and Admission Control for QoS Support in Wireless Ad hoc Networks. IEEE Transactions on Vehicular Technology 53(5):1558–1572

[27] Yang X, Vaidya NH (2002) Priority Scheduling in Wireless Ad hoc Networks. Proceedings of the 3rd ACM international symposium on Mobile ad hoc networking & computing pp 71–79

[28] Zadeh LA, Kacprzyk J (1992) Fuzzy Logic for the Management of Uncertainty. John Wiley & Sons, Inc

[29] Zhu J, Fapojuwo AO (2007) A New Call Admission Control Method for Providing Desired Throughput and Delay Performance in IEEE802.11e Wireless LANs. IEEE Transactions on Wireless Communications 6(2):701–709

[30] Zimmermann HJ (1991) Fuzzy Set Theory and Its Applications. Springer Science & Business Media

Simple Energy-efficient Server Selection Algorithm in a Scalable Cluster

Hiroki Kataoka, Atsuhiro Sawada, Dilawaer Duolikun, Tomoya Enokido, and Makoto Takizawa

Abstract It is critical to reduce the electric energy consumed in server clusters. In our previous studies, the MLPCM and MLCM models are proposed with LEA and MEA server selection algorithms. Here, a server is selected to perform a process by estimating the termination time of every current process. However, it takes time to collect the state of each process and estimate the termination time of each process. In this paper, we propose a simple energy-aware (PEA) algorithm to select a server where only state information on number of processes on each server is used. In the evaluation, we show the computation complexity of the PEA algorithm is $O(1)$, smaller than the other algorithms while the total electric energy consumption of the servers of the PEA algorithm is almost the same as the MEA algorithm and is smaller than the others.

1 Introduction

In information systems, it is now critical to reduce electric energy consumed by servers [12], [24]. There are many hardware-oriented approaches to developing energy-efficient hardware devices like CPU [14]. On the other hand, in our macro-

Hiroki Kataoka
Hosei University, Tokyo, Japan, e-mail: hiroki.kataoka.6v@stu.hosei.ac.jp

Atsuhiro Sawada
Hosei University, Tokyo, Japan, e-mail: atsuhiro.sawada.7n@stu.hosei.ac.jp

Dilawaer Duolikun
Hosei University, Tokyo, Japan, e-mail: dilewerdolkun@gmail.com

Tomoya Enokido
Rissho University, Tokyo, Japan, e-mail: eno@ris.ac.jp

Makoto Takizawa
Hosei University, Tokyo, Japan, e-mail: makoto.takizawa@computer.org

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_56

level approach [5], [7], [8] to reducing the electric energy consumption of servers, we consider the total electric power [W] consumed by a whole server to perform application processes. In this paper, a term *process* means an application process to be performed on a server.

Types of electric power consumption models are proposed to perform computation [5], [8], communication [7], storage [13], and general [7], [9] [26] types of application processes. In this paper, we consider computation processes which mainly use CPU [5], [7], [8]. The multi-level power consumption (MLPC) and computation (MLC) models of a server with a multi-thread CPU [14] are proposed [16], [17], [18]. Here, the electric power consumption of a server depends on the number of active CPUs, cores, and threads. The MLPCM (MLPC with multiple CPUs) model is also proposed [20], [21].

Based on the MLPCM model, the locally energy-aware (LEA) [17], globally energy-aware (GEA) [18], scalably energy-aware (SEA) [19], and modified GEA (MEA) [20] [21] algorithms are proposed to select a server for a request process in a cluster. In the algorithms, we have to estimate the termination time of each process on each server and electric energy to be consumed by each server to perform all the current processes. Here, the state information like laxity of each process on each server has to be collected. Furthermore, it takes time to estimate the termination time of each current process and electric energy consumption of each server. In this paper, we newly propose a *simple energy-aware* (*PEA*) algorithm to select an energy-efficient server in a cluster by getting only the number of current processes on each server.

We evaluate the PEA algorithm compared with other algorithms in terms of the total electric energy and total active time of servers and the average execution time of processes. In our previous studies, the performance and energy consumption parameters of each server are randomly decided. In the evaluation, the parameters of real servers are used. We show the total electric energy consumption of a cluster of the PEA algorithm is almost the same as the MEA algorithm and smaller than the other algorithms while the computation complexity of the PEA algorithm is $O(1)$, smaller than the other algorithm.

In section 2, we briefly present the MLPCM and MLCM models of a server. In section 3, we discuss the PEA algorithm to select a server for each process in a cluster of servers. In section 4, we evaluate the PEA algorithm.

2 Power Consumption and Computation Models

2.1 MLPCM Model

In this paper, a process means a computation type of application process which uses CPU resource. A server s_t is composed of np_t (≥ 1) homogeneous CPUs $cp_{t0}, \ldots,$ cp_{t,np_t-1}. Each CPU cp_{tk} [14] is composed of nc_t (≥ 1) homogeneous cores $c_{tk0}, \ldots,$

c_{tk,nc_t-1} and supports processes with nt_t threads $tr_{tk0}, \ldots, tr_{tk,nt_t-1}$. Each core c_{tkh} supports the same number ct_t of threads. A server s_t supports totally nt_t (= $np_t \cdot nc_t \cdot ct_t$) threads. An *active* thread is a thread where at least one process is performed. An active server is a server where at least one thread is active. Let $CP_t(\tau)$ be a set of processes performed on a server s_t at time τ. The electric power consumption $E_t(\tau)$ [W] of a server s_t to perform processes at time τ is given as follows [17]:

[Multi-level power consumption with multiple CPUs (MLPCM) model]

$$E_t(\tau) = minE_t + \sum_{k=0}^{np_t-1}\{\gamma_{tk}(\tau)\ [bE_t + \sum_{i=0}^{nc_t-1} \alpha_{tki}(\tau)\cdot(cE_t + \beta_{tki}(\tau)\ tE_t)]\}. \quad (1)$$

Here, $\gamma_{tk}(\tau)$ = 1 if at least one core is active on a CPU cp_{tk} at time τ. Otherwise, $\gamma_{tk}(\tau)$ = 0. $\alpha_{tki}(\tau)$ = 1 if a core c_{tki} is active on a CPU cp_{tk} at time τ. Otherwise, $\alpha_{tki}(\tau)$ = 0. $\beta_{tki}(\tau)$ ($\leq ct_t$) is the number of active threads on a core c_{tki}. If $\alpha_{tki}(\tau)$ = 0, $\beta_{tki}(\tau)$ = 0. If $\alpha_{tki}(\tau)$ = 1, $1 \leq \beta_{tki}(\tau) \leq ct_t$. As shown in formula (1), the electric power consumption $E_t(\tau)$ [W] depends on the numbers of active CPUs, cores, and threads.

In Linux operating systems [15], processes are allocated to threads in the round-robin (RR) algorithm. Here, the electric power consumption $CE_t(n)$ [W] consumed by a server s_t to concurrently perform n processes at time τ is given as follows [19], [20]:

[MLPCM model for *n* processes]

$$CE_t(n) = \begin{cases} minE_t\ if\ n = 0. \\ minE_t + n\cdot(bE_t + cE_t + tE_t)\ if\ 1 \leq n \leq np_t. \\ minE_t + np_t\cdot bE_t + n\cdot(cE_t + tE_t)\ if\ np_t < n \leq nc_t\cdot np_t. \\ minE_t + np_t\cdot(bE_t + nc_t\cdot cE_t) + nt_t\cdot tE_t\ if\ nc_t\cdot np_t < n < nt_t. \\ maxE_t\ if\ n \geq nt_t. \end{cases} \quad (2)$$

The electric power consumption $E_t(\tau)$ [W] of a server s_t at time τ is assumed to be $CE_t(|\ CP_t(\tau)\ |)$ in this paper. The total electric energy $TE_t(st,et)$ [J] consumed by a server s_t from time st to time et is $TE_t(st,et) = \sum_{\tau=st}^{et} E_t(\tau)$.

2.2 MLCM Model

A cluster S is composed of servers $s_1, \ldots, s_m$ ($m \geq 1$). Processes are performed on the servers. It takes T_{ti} [sec] to perform a process p_i on a thread of a server s_t. If only a process p_i is performed on a server s_t without any other process, the execution time T_{ti} of the process p_i is minimum, i.e. T_{ti} = $minT_{ti}$. $minT_i$ shows a minimum one $minT_{1i}, \ldots, minT_{mi}$, i.e. $minT_i$ = $minT_{fi}$ on the fastest thread which is on a server s_f. Here, the server s_f is referred to as $fastest$ in a cluster S. We assume one virtual computation step [vs] is performed on a fastest thread for one time unit [sec]. The maximum computation rate $maxCRT_f$ of a fastest server s_f is assumed to be one [vs/sec]. The total number VS_i of virtual computation steps of a process p_i is $minT_i$ [sec] $\cdot$ $maxCRT_f$ [vs/sec] = $minT_i$ [vs] for a fastest server s_t. The maximum computation rate $maxCR_t$ ($\leq nt_t$) of a server s_t is $np_t \cdot nc_t \cdot ct_t \cdot maxCRT_t = nt_t \cdot maxCRT_t$. The maximum computation rate $maxCR_{ti}$ of a process p_i on a server s_t is VS_i / $minT_{ti}$ = $minT_i$ / $minT_{ti}$ (≤ 1). For every pair of processes p_i and p_j on a server s_t, $maxCR_{ti}$ = $maxCR_{tj}$ = $maxCRT_t$. $CR_{ti}(\tau)$ ($\leq maxCR_t$) indicates the computation rate [vs/sec] of a process p_i on a server s_t at time τ is defined as follows:

[Multi-level computation with multiple CPUs (MLCM) model] The computation rate $CR_{ti}(\tau)$ [vs/sec] of a process p_i on a server s_t at time τ is given as follows:

$$CR_{ti}(\tau) = \begin{cases} nt_t \cdot maxCR_t \; / \, |CP_t(\tau)| & if \; |CP_t(\tau)| > nt_t. \\ maxCRT_t & if \; |CP_t(\tau)| \leq nt_t. \end{cases} \quad (3)$$

Figure 1 shows the computation rate $CR_{ti}(\tau)$ of a process p_i on a server s_t with nt_t threads. $CR_{ti}(\tau)$ is constant (= $maxCRT_t$) if a fewer number n = $|CP_t(\tau)|$ of processes than nt_t are performed. If $n \geq nt_t$, $CR_{ti}(\tau)$ = $maxCR_t$ (= $nt_t \cdot maxCRT_t$) / n. Suppose a pair of servers s_u and s_v are composed of nt_u (= nt_t / 2) and nt_v (= nt_t / 4) of threads, where each thread is the same as the server s_t, i.e. $maxCRT_u$ = $maxCRT_v$ = $maxCRT_t$. The dotted lines show the computation rates $CR_{ui}(\tau)$ and $CR_{vi}(\tau)$ of the servers s_u and s_v, respectively. The computation rate $CR_t(\tau)$ of a server s_t at time τ is $\sum_{p_i \in CP_t(\tau)} CR_{ti}(\tau)$. $CR_t(\tau)$ = $|CP_t(\tau)| \cdot maxCRT_t$ if $|CP_t(\tau)| \leq nt_t$, else $maxCR_t$.

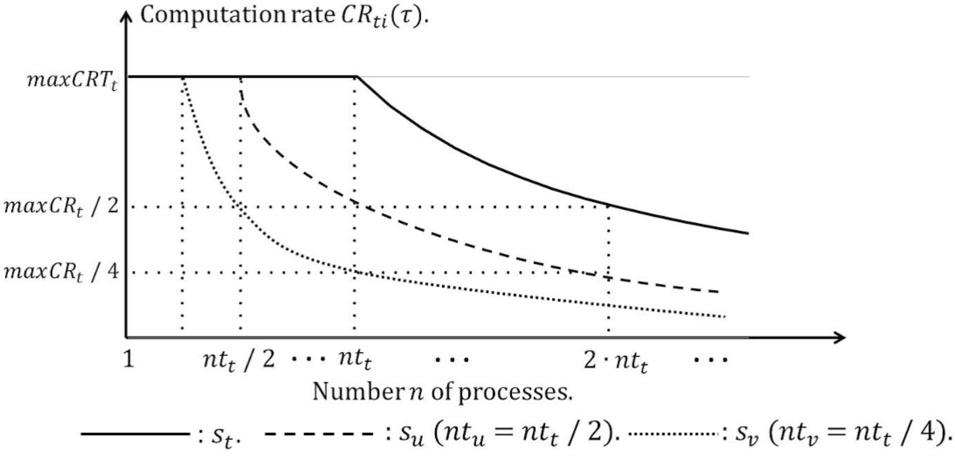


Fig. 1 Computation rate.

Suppose a process p_i on a server s_t starts at time *st* and ends at time *et*. Here, $\sum_{\tau=st}^{et} CR_{ti}(\tau) = VS_i$ [vs] = $minT_i$ which shows the total amount of computation to be performed by a process p_i.

At time τ a process p_i starts, the computation laxity $lc_{ti}(\tau)$ of a process p_i at time τ is VS_i. Then, at each time τ, $lc_{ti}(\tau)$ is decremented by the computation rate $CR_{ti}(\tau)$, i.e. $lc_{ti}(\tau + 1) = lc_{ti}(\tau) - CR_{ti}(\tau)$. If $lc_{ti}(\tau + 1)$ gets 0, the process p_i terminates at time τ.

2.3 Estimation Model

We first assume no process additionally starts on a server s_t after time τ. At each time τ, we can get the expected termination time ET_t and the total expected electric energy EE_t [J] of the server s_t to perform all the current processes in the process set $CP_t(\tau)$ by the following algorithm **EST** (s_t, τ, $CP_t(\tau)$; EE_t, ET_t) [5] [7] [8]:

[Estimation algorithm]

```
EST ($s_t$, $\tau$, $CP$; $EE$, $ET$) {
input $s_t$, $\tau$, $CP$;
output $EE$, $ET$;
  $EE$ = 0;   $x = \tau$;
  while   ($CP \neq \phi$) {
  for each process $p_i$ in $CP$ {
        $lc_{ti}(x+1) = lc_{ti}(x)$ - $CR_{ti}(x)$;
        $EE = EE + CE_t(|CP|)$; /* electric power */
        if $lc_{ti}(x+1) \leq 0$, {     /* $p_{ti}$ terminates*/
            $CP = CP - \{p_i\}$; $ET = x$;        };
   }; /* for end */
   $x = x + 1$; /* time advances */
  }; /* while end */
  $ET = x - \tau$;
};
```

Initially, CP is a set $CP_t(\tau)$ of current processes and x is current time τ. At each time x, the laxity $lc_{ti}(x)$ is decremented by the computation rate $CR_{ti}(x)$ for each current process p_i in CP, i.e. $lc_{ti}(x+1) = lc_{ti}(x) - CR_{ti}(x)$. EE is incremented by the power consumption $E_t(\tau) = CE_t(|CP|)$. If $lc_{ti}(x+1) \leq 0$, the process p_i terminates at time x and is removed from the process set CP, i.e. $CP = CP - \{p_i\}$. Here, x shows the expected termination time of the process p_i. If CP gets empty, every process in $CP_t(\tau)$ terminates at time x - 1. Here, $ET = (x - \tau)$ is the expected termination time of the server s_t and EE is the expected electric energy consumption of the server s_t obtained at time τ.

3 Energy-aware Server Selection Algorithms

3.1 Locally Energy-aware (LEA) Algorithm

A client first issues a request process p_i to a cluster S of m (≥ 1) servers $s_1, \ldots, s_m$ at time τ. In the locally energy-aware (LEA) algorithm [17], the expected electric energy consumption EE_t of each server s_t to perform a new process p_i and every current process in $CP_t(\tau)$ is obtained by the estimation procedure **EST** (s_t, τ, $CP_t(\tau) \cup \{p_i\}$; EE_t, ET_t) [20], [21]. Then, one server s whose expected energy consumption EE is minimum is selected as follows:

```
[LEA algorithm]
LEA ($\tau$, $p_i$; $s$, $EE$, $ET$)
input: $\tau$, /* current time */   $p_i$; /* process issued by a client */
output: $s$, /* server to perform $p_i$ */
   $EE$, /* expected electric energy */ $ET$; /* expected termination time */
   { $EE = \infty$;
   for each server $s_t$ {
      $CP = CP_t(\tau) \cup \{p_i\}$;
      /* $p_i$ starts at time $\tau$ */
      EST ($s_t$, $\tau$, $CP$; $EE_t$, $ET_t$);
      if $EE_t < EE$, {   $s = s_t$;  $EE = EE_t$;  $ET = ET_t$; };
   }; /* for end */
};
```

The process p_i is performed on the selected server s. For each new process p_i, all the servers are checked in a cluster S. Hence, the computation complexity of the LEA algorithm is $O(m)$ for the number m of servers.

3.2 Modified Globally Energy-aware (MEA) Algorithm

In the globally energy-aware (GEA) algorithm [18], one server s is selected for a new process p_i at current time τ, where the expected total electric energy consumption of all the servers is minimum in a cluster S. In the GEA algorithm, all the servers are checked for each server s_t and the expected total electric energy GE_t of all the servers is obtained. Then, a server s_t whose GE_t is minimum is selected. Hence, the computation time of the GEA algorithm is $O(m^2)$. The modified GEA (MEA) algorithm is proposed to reduce the computation time as follows [20], [21]:

```
[MEA algorithm]
MEA ($\tau$, $p_i$ ; $s$, $EE$, $ET$)
input: $\tau$, $p_i$;
output: $s$, $EE$, $ET$;
```

```
{ for each server $s_t$,  {
      EST($s_t$, $\tau$, $CP_t(\tau)$; $EE_t$, $ET_t$);
      EST($s_t$, $\tau$, $CP_t(\tau)$, $\cup$ $\{p_i\}$; $NE_t$, $NET_t$);
  }; /* for end */
  $XET$ = max($\{ET_1, \ldots, ET_m\}$);        $GE = \infty$;
  for each server $s_t$,
      if $ET_t = 0$ /* no process */, $EE_t = minE_t \cdot XET$
      else if $ET_t < XET$, $EE_t = EE_t + minE_t \cdot (XET - ET_t)$;
  select a server $s_t$ such that $GE_t = NE_t + \sum_{s_u \in S(t \neq u)} EE_u$ is minimum;
  $s = s_t$;
  $EE = GE_t$; $ET = NET_t$;        };
};
```

In the MEA algorithm, the expected total electric energy consumption EE_t and expected termination time ET_t of each server s_t are first calculated where only current processes in the set $CP_t(\tau)$ are assumed to be performed. Then, we obtain the expected energy consumption NE_t and expected termination time NET_t of each server s_t where not only all the current processes in $CP_t(\tau)$ but also the new process p_i are to be performed. If the process p_i is performed on a server s_t, $EE_t \leq NE_t$ and $ET_t \leq NET_t$. Here, the total expected energy consumption GE_t of all the servers is $EE_1 + \ldots + EE_{t-1} + NE_t + EE_{t+1} + \ldots + EE_m$. A server s_t where GE_t is minimum is selected to be a server where the process p_i is to be performed. It takes time $O(m)$ to find a server s_t for a process p_i.

3.3 Simple Energy-aware (PEA) Algorithm

In the LEA and MEA algorithms, the expected electric energy EE_t and termination time ET_t of each server s_t have to be calculated. In order to reduce the computation time of the algorithms, we define the energy-computation facter EC_t of each server s_t to be $maxCR_t$ [vs/sec] $\cdot$ $maxE_t$ [W]. We newly introduce a metric, the energy computation factor EC_t multiplied by the number n (= $|CP_t(\tau)|$) of processes performed on a server s_t to show the expected electric energy EE_t to perform the processes at time τ, $EE_t = EC_t \cdot n$. We propose a *simple energy-aware* (*PEA*) algorithm. Here, a server s_t where $EC_t \cdot (|CP_t(\tau)| + 1)$ is minimum is selected to perform a process p_i as follows:

[PEA algorithm]

1. **select** a server s_t where $EC_t \cdot (|CP_t(\tau)| + 1)$ is minimum in a cluster S.

The process p_i is then performed on the selected server s_t.

In the PEA algorithm, only the number $|CP_t(\tau)|$ of current processes is required to be collected from each server s_t to select a server for a process p_i. Thus, the PEA algorithm is simpler than the other algorithms, i.e. the computation time is $O(1)$.

4 Evaluation

4.1 Environment

We evaluate the PEA algorithm in terms of total electric energy consumption [J] of a cluster S and average execution time of processes. In the evaluation, we consider four real servers (m = 4), DSLab4, DSLab, Sunny, and Atria. The servers DSLab4 and DSLab are equipped with four and two Intel Xeon E5-2667 v2 CPUs, respectively, Sunny with an Intel Xeon E5-2620 CPU, and Atria with an Intel Corei7-6700K CPU. Every server s_t is characterized in terms of electric power consumption parameters like $minE_t$, CPU parameters like np_t, and computation parameters like $maxCRT_t$ as shown in Table 1. The servers DSLab4 and DSLab are fastest servers, i.e. $maxCRT_{DSLab4} = maxCRT_{DSLab} = 1$. The maximum computation rate $maxCR_t$ is $nt_t \cdot maxCRT_t$ where nt_t is the total number of threads supported by a server s_t.

Totally n ($\geq$ 1) processes $p_1, \ldots, p_n$ are performed on the four servers. For each process p_i, the starting time st_i is randomly taken from 0 to $xtime$ - 1. One time unit [tu] is assumed to be 100 [msec] since the power consumption of a server can be measured every 100 [msec] [23]. The simulation time $xtime$ is 2,500 [tu], i.e. 250 [sec]. The minimum computation time $minT_i$ of each process p_i is randomly taken out of 5 to 10 [tu], i.e. 0.5 to 1 [sec]. That is, the number VS_i of virtual computation steps of each process p_i is 5 to 10 [vs]. et_i is time when a process p_i terminates. The termination time et_i of each process p_i is obtained in the simulation. Here, the execution time T_i of a process p_i is (et_i - st_i) $\geq$ $minT_i$. The simulation ends at time $etime$ when every process terminates on a server s_t. That is, $etime$ is max ($\{et_1, \ldots, et_n\}$) which may be longer than $xtime$. The process parameters are summarized in Table 2.

At each time τ, if there is a process p_i whose starting time st_i is τ, one server s_t is selected in a selection algorithm. Then, the process p_i is selected on the server s_t. The computation laxity $lc_{ti}(\tau)$ of each current process p_i on each server s_t is decremented by the computation rate $CR_{ti}(\tau)$ at each time τ. The simulation program is implemented in SQL of Sybase [28] and data like termination time et_i of each process p_i and electric energy consumption EE_t of each server s_t obtained in the simulation are stored in tables of the database.

We evaluate the PEA algorithm compared with the round-robin (RR), random (RD), LEA [17], and MEA [20] [21] algorithms to select a server for each process p_i. In the RR algorithm, a server s_t is selected for each process p_i after a server s_{t-1} is selected in the cluster S. Here, the servers in the cluster S are ordered as DSLab4, DSLab, Sunny, and Atria. In the RD algorithm, one server is randomly selected for each process p_i in the cluster S.

Table 1 Parameters of servers.

parameters	DSLab4	DSLab	Sunny	Atria
np_t	4	2	1	1
nc_t	8	8	6	4
nt_t	64	32	12	8
$maxCRT_t$ [vs/tu]	1.0	1.0	0.5	0.7
$maxCR_t$ [vs/tu]	64	32	6	5.6
$minE_t$ [W]	126.1	126.1	87.2	41.3
$maxE_t$ [W]	454.1	301.1	131.2	89.5
bE_t [W]	30	30	16.6	15.9
cE_t [W]	5.6	5.6	3.6	4.7
tE_t [W]	0.8	0.8	0.9	1.1

Table 2 Parameters of processes.

parameters	values
n	number of processes $p_1, \ldots, p_n$ (≥ 0)
$minT_i$ [tu]	minimum computation time of a process p_i
VS_i [vs]	0.5 ∼ 1.0 ($VS_i = minT_i$)
st_i	starting time of p_i ($0 \leq st_i < xtime$ - 1)
$xtime$	simulation time (= 2,500 [tu] (= 250 [sec]))

4.2 Evaluation Results

A process configuration PC_n for n (≥ 1) processes $p_1, \ldots, p_n$ is a tuple $\langle\langle st_1, VS_1\rangle, \ldots, \langle st_n, VS_n\rangle\rangle$ where the starting time st_i and amount of computation VS_i of each process p_i are randomly taken. For n, ten processes configurations PC_n are randomly generated. In the RR, RD, LEA, MEA, and PEA algorithms, the n processes $p_1, \ldots, p_n$ in each process configuration PC_n are performed on the four servers (m = 4) shown in Table 2. Then, the total electric energy consumption EE and average active time AT of the servers and the average execution time ET of the processes are obtained

The ratio of the total electric energy (TER) is given as EE / ($\sum_{t=1}^{m} etime \cdot maxE_t$). Figure 2 shows the ratio of the total electric energy consumption EE of the cluster S of four servers (m = 4) for the number n of processes in the algorithms. The TER of the MEA algorithm is minimum but the computation complexity is $O(m)$. The TER of the PEA algorithm is almost the same as the MEA algorithm and is smaller than the other algorithms and the computation complexity is $O(1)$, smaller than the MEA algorithm.

Figure 3 shows the ratio of the total active time of the servers (TAR) for the number n of processes where there are four servers (m = 4). Let AT_t be total active time of a server s_t when at least one process is performed, i.e. $\sum_{\tau=0}^{etime-1} \delta_t(\tau)$ where $\delta_t(\tau)$ = 1 if $\mid CP_t(\tau) \mid \geq 1$, otherwise 0. TAR is $\sum_{t=1}^{m} AT_t/(etime \cdot m)$. The TAR of the MEA algorithm is minimum and the TAR of the PEA algorithm is almost the same as the MEA algorithm.

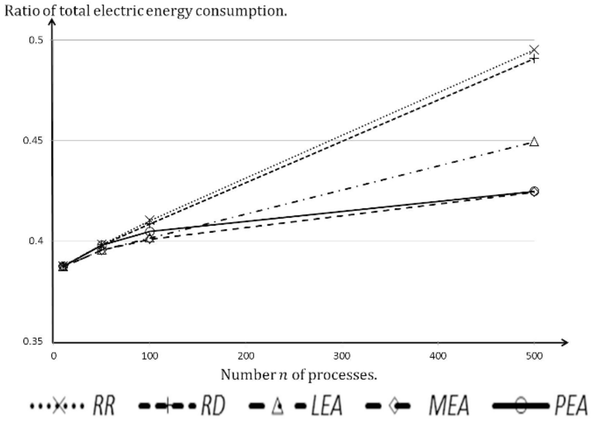


Fig. 2 Ratio of total electric energy consumption (m = 4).

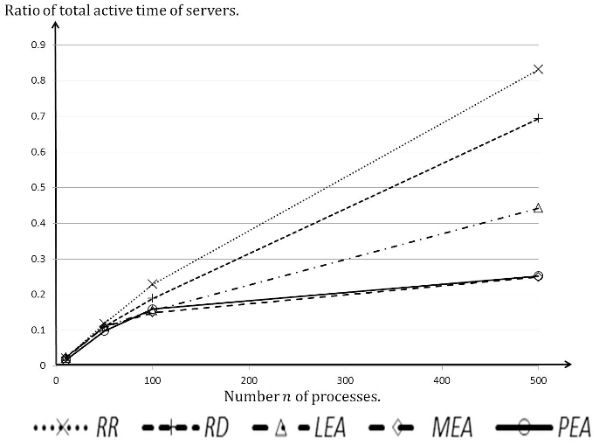


Fig. 3 Ratio of total active time of servers (m = 4).

5 Concluding Remarks

We newly proposed the PEA algorithm to select a server in a cluster based on the MLPCM and MLCM models. In the PEA algorithm, one server is selected for a process issued by a client, where the energy-computation factor multiplied by the number of current processes is minimum. We evaluated the PEA algorithm in terms of the total electric energy consumption, total active time of the servers and the average execution time of the processes compared with the RR, RD, LEA [17], and MEA [20] [21] algorithms. The total electric energy consumption and total active time of servers in the PEA algorithm are almost the same as the MEA algorithm, smaller than the other algorithms. The computation complexity of the PEA algorithm is $O(1)$ and simpler than the LEA and MEA algorithms. We conclude the

PEA algorithm is practical to select a server for a request process in a scalable cluster.

Acknowledgements This work was supported by JSPS KAKENHI grant number 15H0295.

References

1. Bianchini, R. and Rajamony, R.: Power and Energy Management for Server Systems. IEEE Computer, **37**(11), pp. 68-74. (2004).
2. DSlab server: http://h50146.www5.hp.com/products/servers /proliant/system_pdf/dl360pgen8.pdf.
3. Duolikun, D., Aikebaier, A., Enokido, T., and Takizawa, M.: Energy-aware Passive Replication of Processes. Journal of Mobile Multimedia, **9**(1&2), pp. 53-65, (2013).
4. Duolikun, D., Aikebaier, A., Enokido, T., and Takizawa, M.: Energy-efficient Dynamic Cluster of Servers. Journal of Supercomputing, vol.71, no.5, pp. 1647-1656, (2015).
5. Enokido, T., Aikebaier, A., and Takizawa, M.: A Model for Reducing Power Consumption in Peer-to-Peer Systems. IEEE Systems Journal, **4**(2), pp. 221-229, (2010).
6. Enokido, T., Aikebaier, A., and Takizawa, M.: An Integrated Power Consumption Model for Communication and Transaction Based Applications. Proc. of IEEE the 25th International Conference on Advanced Information Networking and Applications (AINA-2011), pp. 627-636, (2011).
7. Enokido, T., Aikebaier, A., and Takizawa, M.: Process Allocation Algorithms for Saving Power Consumption in Peer-to-Peer Systems. IEEE Transactions on Industrial Electronics, **58**(6), pp. 2097-2105, (2011).
8. Enokido, T., Aikebaier, A., and Takizawa, M.: An Extended Simple Power Consumption Model for Selecting a Server to Perform Computation Type Processes in Digital Ecosystems. IEEE Transactions on Industrial Informatics, **10**(2), pp. 1627-1636, (2014).
9. Enokido, T., Aikebaier, A., and Takizawa, M.: Evaluation of the Extended Improved Redundant Power Consumption Laxity-Based (EIRPCLB) Algorithm, Proc. of IEEE the 28th International Conference on Advanced Information Networking and Applications (AINA-2014), pp.940-947, (2014).
10. Enokido, T. and Takizawa, M.: Energy-Efficient Delay Time-Based Process Allocation Algorithm for Heterogeneous Server Clusters. Proc. of IEEE the 29th International Conference on Advanced Information Networking and Applications (AINA-2015), pp. 279-286, (2015).
11. Enokido, T. and Takizawa, M.: Power Consumption and Computation Models of Virtual Machines to Perform Computation Type Application Processes. Proc. of the 9th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2015), pp. 126-133, (2015).
12. Google, Google green, http://www.google.com/green/, (2015).
13. Inoue, T., Aikebaier, A., Enokido, T., and Takizawa, M.: Power Consumption and Processing Models of Servers in Computation and Storage Based Applications. Journal of Mathematical and Computer Modeling, **58**(5&6), pp. 1475-1488, (2013).
14. Intel Xeon Processor 5600 Series: The Next Generation of Intelligent Server Processors, white paper [online]. Available: http://www.intel.com/content/www/us/en/processors/xeon/xeon-5600-brief.html, (2010).
15. Job Scheduling Algorithms in Linux Virtual Server, http://www.linuxvirtualserver.org/docs/scheduling.html, (2010).
16. Kataoka, H., Duolikun, D., Enokido, T., and Takizawa, M.: Power Consumption and Computation Models of a Server with a Multi-core CPU and Experiments. Proc. of IEEE the 29th IEEE International Conference on Advanced Information Networking and Applications Workshops (WAINA-2015), pp. 217-223, (2015).

17. Kataoka, H., Duolikun, D., Enokido, T., and Takizawa, M.: Evaluation of Energy-aware Server Selection Algorithm. Proc. of the 9th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2015), pp. 318-325, (2015).
18. Kataoka, H., Duolikun, D., Enokido, T., and Takizawa, M.: Multi-level Computation and Power Consumption Models. Proc. of the 18th International Conference on Network-Based Information Systems (NBiS-2015), pp. 40-47, (2015).
19. Kataoka, H., Duolikun, D., Enokido, T., and Takizawa, M.: Energy-efficient Virtualisation of Threads in a Server Cluster. Proc. of the 10th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA-2015), pp. 288-295, (2015).
20. Kataoka, H., Sawada, A., Duolikun, D., Enokido, T., and Takizawa, M.: Energy-aware Server Selection Algorithm in a Scalable Cluster. The 30th IEEE International Conference on Advanced Information Networking and Applications (AINA-2016), pp. 565-572, (2016).
21. Kataoka, H., Sawada, A., Duolikun, D., Enokido, T., and Takizawa, M.: Energy-aware Algorithms to Select Servers in Scalable Clusters. Proc. of the 10th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2016), pp. 308-315, (2016).
22. Linux distribution. Available:
http://itpro.nikkeibp.co.jp/article/COLUMN/20120223/382669/.
23. Metaprotocol Corp: ”UWmeter” [online]. Available: http://www.metaprotocol.com/UWmeter /Feautures.html, (2011).
24. Natural Resources Defense Council (NRDC): Data Center Efficiency Assessment - Scaling up Energy Efficiency across the Data Center Industry: Evaluating Key Drivers and Barriers, http://www.nrdc.org/energy/files/data-center-efficiencyassessment-IP.pdf, (2014).
25. Sawada, A., Kataoka, H., Duolikun, D., Enokido, T., and Takizawa, M.: Energy-aware Clusters of Servers for Storage and Computation Applications. Proc. of the IEEE 30th International Conference on Advanced Information Networking and Applications (AINA-2016), pp. 400-407, (2016).
26. Sawada, A., Kataoka, H., Duolikun, D., Enokido, T., and Takizawa, M.: Selection Algorithms to Select Energy-efficient Servers for Storage and Computation Processes. Proc. of the 19th International Conference on Network-Based Information Systems (NBiS-2016), CD-ROM, (2016).
27. Schollmeier, R.: A Definition of Peer-to-Peer Networking for the Classification of Peer-to-Peer Architectures and Applications. Proc. of the First International Conference on Peer-to-Peer Computing, pp. 101-102, (2011).
28. http://www.cultofmac.com/167829/sybasesap-afaria-offers-ios-and-pc-management-options-mobile-management-month/

An Energy-efficient and Reliable Protocol in Wireless Networks

Emi Ogawa, Shigenari Nakamura, and Makoto Takizawa

Abstract In recent years, wireless ad-hoc networks which do not use network infrastructure are getting more important in various applications like vehicle-to-vehicle (V2V) networks. Especially, it is easy for nodes to communicate with one another without network infrastructure like access points and base stations. Here, ad-hoc routing protocols are needed to deliver messages to destination nodes. In wireless ad-hoc networks, neighbor nodes with which each node can directly communicate may be changed, e.g due to movement of nodes and faults of nodes. Besides, since nodes in ad-hoc networks work with their electric batteries, it is required to reduce electric energy consumed by nodes to send and receive messages. It is difficult to always find a best route to a destination node due to complex parameters like bandwidth and message loss ratio. In this paper, we newly propose a reliable one-to-one communication protocol named TBAH (Trustworthiness-Based Ad-Hoc communications) protocol where routes to the destination nodes are found by using the fuzzy logic in wireless networks.

1 Introduction

We are able to take advantage of kinds of service supported by various types of networks like wireless networks. Wireless ad-hoc networks [6] [9] [8] are getting important in various types of applications, especially in disaster environments [10]. In wireless ad-hoc networks, nodes can communicate with other nodes in wireless networks without network infrastructure like access points. Nodes which are

Emi Ogawa
Hosei University, Tokyo, Japan, e-mail: emi.ogawa.2q@stu.hosei.ac.jp

Shigenari Nakamura
Hosei University, Tokyo, Japan, e-mail: nakamura.shigenari@gmail.com

Makoto Takizawa
Hosei University, Tokyo, Japan, e-mail: makoto.takizawa@computer.org

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_57

in wireless communication range of a node are first-neighbor nodes of the node. A node can deliver messages to first-neighbor nodes in a wireless network. If a node p_i would like to send messages to another node p_j which is out of communication range of the node p_i, some first-neighbor node p_k of the node p_i has to forward the messages to the node p_j. Thus, we have to discuss ad-hoc routing protocols to deliver each message to a destination node in wireless ad-hoc networks.

Many routing protocols are so far discussed in wireless ad-hoc networks [6] [8] [9] [11]. The routing protocols can be categorized into reactive and proactive types. In reactive routing protocols like the AODV protocol [9], routes from a source node to a destination node is not *a priori* fixed until the source node is required to send messages. A route is tried to be found on-demand. Here, it takes time to start communication between a pair of source and destination nodes because a source-to-destination route has to be found. However, control messages are not transmitted to maintain a route if the nodes are not communicating. On the other hand, in proactive routing protocols like the DSDV protocol [8] and the OLSR protocol [6], routes to all destination nodes are always kept in the network. Here, the network traffic increases because nodes periodically send control messages to obtain the topology information of network. The advantage of proactive protocols is that protocols can immediately provide the required route when needed.

In this paper, we newly propose a reactive type of ad-hoc routing protocol named trustworthiness-based ad-hoc communication (TBAH) protocol based on the trustworthiness concept [14]. In this paper, the trustworthiness of a neighbor node is defined in the Fuzzy logic [7]. A more trustworthy first-neighbor node is selected to be in a route in the TBAH protocol. We evaluate the TBAH protocol compared with the AODV protocol in terms of energy consumption of nodes and show the total electric energy of nodes can be reduced in the TBAH protocol.

In section 2, we present a system model. In section 3, we discuss the Fuzzy-based Trustworthiness. In section 4, we present the TBAH protocol. In section 5, we evaluated the TBAH protocol.

2 System Model

A group G of n ($\geq$1) nodes $p_1, \ldots, p_n$ are cooperating with one another by exchanging messages in wireless networks like wireless sensor networks [3], [4], [**?**] and wireless ad hoc networks [5]. Let d_{ij} be the distance between a pair of nodes p_i and p_j. In this paper, we make the following assumptions on nodes:

1. The distance d_{ij} between every pair of nodes p_i and p_j are *a priori* known.
2. Each node does not move in networks.

A node p_i sends a message m with electric energy SE_i [J] in a wireless network. Some node p_j can receive the message m in the group G depending on the distance d_{ij} and electric energy SE_i. With the stronger electric energy SE_i a node p_i sends a message, the farther node from the node p_i can receive the message. Let $wd_i(SE_i)$

show the maximum communication range in which every node can receive messages which the node p_i sends with electric energy SE_i. That is, a node p_j can receive a message sent by a node p_i if $d_{ij} \leq wd_i(SE_i)$. The electric energy SE_i is proportional to a square of the distance $wd_i(SE_i)$ [15]. That is, $SE_1/SE_2 = wd_i(SE_1)^2/wd_i(SE_2)^2$ for a pair of electric energy SE_1 and SE_2 of each node p_i. Let $maxSE_i$ be the maximum electric energy [J] of a node p_i to send a message. Let $maxd_i$ show the maximum distance $wd_i(maxSE_i)$. A *first-neighbour* node p_j of a node p_i is a node which is in communication range $maxd_i$ of a node p_i, i.e. $d_{ij} \leq maxd_i$ $(= wd_i(maxSE_i))$. In this paper, we make the following assumptions on the energy consumption of a node p_i:

1. The electric energy SE_i to send a message to a node p_j is d_{ij}^2 where $d_{ij} = wd_i(SE_i)$.
2. Each node p_i can change the electric energy SE_i $(0 \leq SE_i \leq maxSE_i)$ to send a message.
3. Each node p_i does not consume electric energy to receive messages.

3 Fuzzy-based Trustworthiness

It is not easy to find an optimal route from a source node to a destination node since there are many parameters like bandwidth and message loss ratio to obtain an optimal solution. In this paper, we consider the trustworthiness of each node [1] [2] [14]. A node sends a message to a more trustworthy first-neighbour node to deliver to the destination node. We adopt the Fuzzy logic [7] to calculate the trustworthiness of a node. Suppose a node p_j is a first-neighbor node of a node p_i. A peer p_i obtains the trustworthiness Tw_{ij} of a first-neighbor node p_j by directly communicating with the node p_j. Through communication with a neighbor node p_j, a node p_i obtains the following parameters:

1. Battery charge B_j of a node p_j.
2. Forwarding ratio F_{ij} for a nodep_i to a neighbor node p_j.
3. Availability ratio A_{ij} for a nodep_i to a neighbor node p_j.
4. Power consumption P_{ij} of a node p_i to send a message to a node p_j.

The trustworthiness [14] Tw_{ij} of a first-neighbor node p_j for a node p_i is calculated by the Fuzzy logic. A Fuzzy set of each parameter is given as follows:

- H(High).
- M(Moderate).
- L(Low).

The membership functions for the parameters are shown in Figures 1, 2, 3, and 4. A part of the Fuzzy inference rules to calculate the trustworthiness Tw_{ij} from the Fuzzy sets are shown in Table 1.

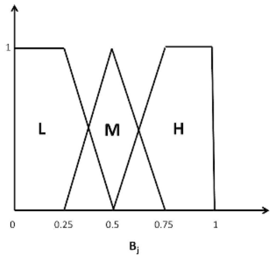


Fig. 1 Battery charge B_j.

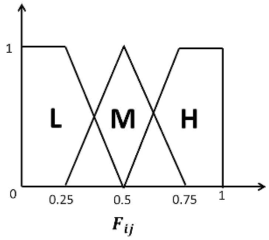


Fig. 2 Forwarding ratio F_{ij}.

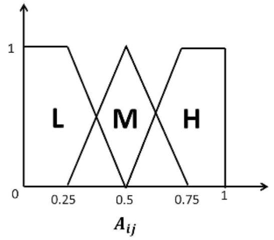


Fig. 3 Availability ratio A_{ij}.

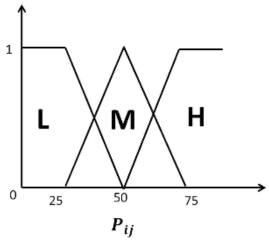


Fig. 4 Power consumption P_{ij}.

Table 1 Inference rules.

	battery charge (B_j)	forwarding ratio (F_{ij})	availability ratio (A_{ij})	power consumption (P_{ij})	trustworthiness (Tw_{ij})
1	*H*	*H*	*H*	*H*	*CT*
2	*H*	*H*	*H*	*M*	*CT*
3	*H*	*H*	*H*	*L*	*CT*
4	*H*	*H*	*M*	*H*	*CT*
5	*H*	*H*	*M*	*M*	*ST*
6	*H*	*H*	*M*	*L*	*ST*
7	*H*	*H*	*L*	*H*	*CT*
8	*H*	*H*	*L*	*M*	*ST*
9	*H*	*H*	*L*	*L*	*M*
10	*H*	*M*	*H*	*H*	*CT*
11	*H*	*M*	*H*	*M*	*ST*
12	*H*	*M*	*H*	*L*	*ST*
13	*H*	*M*	*M*	*H*	*ST*
14	*H*	*M*	*M*	*M*	*M*
15	*H*	*M*	*M*	*L*	*M*
16	*H*	*M*	*L*	*H*	*M*
17	*H*	*M*	*L*	*M*	*M*
18	*H*	*M*	*L*	*L*	*DU*
19	*H*	*L*	*H*	*H*	*CT*
...	...	...	...	...	...
78	*L*	*L*	*M*	*L*	*CU*
79	*L*	*L*	*L*	*H*	*CU*
80	*L*	*L*	*L*	*M*	*CU*
81	*L*	*L*	*L*	*L*	*CU*

Suppose a node p_j is a first-neighbor node of a node p_i and a node p_k is a first-neighbor node of the node p_j [Figure 5]. The transitive trustworthiness Tw_{ik} of the node p_k for the node p_i is a composition $Tw_{ij} \circ Tw_{jk}$.

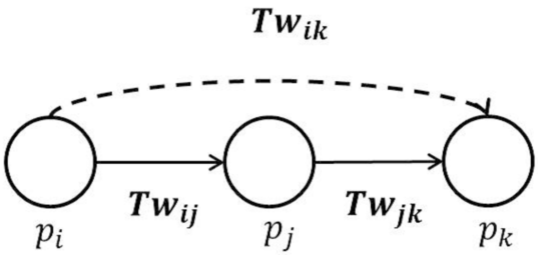


Fig. 5 Transitive Trustworthiness.

4 TBAH Protocol

Suppose a network N is composed of nodes $p_1, \ldots, p_n$ ($n \geq 1$) which are intercorrelated in wireless communication links. Each node p_i manipulates variables $p_i.l$ and $p_i.N$. Initially, $p_i.l$ = 0 and $p_i.N$ = ϕ. First, a source node p_s would like to deliver messages to a destination node p_d in the network N. The source node p_s first sends an RQ message q to every first-neighbor node p_i with the maximum electric energy SE_i = $maxSE$. Here, the RQ (request) message q brings a level parameter $q.l = p_i.l = 0$ and first-neighbor nodes $q.N = p_s.N = \phi$. If a node p_i receives an RQ message q from a node p_j, the level parameter l is incremented by one in a node p_i, i.e. $p_i.l$ = $q.l + 1$. $p_i.N$ = $p_i.N \cup \{p_j\}$, i.e. p_i finds the node p_j is a first-neighbor node of p_i. Thus, each node p_i recognizes what nodes are first-neighbor nodes of the node p_i. Then, the node p_i sends the RQ message q where $q.l$ = $p_i.l$ and $q.N$ = $p_i.N$. If the node p_i already receives the RQ message q, the node p_i neglects the RQ message q. Here, the node p_i just recognizes the node p_j to be a first-neighbor node, $p_i.N = p_i.N \cup \{p_j\}$. In addition, the node p_i keeps in record of $q.N$, i.e. a set $p_j.N$ of first neighbor nodes.

Eventually, the destination node p_d receives an RQ message q. A first-neighbor node p_i whose $p_i.l$ is minimum is selected similarly to the AODV protocol. If there is another neighbor node p_k where $p_k.l > p_j.l$ and $p_j \in p_k.N$, i.e. p_j is a first-neighbor node of p_k, the trustworthiness is checked. Here, if $Tw_{ij} < Tw_{ik} \circ Tw_{kj}$, the node p_k gets a node between p_i and p_j in a route. Otherwise, p_j is a next node to the node p_i. Then, a node p_k which is next to p_i is taken. If there are multiple next nodes, a node p_j is selected in $p_i.N$ where Tw_{ij} is maximum. Here, suppose a node p_j is selected to be next to the node p_i. The node p_i sends an RC (request confirmation) message C, whose destination $c.dest$ is p_j, with the electric energy $SE_i = d_{ij}^2$. On receipt of an RC message C from a node p_i, a node p_j where $c.dest$ = p_j recognizes that the node p_i is next from the node p_j. The node p_j does the same procedure. Then, the source node p_s eventually receives an RC message C from a node p_i. Here, a route from the source node p_s to the destination node p_d is established.

5 Evaluation

We evaluate the TBAH protocol in terns of electric energy consumption of nodes compared with the AODV protocol. In the evaluation, the trustworthiness Tw_{ij} of a first-neighbor node p_j for a node p_i is assumed to show the electric energy d_{ij}^2. That is Tw_{ij} = d_{ij}^2 and $Tw_{ik} \circ Tw_{jk}$ = Tw_{ik} + Tw_{jk}. First, n nodes $p_1 \dots p_n$ are deployed on $m \cdot m$ meshes. In the evaluation, $m = 5$. A network N is composed of seven nodes p_0, …, p_6, which is shown in Figure 6. A link between a pair of nodes shows a wireless link. Here, $maxd_i$ of each node p_i is 2.0 and $maxSE_i = maxd_i^2 = 4.00$. The node p_0 is the source node and the node p_6 is the destination node. A route from p_s to p_d is selected so that the number of hops is minimum in the AODV protocol. Hence, the node p_3 is chosen to be an intermediate node between the source node p_0 and the destination node p_6. The number of hops is 2 in the route $p_0 \rightarrow p_3 \rightarrow p_6$. The electric energy SE_0 is $d_{03}^2 = 2.00^2$ and SE_3 is $d_{36}^2 = 1.98^2$. The total electric energy consumption of the nodes p_0 and p_6 in the AODV protocol is $2.00^2 + 1.98^2 = 7.94$ as shown in Table 2. The node p_2 is between the node p_0 and the node p_3. The electric energy consumption SE_0 of the node p_0 to deliver a message to the node p_2 is $d_{02} = 1.43^2$ and SE_2 to deliver a message to the node p_3 $d_{23} = 0.92^2$. Here, d_{03}^2 $(= 2.00^2) > d_{02}^2$ $(= 1.43^2)$ + d_{23}^2 $(= 0.92^2)$. Thus, a message will not be directly transmitted to the node p_3 from the node p_0. The node p_2 is selected as an intermediate node between the node p_0 and the node p_3. In the same way, the node p_1 is selected as an intermediate node between the nodes p_0 and p_2. There are three routes from p_3 to p_6. Since $d_{34}^2(= 1.39^2) + d_{46}^2(= 2.00^2) > d_{36}^2(= 1.98^2) > d_{35}^2(= 1.80^2)$ + $d_{56}^2(= 0.30^2)$, nodes in the route $p_3 \rightarrow p_5 \rightarrow p_6$ to the node p_3 from the node p_6 via the node p_5 totally consume the smallest electric energy. Thus, the node p_5 is selected as an intermediate node. The route from the source node p_0 to the destination node p_6 is $p_0 \rightarrow p_1 \rightarrow p_2 \rightarrow p_3 \rightarrow p_5 \rightarrow p_6$ in the TBAH protocol. The number of hops is 5. The total electric energy consumed by nodes in the TBAH protocol is $d_{01}^2(= 0.50^2)$ + $d_{12}^2(= 1.30^2)$ + $d_{23}^2(= 0.92^2)$ + $d_{35}^2(= 1.80^2)$ + $d_{56}^2(= 0.30^2)$ = 6.14 as shown in Table 2. Table 2 shows the electric energy consumption and the number of hops of the AODV protocol and the TBAH protocol. As the result, according to Table 2, the number of hops in the TBAH protocol is larger than the AODV protocol. However, the electric energy consumption of nodes in the TBAH protocol is smaller than the AODV protocol.

Table 2 TBAH and AODV protocols.

	AODV	TBAH
Electric energy [J]	7.94	6.14
Number of hops	2	5

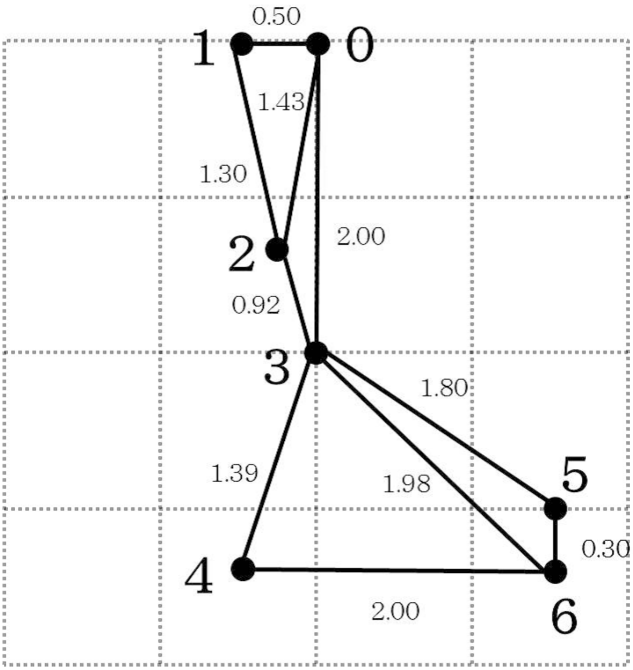


Fig. 6 Network *N*.

6 Concluding Remarks

In this paper, we proposed the trustworthiness-based ad-hoc routing (TBAH) protocol in wireless ad-hoc networks. The trustworthiness of first-neighbor nodes is obtained in the Fuzzy logic. We showed nodes in a route obtained in the TBAH protocol consume smaller electric energy than the AODV protocol.

Acknowledgment

This work was supported by JSPS KAKENHI grant number 15H0295.

References

1. Aikebaier, A., Enokido, T., and Takizawa, M.: Trustworthy Group Making Algorithm in Distributed Systemsl. Humancentric Computing and Information Sciences (HCIS), **11**(1), pp. 6:1–6:15, (2009).

2. Aikebaier, A., Enokido, A., and Takizawa, M.: Reliable and Efficient Way to Broadcast Messages in a Group by Trust-Based Broadcast (TBB) Scheme. Computing and Informatics (CAI), **30**(6), pp. 1001–1015, (2011).
3. Akyildiz, I. F. and Kasimoglu, H.: Wireless Sensor and Actor Networks: Research Challenges. Ad-Hoc Networks, **2**(4), pp. 351–367, (2004).
4. Akyildiz, I. F. and Kasimoglu, H.: A Survey on Sensor Networks. IEEE Communications Magazine, **40**(8), pp. 102–114, (2002).
5. Clausen, T., Hansen, G., Christensen, L., and Behrmann, G.: The Optimized Link State Routing Protocol, Evaluation through Experiments and Simulation. Proc. of the IEEE Symp, Conferenced on Wireless Personal Mobile Communications, (2001).
6. Jacquet, P., M´uhlethaler, P., Clausen, T., Laouiti, A., Qayyum, A., and Viennot, L.: Optimized Link State Routing (OLSR) Protocol for Ad Hoc Network. https://www.ietf.org/rfc/rfc3626.txt, (2003).
7. Mukaidono, M.: Fuzzy Logic for Beginners. World Scientific Publishing, 116 pages, (2001).
8. Perkins, C. E. and Bhagwat, P.: Highly Dynamic Destination Sequenced Distance Vector Routing (DSDV) for Mobile Computers. Proc. of ACM SIGCOMM'94, pp. 234–244, (1994).
9. Perkins, C. E., Royer, E. M., and Das, S. R.: Ad hoc On-demand Distance Vector (AODV) Routing. https://www.ietf.org/rfc/rfc3561.txt, (1997).
10. Sato, G., Uchida, N., Shiratori, N., and Shibata, Y.: Research on Never Die Network for Disaster Prevention Based on OpenFlow and Cognitive Wireless Technology. Proc. of IEEE the 30th International Conference on Advanced Information Networking and Applications (AINA-2016), pp. 370–375, (2016).
11. Sugino, M., Nakamura, S., Enokido, T., and Takizawa, M.: Energy-efficient Broadcast Protocols in Wireless Networks. Proc. of the 18-th International Conference on Network-Based Information Systems (NBiS-2015), pp. 357–364, (2015).
12. Sugino, M., Nakamura, S., Enokido, T., and Takizawa, M.: Protocols for Energy-efficiently Broadcasting Messages in Wireless Networks. Proc. of the AINA-2016 Workshops (WAINA-2016), pp. 286–293, (2016).
13. Waluyo, A. B., Taniar, D., Rahayu, W., Aikebaier, A., Takizawa, M., and Srinivasan, B.: Trustworthy-based Efficient Data Broadcast Model for P2P Interaction in Resource Constrained Wireless Environments. Journal of Computer and System Sciences, **78**(6), pp. 11716–1736, (2012).
14. Watanabe, K., Nakajima, Y., Enokido, T., and Takizawa, M.: Ranking factors in Peer-to-Peer Overlay Networks. ACM Transactions on Autonomous and Adaptive Systems, **2**(3), pp. 11:1–11:26, (2007).
15. Zhao, F. and Guibas, L.: Wireless Sensor Networks: An Information Processing Approach. Morgan Kaufmann Publishers, (2004).

Proposal of Collaborative Object Tracking Methods by Multi-Drones for Flight Surveillance Systems

Tomoki Okutake[1], Noriki Uchida[2], Noriyasu Yamamoto[3]
[1] Graduate School of Engineering, Fukuoka Institute of Technology, 3-30-1 Wajirohigashi, Fukuoka Higashi-ku, Fukuoka 811-0214, Japan
Emails: mgm16102@bene.fit.ac.jp
[2] Fukuoka Institute of Technology, 3-30-1 Wajirohigashi, Fukuoka Higashi-ku, Fukuoka 811-0214, Japan
Emails: n-uchida@fit.ac.jp
[3] Fukuoka Institute of Technology, 3-30-1 Wajirohigashi, Fukuoka Higashi-ku, Fukuoka 811-0214, Japan
Emails: nori@bene.fit.ac.jp

Abstract. Although there are rapid growth the drone's applications such as surveillance systems and delivery systems in recent years, the problems of the accidents are also increasing conversely. Therefore, this paper discusses the collaborative object tracking methods by multi-drones. The proposed methods consist of the function of the motion tracking and the WiFi detections for each drone, and choose a proper level of emergent procedures in order to avoid the drone's accident in various circumstances. Then, the prototype system of the proposed methods is introduced and discussed for the future works.

1 Introduction

Although there are rapid growth the drone's applications such as military uses, surveillance systems, and pesticide sprayings for harvests other than hobby usages in recent years, the problems of the accidents are also increasing conversely. For instance, the drone flew into the White House in the U.S. and it crashed on the ground in January 2015 [1]. The accidents brought the strong shocks through the whole over the world and it influenced to the regulation of the drone in the U.S. Also, in Japan, the crashed drone was found on the rooftop of the prime minister's official residence in April 2015 [2], it became the significant issues in the country. Moreover, only in Japan, there have been many accidents that caused from the miss controls or malfunctions of drones like that the uncontrollable drone crashed in to the snow pavement in February 2014[3] and that the uncontrollable military drones moved away military zone in March 2016 [4].

Therefore, this paper discusses the collaborative object tracking methods by multi-drones. The proposed methods mainly consist of the motion tracking and the WiFi

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_58

detections of each drone and the collaborative drone's controls by multiple drones attempt to avoid the emergent uncontrollable situations. The paper also introduces the prototype system and discussed for the future work.

In the followings, the section II discuss about the related previous works of drone's accidents and its researches, and the proposed systems are introduced in section III. Then, the prototype system is explained in section IV, and section V deals with the experimental reports, and the conclusion and future study is discussed in section VI.

2 Related Works

The merchant drones (Quad-copter, UAV) such as figure 1 [5] are rapidly spread over the whole world in the recent. They usually consist of multi-propeller in order to easy control in the air. Also, the wireless IP network, GPS, and cameras are usually equipped for the hobby usages, and so the drones can be controlled by smartphones or programmed procedures. Thus, these drones are recently used by the purpose of not only hobbies but also various researches such as surveillance system or ad-hoc network.

Fig. 1. The picture of the merchant drone

For example, there are some previous approaches [6] that drones are controlled by the pattern recognition with cameras. The paper introduced the implementations of the specific signs of the images and it discussed about the additional necessities for the drone controls.

Besides, the papers such as [7] proposed the guidance control with collision avoidance for multiple drones under communication restricted. In the paper, if drones are out of communication ranges, drones acquire GPS points and avoid other drones by using distributed nonlinear model predictive control.

Also, Bills and at el. [8] discussed the problem of autonomously flying drones in indoor environments such as home and office buildings. In the paper, the primary long range sensor in these drone is a miniature camera. The method neither attempts to build nor requires a 3D model. Instead, the method first classifies the type of indoor environment the drone is in, and then uses vision algorithms based on perspective cues to estimate the desired direction to fly.

Then, Ito and at el. [9] introduced a highly efficient space searching method that is flexible to environment and independent of data-handling capacity. With onboard camera and image processing technique, Drones can locate its position. The method is applicable to multiple Drones and by deploying them.

However, the accidents by the drones are increasing conversely with the expectation of these new kinds of application, and it is necessary to consider the function of safety managements for the usages.

3 The Proposed Methods

This paper proposes the collaborative safety flight control system for multiple drones. In the case of two drone operations, each drone observes another by the status of the pattern recognition of other drones and the wireless connections. Fig. 2 shows the example operations for the surveillance systems by the two drones, and one is set as the host drone as the observing the status of the operation and another is set as the child drone to be observed.

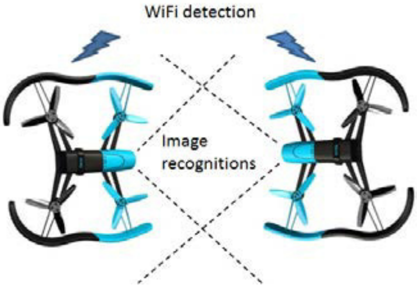


Fig. 2. The example of two drone operations

First of all, the failure level is decided by the status of the pattern recognition of another drone and the wireless connections. The following Table 1 is considerable three levels of the drones' operations for the safety flight controls.

Table 1. The considerable failure levels collaborative flight

Levels	Details
High	Both the camera and the wireless communication do not work
Middle	The camera doesn't work while the wireless communication works Out of battery
Low	Both the camera and the wireless communication work

In Table 1, when the child drone is in the serious conditions that there is no wireless connection and the drone is out of the view, the proposed system becomes the high emergent operation. In the emergent, the observed and malfunctioned drone beeps the alert sound in order to notify the emergency for the person where is near the drone. Then, the drone automatically seeks safety landing spot if it is possible, and it confirms the emergent landing. These functions are implemented in the separated circuit in the drones, and so the system reduces the risk of main circuit controls. On the other hands, the host drone firstly moves to the GPS point where the target drone was disappeared. Then, the host seeks the missing drone, and the emergent notification is also transmitted to the user. Then, if the host drone can find the missing drone by the camera sensor, the GPS location and the camera image are transmitted to the user. Thus, the user can notice the situation of the missing drone.

Secondly, in the middle emergent level, if there is a wireless connection but out of camera's view, the observed drone acquire the current GPS location and the data is transmitted to the host drone. Then, the host drone receives the GPS point, and the received GPS point is used for the modification of the observed drone's location. After the modification of the observed drone's location, the operation becomes the emergent operation if the modification is unable.

Thirdly, if one of drones is out of battery, both drones return to the charge station where the location is previously installed. TABLE2 shows the summary of the middle emergent level controls.

Table 2. The summary of the middle emergent level controls

Status	Host drone	Child drone
When there is no wireless connections	The drone move to the GPS point where the target drone was disappeared. Then, the drone seeks the missing drone, and the alert is also transmitted to the user.	The drone beeps an alert signal, and it confirms an emergent landing.
When the target is the out of cam's view	The received GPS point is used for the modification of the observed drone's location	The GPS point is collected, and the data is transmitted to the host drone.
When one or two drones are out of battery	Both drones return to the charge station where the location is previously installed.	Both drones return to the charge station where the location is previously installed.

Then, Figure 2 is about the flow chart of the modules of the proposed system in details. Firstly, the drone recognition is worked for the collaborative flight, and then the emergent level is detected if the drone is out of camera view as the flow chart.

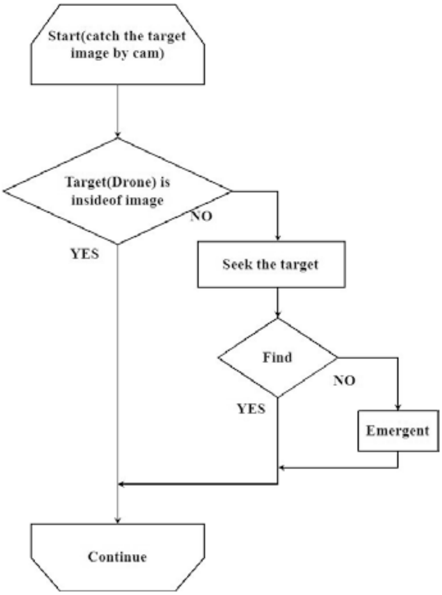

Start(catch the target image by cam)
Target(Drone) is insideof image
NO
YES
Seek the target
Find
NO
YES
Emergent
Continue

Fig. 3. Flow chart of motion tracking

These operational methods are basically applied to the three or more drones' control system in the proposed system as the setting of the host drone and the child drones.

4 The Prototype System

For the effectiveness of the proposed methods, the implementations of the prototype system are under working. Table 3 shows the specification of the drone. In the system, the drones consisted of wireless interfaces such as IEEE802.11a/b/g/n/ac, and all drones are controlled from tablets or note PCs.

Table 3. The specifications of the prototype system

	Specification
Wireless	IEEE 802.11 a/b/g/n/ac
Resolution	4096×3072 px
Memory	8GB
Battery	Lithium Polymer1200mAh
CPU	Parrot P7 dual-core CPU Cortex9
OS	Linux
Sensors	3-axes magnetometer, 3-axes gyroscope, 3-axes accelerometer
Size	28×32×3.6 cm
Weight	400g

For the controls of the drones, the modules of Katarina Bebop Drone [17] and Selfie Dronie [18] are used for the implementations, and OpenCV [19] is introduced for the pattern recognition by drone's cameras. Those codes are mainly written by python.

Although the implementations of the collaborative safety flight are under developing, the proposed tracking functions by two drones are experimented for the effectiveness of the proposed methods. The experiment was held in the Fukuoka Institute of Technology, Japan as shown in Fig. 4.

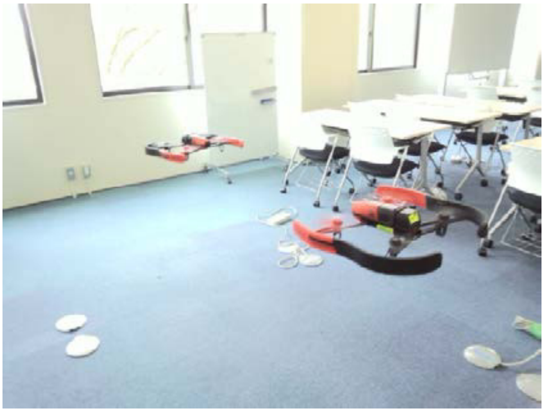


Fig. 4. The Experiments

The results show that the drone tracking worked under the slow movements, but there were some problems under the rapid movements. The considerable reason is firstly the resolution of the camera is smaller than the required functions. In other words, because the drones are small to identify for the tracking, the application did not work when each drone separate at certain distances. Probably, the larger objects will be needed for the future works.

Secondly, the drones usually move a specific direction because the balance of drones or adjustment of rotors, and the observed location is not so accurate. Therefore, the accurate controls toward the specific locations are difficult by the autonomous flights. In the experiments, the child drone was tried to adjust within the range of the camera view by manual controls, the controls were not well functioned.

5 The Conclusion and Future Study

Although there are rapid growth the drone's applications such as military uses, surveillance systems, and pesticide sprayings for harvests other than hobby usages in recent years, the problems of the accidents are also increasing conversely. Therefore, this paper discusses the collaborative object tracking methods by multi-drones. The proposed methods consist of the function of the motion tracking and the WiFi detections for each drone, and choose a proper level of emergent procedures in order to avoid the drone's accident in various circumstances.

In details, the emergent level is decided by the observed conditions of the image recognition of the other drones and the WiFi detections, and the proper emergent procedures are confirmed by the emergent level.

Then, the prototype of the proposed methods is introduced, and the experimental results are discussed for the subjects of the future works. The subjects include the additional functions such as the camera resolutions or the balance adjustments for accurate location point controls. Now we are planning for additional implementations and field experiments for the future works.

References

1. Nihon Keizai Shinbun.Inc: http://www.nikkei.com/article /DGXLASGM15H1A_10C15A5EAF000/
2. Sankei News: http://www.sankei.com/affairs/news/150422/afr1504220016-n1.html
3. Center for Computational Science & e-Systems: http://fukushima.jaea.go.jp/pdf/2014-0220.pdf
4. Asahi Shinbun DIGITAL: http://www.asahi.com/articles /ASG5W6HLKG5WTLVB00K.html
5. Bebop Drone: http://www.parrot.com/jp/products/bebop-drone/
6. Yoshida, J., Kashima, M., Sato, K., Watanabe, M.: Study on the automatic control of the flying robot by the aerial image analysis, FY2011 Electrical and Related Engineers Kyusyu Branch Union tournament, March 5,2013
7. Aida, Y., Fujisawa, Y., Suzuki, S., Iiduka, K., Kawamura, T., Ikeda, Y.: Guidance Control with Collision Avoidance for Multiple UAVs under Communication Restricted(Cooperation Control of MultiRobots), JSME annual Conference on Robotics and Mechatronics, May 24,2011
8. Bills, C., Chen, J., Saxena, A.: Autonomous MAV Flight in Indoor Environments using Single Image Perspective Cues, International Conference on Robotics and Automation - ICRA,pp5776-5783,2011

9. Ito, M., Hori, K.: Cooperative Space Searching Method by Multiple Small UAVs, The Japanese Society for Artificial Intelligence 27,January 3,2013
10. Takuya, S., Kenichi, M.: Development of Drone Pilot. NET, a SDK Supports Both Managed and Unmanaged Code for AR.Drone, PROCEEDING OF THE 2013 IEICE COMMUNICATIONS SOCIETY CONFERENCE,pp426,September 3,2013
11. Muller, M., Lupashin, S., D'Andrea, R.: Quadcopter Ball Juggling", IEEE/RS J International Conference on Intelligent Robots and Systems, pp5113-5120,September 25-30,2011
12. Nonami, K., Iwakura, D., Song, Y.: Teleoperation Technology of Multi-Rotor Electric Power Helicopter for Industrial Application, The Robotics Society of Japan 30(6),pp574-577,July 15,2012
13. Ayusawa, H., Nemoto, T., Iwakura, D., Nomura, K.: Autonomous Battery Exchanging System for Industrial Multi-rotor Helicopters, Dynamics and Design Conference 2013(13),August 25,2013
14. S. Suzuki, "Technical and Institutional Directions for the Safe Use of Small Unmanned Aircraft", The Robotics Society of Japan 34(1),pp24-27,January 2011
15. Takaishi, D., Nishiyama, H., Kato, N., Miura, R.: A Study on the Trajectory Control Algorithm in UAS Network, IEICE Technical Report 113(152),pp5-10,July 24-25,2013
16. Tanaka, N., Koji, H., Shibata, Y., Uchida, N.: Video Surveillance and Transmission System by Mesh Network Base on UAV, The 77th Annual Conference Proceeding, pp87-89,March 17,2015
17. Katarina: https://github.com/robotika/katarina
18. Selfie Dronnie: http://www.neurala.com/products/selfie-dronie-drone/
19. Opencv: http://opencv.jp/cookbook/

Performance Evaluation of a DTN Based Multi-hop Network for Disaster Information Transmission by Smart Devices

Shinya Kitada[1], Goshi Sato[1], Yoshitaka Shibata[1]
[1] Iwate Prefectural University, 152-52 Sugo, Takizawa, Iwate, Japan 020-0193

g231n007@s.iwate-pu.ac.jp
sato_g@ipu-office.iwate-pu.ac.jp
shibata@iwate-pu.ac.jp

Abstract. In Recent, there have been serious larger disasters, such as earthquakes tsumami, typhoons have around the world. Once disaster occurred, communication network infrastructure is often seriously damaged and network traffic is heavily congested. Particularly since the communication network infrastructure of mountainous and coast areas are not well developed, it is difficult to normally use the communication means once the disaster happened. For those cases, Delay / Disruptions Tolerant Network (DTN) can realize mobile devices communication without constructing infrastructure. In this paper, we propose a DTN based Multi-hop network for temporal network in the case of disaster and considered transmission disruption and transmission delay. In order to evaluate the usefulness of our proposed, we constructed an experimental prototype to by smart devices and evaluate the performance.

Keywords: *Disaster Information; DTN; Multi-hop Network*

1. Introduction

A large scale earthquakes have frequently occurred around the world. Once disaster occurred, communication network infrastructure is often seriously damaged and network traffic is heavily congested. Particularly since the communication network infrastructure of mountain areas are not well developed, it is difficult to normally use the communication means once the disaster happened. From the previous researches [1][2], it is clear that the wireless and mobile networks are very effective as disaster

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_59

information communication means. However, it is difficult to use cellular network in case of serious disasters because the cellular base stations are damaged by the secondary disaster such as tsunami and landslide. For this reason, the network traffic of cellular network which is concentrated to the base station is congested and cannot be transmitted to the destination.

In this paper, in order to resolve those problems, we propose a Epidemic DTN [3] based Multi-hop network corresponding to challenge network environment. In our system, the message issued from the source mobile node can be finally reached to the destination node or the gateway to Internet by multi-hopping on the intermediate mobile nodes. At the same time, we suppose vehicle-to-vehicle (or walker) communication where vehicle speed is not constant, eventually the vehicle can move out of communication range while transmitting data. So we use acceleration sensors mounted smart devices. In our system, by reducing redundant message transmission, battery energy can be saved.

2. Rerated Works

From the previous research [1][2], wireless cognitive networks, wireless adhoc networks, multi-hop networks are effective to realize robust and resilient network for disaster. In particular, multi-hop network can be deployed in the disaster areas without constructing network infrastructure. However, since the people move with mobile terminal in disaster area, it is difficult to constantly keep connection and transmission of information data. Delay / Disruptions Tolerant Network, simply DTN uses store and forward transport method. The mobile node stores the data to storage device if the link to the neighbor mobile node is not existed. If the link to the neighbor mobile node can be found, then the mobile node sends the stored data to the neighbor mobile nodes. There are many researches concerned with designing the functions of DTN protocol and running them on network simulators, but there is few researches concerned DTN protocol using actual network devices.

DTN is a relay transmission technology to achieve end-to-end transmission under poor and unstable communication environment. By using store-and-forward transmission method, reliable end-to-end communications can be possible. There are several works using such as sensor data and metadata to achieve higher performance [4] [5] [6]. In those works, DTN Protocol is developed on the network simulator and demonstrate in variety of parameter environments.

On designing DTN protocol, it is important to consider the following issues such as unstable data communication period between mobile nodes, amount of data transmission to send/receive and limited battery energy resource. In addition, the development of DTN protocol on the actual prototype system depends on network hardware and OS specification. Under those conditions, protocol overhead in limited environment has to be reduced in the system design. There are only a few related works with DTN protocol using the actual prototype system.

In the work [7], the proposed function of the system combining MANET and DTN can switch between MANET and DTN as necessary to avoid network resources consumption. DTN MapEx [8] integrates a DTN protocol and distributed computing function and generates maps for disaster information to share data for decision-making.

3. The proposed system

In this section, we propose an Epidemic DTN [3] based multi-hop network which can realize vehicle-to-walker communication using smart devices. On the occurrence of disasters, since many residents evacuate with smart devices, the limitation of battery energy resource should be considered. We manage the node list and acceleration sensor data to reduce the redundant transmission and communication.

Figure 1 shows our proposed network system configuration at the time of disaster where inland, coastal and mountains areas are assumed in the following conditions; 1) cellular network is congested or cannot be functioned in the disaster areas, 2) most of the residents move carrying with smart devices move in disaster area, 3) the relay nodes are deployed each area just after occurrence of disaster.

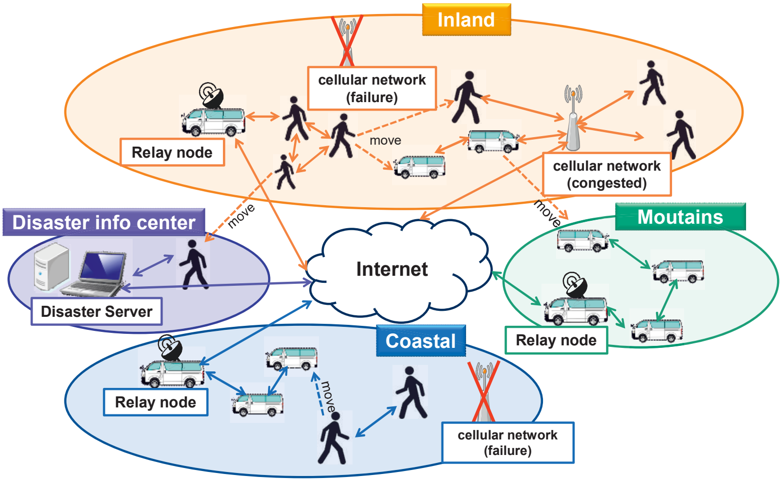


Fig. 1 System Configuration

On the above conditions, DTN based Multi-hop network is configured using the smart devices by residents. The data from a smart device as mobile node are transmitted to the neighbor nodes by DTN protocol. This transmission is repeated until arriving at the destination node or the relay node as a network gateway to Internet. The relay node is assumed the following conditions; 1) the relay node by wireless cognitive access network with long distance can connect to Internet, 2) the relay node has a mobile server that collects the received data from other mobile nodes. We regard the transmitted data as *Message* in our proposed network.

A Message consists of sender address, destination address, message ID, disaster information, type of *Message*, *Message* priority, Date and Message Payload as shown in Figure 2. We use the specific device name as *Sender Address* and *Destination Address*.

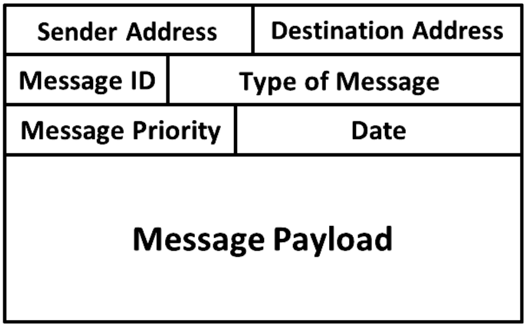


Fig. 2 Message Structure

Disaster victims create *Message* and decide urgency using *Message* priority. And *Message* is transmitted based on this priority. The relay node can send/receive those messages to/from the disaster information center in the other areas through Internet. On the other hand, the disaster information center is assumed in the following conditions; 1) the disaster information center is allocated at the place where there is no damage of tsunami and landslide, 2) an access to Internet is always possible. Thus, in our proposed system, the communication between the mobile node and the relay node, and the mobile modes at the different areas can be performed in the same manner.

Figure 3 shows our proposed conceptual network model. When the destination node exists in the same area as the source node, the source node spreads *Messages* to the neighbour nodes, n1, n2 and n3. The received neighbour nodes spread the received *Messages* to the other neighbour nodes. By repeating this process, the *Messages* eventually arrive at the destination node.

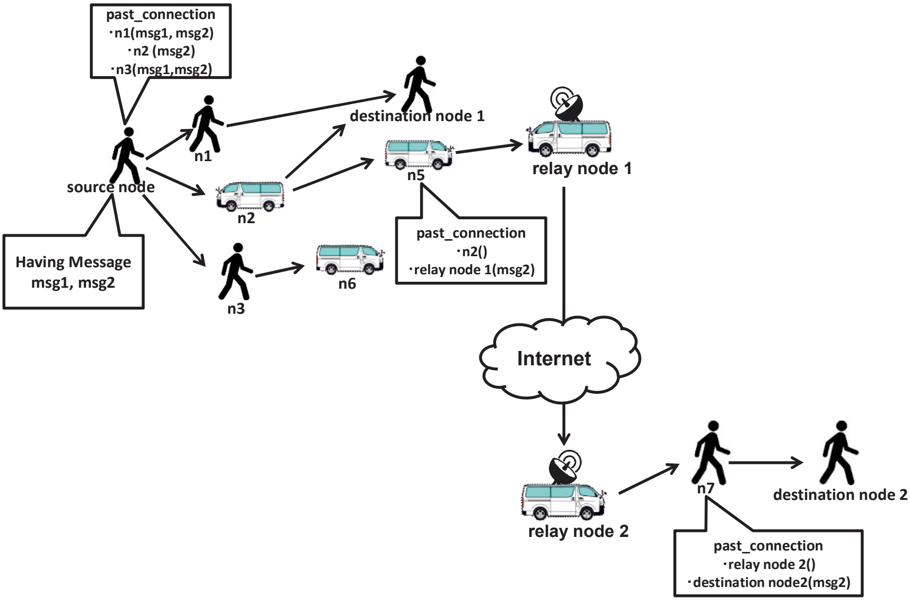


Fig. 3 Conceptual Network Model

In the case that the source node, n1 and n3 are walkers, their moving speeds are slow. Thus the source node transmits the *Message* and can expect that all of the Messages in his device can arrive at those neighbour nodes. If the source node has many *Messages*, it transmits higher priority *Message* first. In the case that the node n2 is a vehicle and its moving speed is faster, the source node may not transmit all of the *Messages* in his device storage and may derive to link down. To prevent the link down while transmission of *Messages*, the source node selects proper *Message* and transmits based on the relative moving speed between the source node and n2, and *Messages* priority. Thus proper *Message* can arrive at the destination node 1 through node n2.

On the other hand, when we need to send *Messages* to the destination node 2 across Internet, the relay node 1 performs the role of temporary destination as a gateway to Internet. The relay node 1, after receiving *Message* from node n5, transmits to relay node 2 through Internet. Then relay node 2 transmits to node n7. Thus the *Message* from the source node can be finally arrived at the destination node 2 across Internet.

4. Network protocol

The network routing protocol of our proposed system is based on flooding based DTN method such as Epidemic routing [3]. In our routing protocol, each node manages connection nodes using *neighbour_list* keeps neighbor nodes and *past_connection* list which keeps past connection state. First, the transmission node and neighbour nodes exchange Probe Request and Probe Response while changing *Listen* mode and *Search* mode mutually. If the mode is *Search*, the node sends *Probe Request*. Also if the mode is *Listen*, the node waits for *Probe Request* and replies *Probe Response*. When the transmission node received *Probe Response*, then adds the responded node to *neighbor_list*. Next, the transmission node checks *past_connection_list* whether the list is empty or not**.**

Figure 4 shows an example of *past_connection* and *neighbor_list*. In this example, the transmission node has *Message*1 (msg1) and msg2 to the neighbor n1. First, the transmission node checks the neighbor node1 (n1) which is the top of *neighbor_list* whether there is n1 in *past_connection*. In this example, since n1 exists in *past_connection*, the transmission node passes n1 and goes to n2 which is the second of *neighbor_list*. In this case, since n2 exists in *past_connection*, but the transmission node had not transmitted all *Message* in the storage. So the transmission node decides to connect n2 and transmits msg2. If all of the nodes in *neighbour_list* exist in *past_connection* and the transmission node had transmitted all *Message* to all nodes in *past_connection*, then the transmission node continues to discover new neighbor node to update *neighbour_list*. This step is repeated for the all of the existing nodes.

```
Algorithm Managing list to routing
If past_connection is empty;
    The source node sends connection request to the first node in
    neighbor_list.
    Else;
        The source node compares neighbor_list with the
        past_connection and focuses on the i-th node in neighbor_list.

If node i in neighbour_list  ∉ past_connection;
    The source node sends connection request to the node i
        Else if the source node had not  transmitted all Messages in
        the storage;
            The source node sends connection request to the node
        Else;
            When the node i is existed and the source node had transmitted
            all all of Messages,
            then tries the next node i+1.
```

Fig. 4 Compare Lists Example

Next, both nodes select their own acceleration data form acceleration and gyro sensors, and exchange those data. Both nodes determines the proper *Message* from the storage based on the relative speed and message propriety. Finally, the transmission node transmits the selected *Message*. Then node n2 repeats the message transmission to other nodes in the same manner.

Thus, our algorithm uses the *neighbor list* to manage the neighbor nodes, the *past_connection* to manage the *messages* whether those are transmitted or not, moreover exchanging acceleration and gyro sensor data before transmission data. They lead to save duplication of data and reduce battery energy consumption by the unnecessary communication.

5. Prototype system

In order to verify the effect of our proposed system, we construct a prototype system using commercially available smart devices as shown in Figure 5 and evaluate its functional and performance. The prototype is consisted of multiple different devices including smart terminals for human carrying communication devices, relay node with smart device and Mobile server and a disaster server as a destination server.

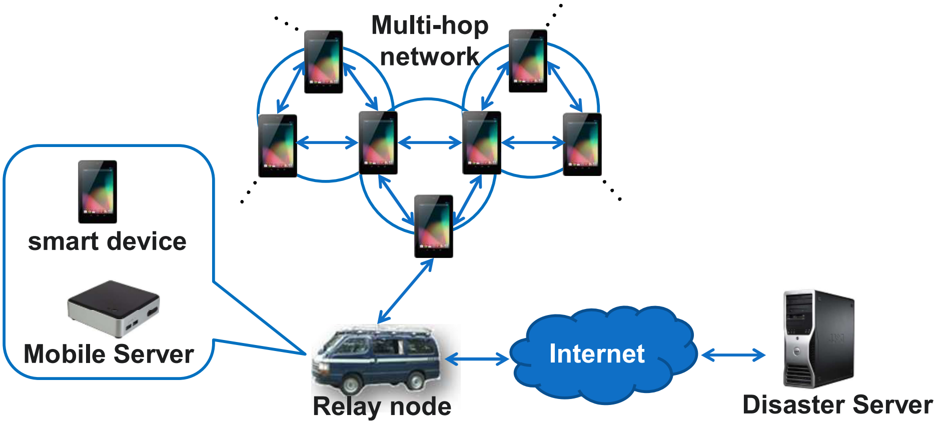


Fig. 5 Prototype System

Table 1 shows the details of hardware system of our prototype system. We develop our DTN function module and routing function module for as an application of the Android OS based smart terminal using Android SDK. We construct Multi-hop network by those smart terminals using WLAN IEEE 802.11n which is mounted as common wireless network devices. For Mobile server and Disaster server, mobile PCs operated by Ubuntu OS are introduced. JAVA and MySQL are used for software system development of our architecture. All of the *Message*s are managed by SQLite in the message storage. Figure 6 shows a test Graphical User Interface (GUI) of our application. This GUI shows a view at the time of the first login on the application of the smart device.

Table 1 Hardware Specification of Prototype

Smart Devices	Nexus 7 (2013) OS: Android 5.0 CPU:APQ8064 QuadCore1.5GHz MEM: RAM 2GB Develop. SDK. : Android SDK NIC: IEEE 802.11n
Mobile Server	OS: Ubuntu 14.04 CPU: Core i3 4010U MEM: RAM 4GB Develop. SDK : JDK7
Disaster Server	OS: Ubuntu 14.04 CPU: Core 2 Extreme MEM: RAM 4GB Develop. SDK: JDK7

6. Performance Evaluation and Discussions

The preliminary test is executed to evaluate the performance of the basic application system on the smart terminals. As performance index, the throughput and the packet loss rate are measured at the outdoor of our university by changing the transmission distance between the two smart terminals. As network parameter values, packet is 1KB, Message size is 1MB, Message transmission protocol is TCP/IP and DTN routing is Epidemic routing method. The performance test was executed 5 times by changing the transmission distance between smart devices

Figure 6 shows the result of the performance of the throughput and the packet loss rate for the basic application system between the smart terminals. As can be seen, at the transmission distance from 0~40m, both the throughput and the packet loss rate are almost stable and constant, about 7 Mbits/sec and 10 %, respectively. At the transmission distance from 40~80m, the packet loss rate gradually increases and throughput decreases, and finally all of the packets are almost lost and throughput reached to 0. From this result, it is founded that the possible transmission distance is within 80 m and the throughput is under 7 Mbits/sec. Therefore, in the actual disaster situation, if the people are located outdoor with smart terminals within 0~80m area, the message transmission can be possible in the area without DTN function. If people are mutually far away, messages are temporally stored in his storage by DTN function. Those messages can be delivered to the neighbor smart terminal by mutually moving.

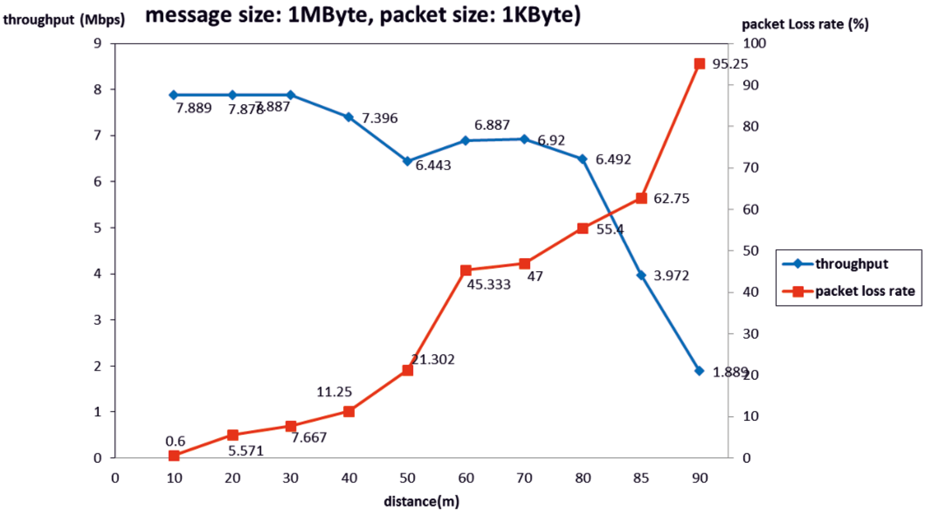


Fig. 6 Throughput and Packet Loss Rate

Next, we examined the performance evaluation of multi-hop network environment. Table 3 shows summery of the test network parameters and their values for the performance evaluation. The end-to-end transmission time form the source smart terminal to the destination disaster server was measured for three different message sizes when the number of the hops is increased. Each smart terminal node was placed every 3 m at distance. The test was repeated 5 times and the observed end-to-end transmission time were to derive the average values.

Figure 7 shows the result of end-to-end transmission time form the source smart terminal to the destination disaster server for three different message sizes when the number of the hops is increased from 1 to 5 depending on the size of the messages. Therefore by considering this result, it is possible to estimate the end-to-end transmission time even when the number of the hops increase more than 5.

On the other hand, the differences of the end-to-end transmission times for three message sizes are small. This reason is that the packet processing, such as packet connection establishment on TCP on the smart terminal node is dominant for the end-to-end transmission time compared with the number of the packet transmission. As result, the message transmission by TCP should be improved by using more simple reliable transmission protocol or introducing the larger packet size more than 1 Kbyte to relatively reduce the influence of the protocol processing time.

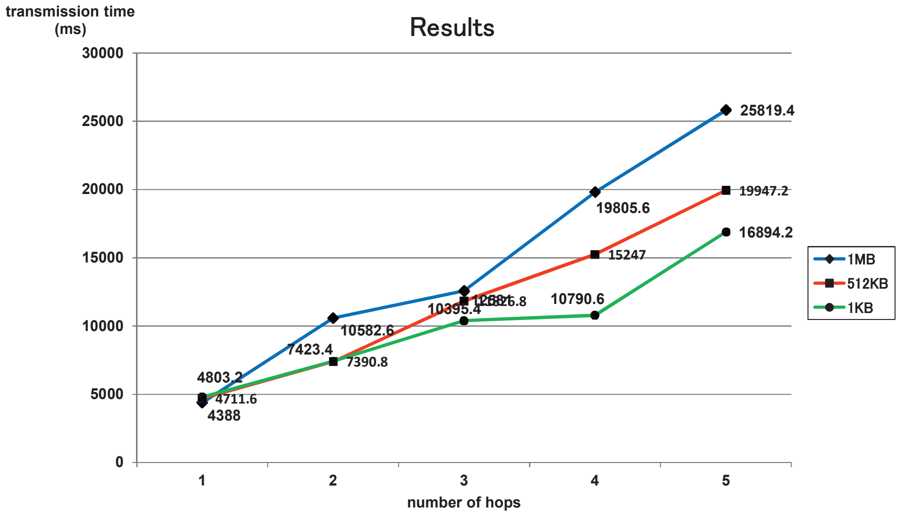


Fig. 7 Result of Measure Test

7. Conclusions

In this paper, we proposed an effective message transmission method by DTN based Multi-hop network for the case of infrastructure failure at time of disaster. It is possible to consider the message transmission and communication interruption using DTN protocol function. Also we used the neighbor list and the *past_connection_list* to manage routing, and acceleration sensor data to reduce link down while transmitting data. Furthermore communication between disaster areas also becomes possible by using relay node capable of Internet access. In order to verify effect of the proposed system, we built a prototype system and experimented the performance evaluation. Through the measurement test with actual devices, we found that the current prototype provide reasonable performance but need to improve message transmission time when the size of the multi-hop is larger.

As our future works, we will examine the message transmission on vehicle-to-vehicle communication. Then we will design more effective message transmission method using acceleration sensor data. Next, we plan to improve

DTN protocol which can ensure to reach to the relay node, and attempt the experimental test with various metrics (duplicated data ratio, delivery ratio etc.) while we compare existing protocols. Also since performance changes due to the mobility of nodes, we will work for an experimental test for dynamic change of the nodes.

Acknowledgement The research was supported by SCOPE (Strategic Information and Communications R&D Promotion Program) Grant Number 142302010 by Ministry of Internal Affairs and Communications in Japan.

References

[1] Goshi Sato, Noriki Uchida, and Yoshitaka Shibata "Performance Evaluation of Software Defined and Cognitive Wireless Network Based Disaster Resilient System", The 29th IEEE International Conference on Advanced Information Networking and Applications, (AINA2015), 741-746, March 2015.

[2] Yoshitaka Shibata, Noriki Uchida, Norio Shiratori,"Analysis and Proposal of Disaster Information Network from Experience of the Great East Japan Earthquake," IEEE Communications Magazine, Vol. 52, No. 3, pp.44-48, March 2014.

[3] Amin Vahdat, David Becker , "Epidemic Routing for Partially-Connected Ad Hoc Networks," Technical Report CS-2000-06, Department of Computer Science, Duke University, Apr., 2000.

[4] R. Ramanathan, R. Hansen, P. Basu, R. Rosales-Hain, and R. Krishnan, "Prioritized epidemic routing for opportunistic networks," San Juan, Puerto rico, pp. 62-66, 2007.

[5] Chengping Tian, Linlin Ci, Bin Cheng, Xuanya Li "A 3D Location-Based Energy Aware Routing Protocol in Delay Tolerant Networks," Dependable, Autonomic and Secure Computing (DASC 2014), pp. 485-490, Aug., 2014.

[6] Jovilyn Therese B. Fajardo, Keiichi Yasumoto, Naoki Shibata, Weihua Sun, Minoru Ito "DTN-Based Data Aggregation for Timely Information Collection in Disaster Areas," Wireless and Mobile Computing, Networking and Communications (WiMob 2012), pp. 333-340, Oct., 2012.

[7] Masaya Ito, Hiroki Nishiyama, and Nei Kato "A Novel Communication Mode Selection Technique for DTN over MANET Architecture," International Conference on Computing, Networking and Communications (ICNC 2014), Feb., 2014.

[8] Edgar Marko Trono, Yutaka Arakawa, Morihiko Tamai, Keiichi Yasumoto "DTN MapEx: Disaster Area Mapping through Distributed Computing over a Delay Tolerant Network," Mobile Computing and Ubiquitous Networking (ICMU), pp. 179-184, Aug., 2015.

OpenFlow Mesh for Metabolic Computing

Minoru Uehara[1]
[1] Dept. of Info. Sci. and Arts, Toyo Univ., 2100 Kujirai, Kawagoe, Saitama, 350-8585 Japan
uehara@toyo.jp

Abstract. Renewable computing systems have recently become important. We propose a metabolic architecture that is suitable for renewable systems. Here, we consider a metabolic architecture based system that can exchange all of its elements dynamically like a multicellular organism. It can maintain homeostasis and adapt to environmental changes. However, there are many issues when realizing metabolic systems. In this paper, we propose an OpenFlow mesh based metabolic architecture and discuss its potential realization. An OpenFlow mesh is mesh network based on OpenFlow. In this paper, we realize an OpenFlow mesh using an ARM SBC (single board computer) and evaluate its renewability. Then, we show that an OpenFlow mesh can be used to realize metabolic systems.

1 Introduction

In the future, our society must be sustainable. To realize this sustainable society, all social systems related to social development, the economy, and the environment should be sustainable. Therefore, we believe that information systems should also be sustainable.

Academics have recently started to focus on sustainable computing. Sustainable computing is defined as a computer science approach to realize a sustainable society. For example, it considers life-cycle management, total cost of ownership, and energy-awareness.

In previous work [1], we proposed renewable computing. Renewable computing is defined as recoverable and evolutionary sustainable computing. Even if a society is sustainable, if it is not developed is not optimal. Developments occur during revolutions. Throughout history, society has been developed by several industrial revolutions. Unfortunately, revolutions are destructive. Societies have recovered from several destructive periods such as economic crises, war, and natural disasters. The dynamic nature of renewable computing is essential to sustainable computing.

We propose metabolic computing and its architecture for renewable computing [1]. In metabolic architecture, homeostasis is maintained by exchanging elements over a period. Generally, if there are no initial faults, new parts are very reliable. Reliability decreases over time and metabolic architecture helps to alleviate decreasing reliability. In metabolic architecture, parts are periodically checked and exchanged. They are

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_60

recycled if they fail, or are reused to reduce costs. Additionally, metabolic architecture has minimal requirements and is dependable.

We are developing a renewable cloud based on metabolic architecture. In [1], we designed the first version of metabolic architecture. However, it contained many technical parts, so it was hard to implement. In [2], we simplified the design to make it easier to implement. However, the performance also drastically decreased. In this paper, we propose a practical metabolic architecture using an OpenFlow mesh with the objective of producing a reasonable performance and an easy implementation.

An important issue in a cloud environment is power consumption. In the future, renewable clouds will only be run using renewable energy. Our objective is to construct a cloud environment using ARM processors to maximize energy efficiency. In this paper, we used a Raspberry Pi as an ARM SBC (single board computer), which is used as a node of the cloud. However, a Raspberry Pi is not sufficient powerful for use as a cloud node. Therefore, in the future we will develop an ARM SBC that is a suitable cloud node.

Generally, we must virtualize the hardware and network to realize a cloud infrastructure. ARM lacks a hardware virtualization mechanism such as Intel VT (Virtualization Technology), so it is hard to virtualize hardware. However, we can use a LXC (Linux container) instead of hardware virtualization. Alternatively, the network can be virtualized as a SDN (software defined network) such as OpenFlow. In [3], we developed an OpenFlow mesh to realize the network infrastructure of an ARM Linux based cloud. In this paper, we discuss the metabolic architecture based on an OpenFlow mesh.

The remainder of this paper is organized as follows. We introduce the metabolic computing model in Section 2, and propose a propose metabolic architecture based on the OpenFlow mesh in Section 3. Section 4 contains our conclusions.

2 Metabolic Computing

In previous work [1], we proposed the metabolic computing model as an architecture for renewable computing systems. The metabolic computing model is the basis of the metabolic ring. In that paper, we defined the concept of metabolic computing. Here, we describe the concept of the metabolic computing model illustrated in Fig. 1. This model contains five elements: metaboloid, slot, power queue, recycle unit, and delivery system.

Metaboloid: A metaboloid is a small recyclable computer that is passive. For example, it cannot move by itself nor connect to another metaboloid by itself. The mechanism is simple and the metaboloid can be recycled easily. Metaboloids are passive, otherwise failed metaboloids would move and be hard to find. The shape of each metaboloid is a cube, and each surface has a communication terminal. When a metaboloid is moved its direction may change, so it must be able to recognize direction. Metaboloids communicate with each other through contact with neighboring communication terminals. A metaboloid has no power unit, and instead the power is provided by the power queue.

Slot: This is a fixture that holds a set of metaboloids. It also supports the movement and connection of metaboloids. A slot is not an essential part of the metabolic computing model, but it is useful. A slot can be easily recycled because it is simpler than a metaboloid.

Power Queue (PQ): This is the computational environment where the metaboloids compute. It also supplies power to the metaboloids. The PQ is a queue of slots. A slot that has remained for a certain period is removed from the PQ, and a new slot is added. We call this action a "shift." In this way, metabolism in this model is represented by the behavior of the PQ. The PQ may also be regarded as a 2D array of metaboloids. The position of a metaboloid in the PQ is determined by its row and column. This position changes due to metabolism. Furthermore, the PQ communicates with metaboloids to enable input and output processes. The PQ is a complex part of this model, so it is difficult to recycle.

Recycle Unit (RU): RU reuses metaboloids and slots if possible, and recycles them otherwise. If the RU cannot recycle some resource then it will be rejected, because the resource cannot be used. There are two kinds of RU: slot RU and metaboloid RU. Slot RUs recycle slots and metaboloid RUs recycle metaboloids. An RU cannot recycle itself.

Delivery System (DS): The DS connects all elements of the metabolic computing model to each other. It also delivers parts to the appropriate places. For example, the DS collects a number of metaboloids from the metaboloid RU and packs them into a slot. The DS adds a slot to the PQ if there is space. Additionally, the DS opens a slot that is removed from the PQ and then splits it into the slot and the metaboloids. Finally, a DS delivers slots and metaboloids to the appropriate RUs. It is difficult to recycle a DS.

It is hard to realize RUs and DSs because they are not purely computer based systems. In [4], we mainly focused on metaboloids, slots, and PQs. In this paper, we also consider these elements.

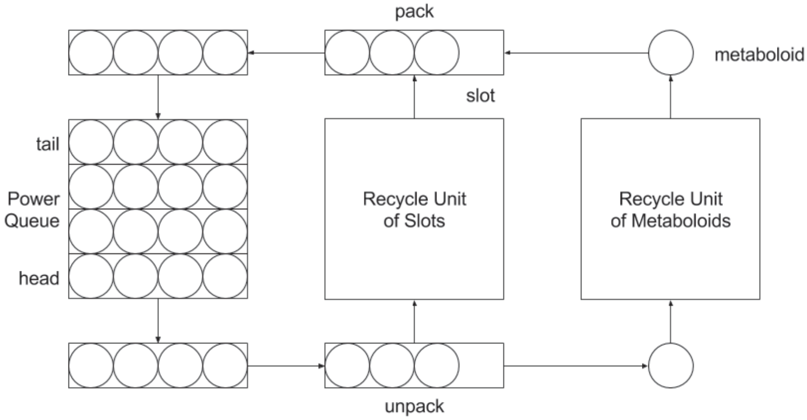


Fig. 1. Overview of the metabolic computing model

3 Design of OpenFlow Mesh based Metabolic Architecture

3.1 Requirements

The requirements of this research are as follows.

Low power: The PQs and metaboloids consume the most power, except the RUs and DSs. We can reduce the power consumption of the metaboloid using an ARM SBC. The PQ consumes a lot of power because of its shifting mechanism, which moves many mechanical slots. Here, shifting means to individually move all of the PQ slots. This shifting behavior can be represented by the following pseudocode.

```
Pull Slot[0]
Slot[i] := Slot[i+1] (i=0..n-2)
Push Slot[n-1]
```

Here, if *m* is the weight of a slot and *v* is the shifting speed, the energy required to shift *n* slots is nmv^2. If we have a shift-less PQ, the required energy is mv^2, but we require random access to the slots, which is achieved using a robot arm. The energy required to move a robot arm is smaller than the energy required by an entire PQ with shifting.

Efficiency: We must migrate tasks running on metaboloids to a slot that will be pulled in a shifting PQ. Additionally, we need the ability to migrate task to balance the load. Generally, task migration generates a heavy load. However, in a shift-less PQ, we do not need to migrate any tasks.

Easy Implementation: A slot is a container that contains more than one metaboloid. If a unit of metaboloids in a PQ is a slot, we need packing/unpacking processes. Currently, these are manual processes, so, in this paper, we only consider slots that contain one metaboloid. This means that a slot is not used or it is equivalent to a metaboloid.

3.2 OpenFlow 2+1D Mesh

An OpenFlow mesh is a 2+1D mesh of an OpenFlow Switch (SW). Figure 2 illustrates an OpenFlow mesh. There are two types of switches: SW_{xy} and SW_z. A 1D mesh consists of one or more of SW_z. A 2D mesh consists of 1D meshes, and is organized with one or more SW_{xy}. LXC containers run in SW_z. The X, Y, and Z axes represent the width, depth, and height, respectively. Here, if the numbers of SWs in the X, Y, and Z directions are *w*, *d*, and *h*, respectively, there are *wd* SW_{xy}s and *wd(h-1)* SW_zs. The number of LXC containers is proportional to the number of SW_zs. However, the number of SW_zs is different for every 1D mesh. There are SW_{xy} in the *Z=0* plane. SW_zs are only connected to each other in the Z-direction. SW_z is connected to SW_{xy} in *Z=0* plane. Therefore, the maximum delay will be proportional to *w+d+2h*.

Figure 2 also shows the structure of the OpenFlow mesh SW. The role of the SW is split into SW_{xy} and SW_z. However, the structures of the two types are the same. SW consists of an ARM SBC (Raspberry Pi) and 6 USB NICs. Each port is connected to a

USB NIC. Additionally, the SW is connected to the OpenFlow controller by a wireless network. In SW_{xy}, 5 of 6 ports are connected to the N(orth), S(outh), W(est), E(ast), and U(p) ports. In Figure 3, if the coordinate of SW_{xy} in the *z=0* plane is (x, y, 0), the coordinates of SW_n, SW_s, SW_w, SW_e, and SW_u are (x, y-1, 0), (x, y+1, 0), (x-1, y, 0), (x+1, y, 0), and (x, y, 1), respectively. SW_u is not shown in Figure 3, but it overlaps on SW_{xy}. In SW_z, 2 of 6 ports are connected to the U(p) and D(own) ports. Additionally, the local port is connected to the LXC Bridge.

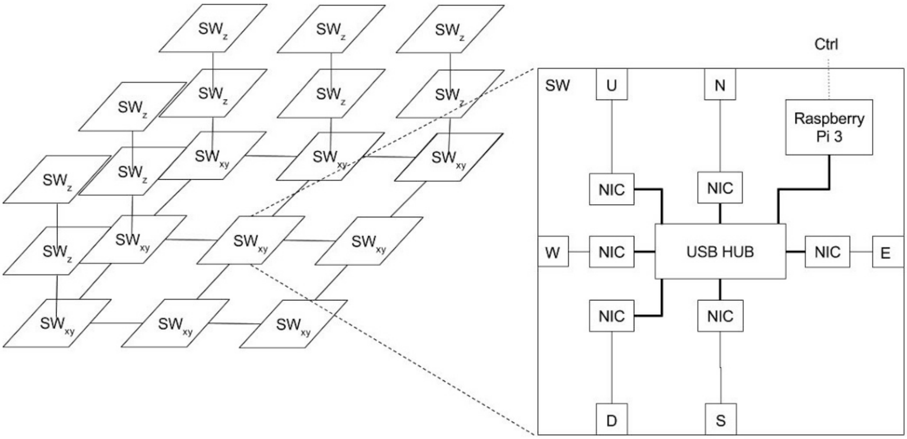


Fig. 2. OpenFlow 2+1D Mesh and Switch

3.3 OpenFlow Mesh based PQ

The OpenFlow 2+1D mesh is organized as a 2D mesh of 1D meshes.

The 1D meshes are organized as stacked SWs. Figure 3 shows the stacking SWs. To ensure a fail-safe design, stacking is passive and mechanical. This means that a SW does not move by autonomously. Guide poles are used to fix the SWs so that they are robust during disasters such as Earthquakes. When a SW is stacked at the top, it is connected to the top SW by its own weight. There are no wires between SWs. They are directly connected to each other with connectors. Here, W, S, D, V, U, N, and E in Figure 4 represent connectors to the west port, south port, down port, V (Power supply), up port, north port, and east port, respectively. There is a base at the bottom of the stack. A base is directly connected to neighboring bases with wires. We assume that the failure rate of the wires is very small.

In the 1D meshes, metabolism is achieved as follows. The elevator is set on the base of a stack. It lifts all the SWs in the stack. Old SWs are removed as follows. The elevator lifts an old SW up to the top of the guide pole. The robot arm removes the SW from the top of the stack and moves it to the DS. This is repeated until there are no SW in the stack. New SWs are installed in the following ways. The position (*x*, *y*) at which new SWs are installed is determined by the height of the elevator. After the DS moves a new SW, the robot arm installs it to the top of the stack. The elevator lifts it down it after a SW is installed on the top of the stack. This process repeats until the stack is filled.

A 2D mesh is a mesh network of 1D meshes. It is organized by all the SW_{xy}s at the bottom of the stacks. The metabolism of the 2D mesh is carried out in the following ways. The order is statically scheduled. For example, if SW_{xy} is located at (x, y) in the 2D mesh, its order is $(wy+x)$-th. Here, w is the width of the 2D mesh.

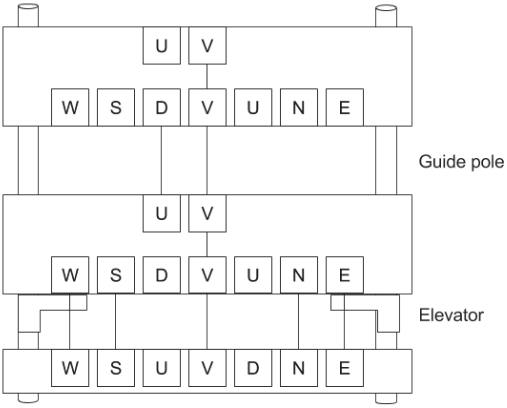


Fig. 3. Stacking in 1D Mesh

3.4 Routing

The topology of the mesh network changes when an elemental SW fails, is removed, or is installed. The change in topology can be detected using LLDP (link layer discovery protocol). In an OpenFlow mesh, all LLDP packets are sent to the OpenFlow controller. So, the OpenFlow controller manages the neighbor matrix of the OpenFlow mesh and computes the network model. In this way, the OpenFlow mesh can be monitored during run time.

OpenFlow virtualizes the connection between any two switches, SW_1 and SW_2. Here, assume that the coordinates of SW_1 and SW_2 are (x_1, y_1, z_1) and (x_2, y_2, z_2), respectively. If $x_1=x_2=x$ and $y_1=y_2=y$ then a virtual flow is created on the path (x, y, z_1) - (x, y, z_2) in the 1D mesh (x, y). Otherwise, a virtual flow is created on the path (x_1, y_1, z_1) - $(x_1, y_1, 0)$ - $(x_2, y_2, 0)$ - (x_2, y_2, z_2), which is connected between different 1D meshes (x_1, y_1) and (x_2, y_2). Here, there is more than one path connecting $(x_1, y_1, 0)$ and $(x_2, y_2, 0)$.

There are many types of routing methods for a 2D mesh. For example, in surrounding XY DOR (dimension order routing)[4], the route between $(x_1, y_1, 0)$ - $(x_2, y_2, 0)$ is statically determined as the path $(x_1, y_1, 0)$ - $(x_2, y_1, 0)$ - $(x_2, y_2, 0)$. XY DOR is not fault tolerant, but surrounding XY DOR is. These routing methods are used in NoC (Network on Chips), and only use local connection information. However, in an OpenFlow mesh, the OpenFlow controller has global connection information, so we can use Dijkstra's shortest path algorithm.

3.5 Extending

The advantage of using a cloud environment is its scalability. There are two types of scaling methods: scale-up and scale-out. In scale-up, we upgrade each element. In scale-out, we increase the number of elements.

Scale-up methods for an OpenFlow mesh are (1) improving the SoC (system on chip) of the ARM SBC, (2) increasing the number of ARM SBC in a metaboloid, and (3) increasing the number of ARM SBC in a slot. However, (3) is not appropriate for the proposed architecture.

In (2), additional ARM SBCs are connected to a SW. Local ports are assigned to all additional ARM SBCs in the SW.

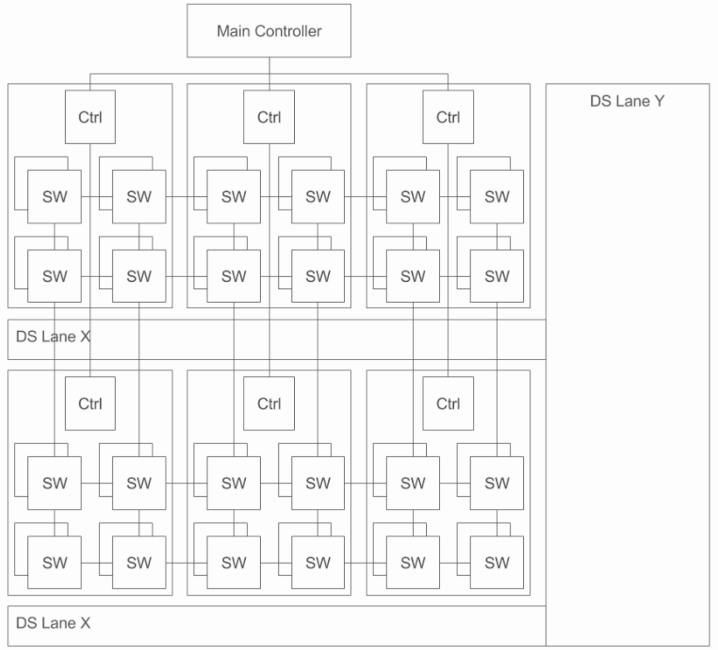


Fig. 4. Large-scale OpenFlow Mesh

To scale-out an OpenFlow mesh, multiple OpenFlow meshes are arrayed. Small scale meshes are suited to robot arms and elevators, and wireless communication between SWs and the OpenFlow controller. A rack is the suitable scale. Figure 4 shows the structure of a large-scale OpenFlow mesh. In this system, an OpenFlow mesh is organized in each rack. Each controller (Ctrl) is a bridge connected to the main controller via a wire. A mesh is connected its neighboring meshes through bases. The width of DS Lane X is equal to the width of a metaboloid. The width of DS Lane Y is equal to the width of a rack.

Here, we assume that a metaboloid is the same size as an A4 file box (100 mm × 320 mm × 250 mm). The size of 42U rack is approximately 600 mm × 1000 mm × 2088 mm. So, a rack can contain 5 × 3 × 8 = 120 metaboloids. A metaboloid can contain 1 × 2 × 8 = 16 ARM SBCs (Raspberry Pi). At this time, the accommodation rate is 58%. This is sufficiently large to contain the other parts of a metaboloid. Therefore, a rack can theoretically contain 1920 ARM SBCs.

A Raspberry Pi weighs 45 g, so 20 weigh less than 2 kg. Even if the weight of the other parts is estimated to be 3 kg, the total weight is less than 5 kg and it can be moved by the considered robot arm. A different robot arm can move a 40-kg object, so it can be used as the elevator that lifts an entire 1D mesh including 8 SWs. If we cannot prepare an elevator for each 1D mesh, the robot arm can individually move all the 1D meshes. However, this reduces availability because the unit of metabolism is a rack.

4 Conclusions

In this paper, we described a metabolic architecture based on an OpenFlow mesh. The proposed architecture requires less power, is very efficient, and is easily implemented. Furthermore, it can achieve a dense, autonomous, and maintenance free system. These features are important for cloud infrastructures, so this architecture is useful.

In the future, we will: design an inter SW_z wireless communication method such as short-range wireless and visible light communication; package metaboloids, develop a small elevator that can lift heavy weights; design a three-axis robot arm that can install and remove metaboloids; and propose a cart robot that acts as a DS.

References

1. Minoru Uehara: "Metabolic Computing: Toward Truly Renewable Systems", IGI Global, International Journal of Distributed Systems and Technologies (IJDST), 3(3), pp.27-39, (July-September 2012)
2. Minoru Uehara, Hideki Mori: "Metabolic Ring: Tape based Renewable System", In Proc. of the 15th International Symposium on Multimedia Network Systems and Applications (MNSA2013) in conjunction with the 8th International Conference on Broadband, Wireless Computing, Communication and Applications (BWCCA2013), pp.453-458, (Compiegne, France, 2013.10.28-30)
3. Suguru Yasui, Minoru Uehara: "OpenFlow Mesh based Virtual Crossbar Network", In Proc. of the 10th International Conference on Complex, Intelligent and Software Intensive Systems (CISIS2016), pp.1-6, (Fukuoka, Japan, 6-8 July 2016)
4. C. Bobda, A. Ahmadinia, M. Majer, J. Teich, S. Fekete, J. van der Veen: DyNoC: A Dynamic Infrastructure for Communication in Dynamically Reconfigurable Devices. International Conference on Field Programmable Logic and Applications, 24–26 August 2005, pages: 153–158.

Part III
The 9-th International Workshop on Next Generation of Wireless and Mobile Networks (NGWMN-2016)

Performance Evaluation of an AmI Testbed for Improving QoL: Evaluation Using Clustering Approach Considering Parallel Processing

Ryoichiro Obukata, Tetsuya Oda, Donald Elmazi, Makoto Ikeda and Leonard Barolli

Abstract Ambient intelligence (AmI) deals with a new world of ubiquitous computing devices, where physical environments interact intelligently and unobtrusively with people. AmI environments can be diverse, such as homes, offices, meeting rooms, schools, hospitals, control centers, vehicles, tourist attractions, stores, sports facilities, and music devices. In this paper, we present the design and implementation of a testbed for AmI using Raspberry Pi mounted on Raspbian OS. We analyze the performance of *k*-means clustering algorithm considering sensing data. For evaluation we considered respiratory rate and heart rate metrics. We speeded up the *k*-means clustering algorithm by using parallel processing.

1 Introduction

Ambient Intelligence (AmI) is the vision that technology will become invisible, embedded in our natural surroundings, present whenever we need it, enabled by simple and effortless interactions, attuned to all our senses, adaptive to users and context and autonomously acting [1]. High quality information and content must be available to any user, anywhere, at any time, and on any device.

In order that AmI becomes a reality, it should completely envelope humans, without constraining them. Distributed embedded systems for AmI are going to change the way we design embedded systems, in general, as well as the way we think about

Ryoichiro Obukata and Donald Elmazi
Graduate School of Engineering, Fukuoka Institute of Technology (FIT), 3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811–0295, Japan, e-mail: obukenkyuu@gmail.com, donald.elmazi@gmail.com

Tetsuya Oda, Makoto Ikeda and Leonard Barolli
Department of Information and Communication Engineering, Fukuoka Institute of Technology (FIT), 3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811–0295, Japan, e-mail: oda.tetsuya.fit@gmail.com, makoto.ikd@acm.org, barolli@fit.ac.jp

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_61

such systems. But, more importantly, they will have a great impact on the way we live. Applications ranging from safe driving systems, smart buildings and home security, smart fabrics or e-textiles, to manufacturing systems and rescue and recovery operations in hostile environments, are poised to become part of society and human lives.

There are a lot of works done on testbed for AmI. In [2], the authors present a simulation environment that offers a library of Networked Control Systems (NCS) blocks. Thus, the constraints can be considered and integrated in the design process. They describe a real process, an inverted pendulum, which is automated based on Mica nodes. These nodes were designed especially for AmI purposes. This real NCS serves as a challenging benchmark for proving the AmI suitability of the controllers.

In [3], the authors present the development of an adaptive embedded agent, based on a hybrid PCA-NFS scheme, able to perform true real-time control of AmI environments in the long term. The proposed architecture is a single-chip HW/SW architecture. It consists of a soft processor core (SW partition), a set of NFS cores (HW partition), the HW/SW interface, and input/output (I/O) peripherals. An application example based on data obtained in an ubiquitous computing environment has been successfully implemented using an FPGA of Xilinx's Virtex 5 family [4].

In [5], the authors describe a framework to Context Acquisition Services and Reasoning Algorithms (CASanDRA) to be directly consumed by any type of application needing to handle context information. CASanDRA decouples the acquisition and inference tasks from the application development by offering a set of interfaces for information retrieval. The framework design is based on a data fusion-oriented architecture. CASanDRA has been designed to be easily scalable; it simplifies the integration of both new sensor access interfaces and fusion algorithms deployment, as it also aims at serving as a testbed for research.

In this work, we implement a AmI testbed and investigate the performance of k-means clustering algorithm based on Python. For evaluation, we considered sensing data. As evaluation metrics we considered respiratory rate and heart rate. We speeded up the k-means clustering algorithm by using parallel processing.

The structure of the paper is as follows. In Section 2, we present a short description of AmI. In Section 3, we give a brief introduction of k-means clustering algorithm. In Section 5, we show the description and design of the testbed. In Section 6, we discuss the experimental results. Finally, conclusions and future work are given in Section 7.

2 Ambient Intelligence (AmI)

In the future, small devices will monitor the health status in a continuous manner, diagnose any possible health conditions, have conversation with people to persuade them to change the lifestyle for maintaining better health, and communicates with the doctor, if needed [6]. The device might even be embedded into the regular clothing fibers in the form of very tiny sensors and it might communicate with other

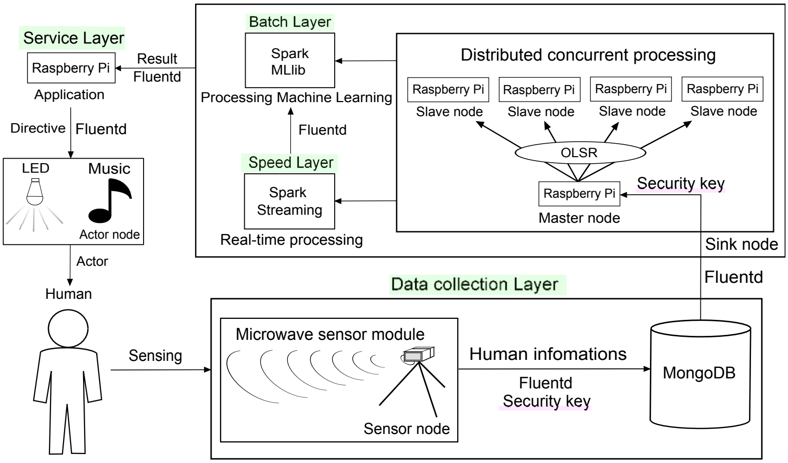


Fig. 1 Structure of AmI testbed.

devices including the variety of sensors embedded into the home to monitor the lifestyle. For example, people might be alarmed about the lack of a healthy diet based on the items present in the fridge and based on what they are eating outside regularly.

The AmI paradigm represents the future vision of intelligent computing where environments support the people inhabiting them [7, 8, 9]. In this new computing paradigm, the conventional input and output media no longer exist, rather the sensors and processors will be integrated into everyday objects, working together in harmony in order to support the inhabitants [10]. By relying on various artificial intelligence techniques, AmI promises the successful interpretation of the wealth of contextual information obtained from such embedded sensors, and will adapt the environment to the user needs in a transparent and anticipatory manner.

3 The *k*-means Clustering Algorithm

Here, we briefly describes the standard k-means algorithm [11]. The *k*-means is a typical clustering algorithm in data mining and which is widely used for clustering large set of data. The *k*-means algorithm is one of the most simple, non-supervised learning algorithms, which was applied to solve the problem of the well-known cluster [12]. It is a partitioning clustering algorithm, this method is to classify the given date objects into *k* different clusters through the iterative, converging to a local minimum. So the results of generated clusters are compact and independent. The algorithm consists of two separate phases. The first phase selects *k* centers randomly, where the value *k* is fixed in advance. The next phase is to take each data object to the nearest center [13]. Euclidean distance is generally considered to determine the distance between each data object and the cluster centers. When all the data objects are included in some clusters, the first step is completed and an early grouping is done. Recalculating the average of the early formed clusters. This iterative process continues repeatedly until the criterion function becomes the minimum. Supposing that the target object is x, x_i indicates the average of cluster C_i, criterion function is defined as follows:

$$E = \sum_{i=1}^{k} \sum_{x \in C_i} |x - x_i|^2 . \tag{1}$$

where E is the sum of the squared error of all objects in database. The distance of criterion function is Euclidean distance, which is used for determining the nearest distance between each data object and cluster center. The Euclidean distance between one vector $x = (x_1, x_2, \ldots, x_n)$ and another vector $y = (y_1, y_2, \ldots, y_n)$, The Euclidean distance $d(x_i, y_i)$ can be obtained as follow:

$$d(x_i, y_i) = \left[\sum_{i=1}^{n} (x_i - y_i)^2 \right]^{\frac{1}{2}} . \tag{2}$$

The process of k-means algorithm in Algorithm 1. The k-means clustering algorithm always converges to local minimum. Before the k-means algorithm converges, calculations of distance and cluster centers are done while loops are executed a number of times, where the positive integer t is known as the number of k-means iterations. The precise value of t varies depending on the initial starting cluster centers [14]. The distribution of data points has a relationship with the new clustering center, so the computational time complexity of the k-means algorithm is $O(nkt)$. n is the number of all data objects, k is the number of clusters, t is the iterations of algorithm. Usually requiring $k \ll n$ and $t \ll n$.

Algorithm 1 The process of k-means algorithm.

```
Input: Number of desired clusters, $k$, and a database $D = d_1, d_2, \ldots, d_n$ containing $n$ data objects;
Output: A set of $k$ clusters;
Randomly select $k$ data objects from dataset $D$ as initial cluster centers;
Calculate the distance between each data object $d_i$ $(1 \le i \le n)$ and all $k$ cluster centers $c_j$ $(1 \le j \le k)$ and assign data object $d_i$ to the nearest cluster;
For each cluster $j$ $(1 \le j \le k)$, recalculate the cluster center;
Until no changing in the center of clusters;
```

4 Parallel Processing

The parallel processing is the simultaneous use of multiple compute resources to solve a computational problem [15]. A problem is broken into discrete parts that can be solved concurrently. Each part execute simultaneously on different processors.

Table 1 Simulation parameters.

Parameters	Values
Number of clusters	3
Initial centroids	random
Precompute distance	true

Table 2 Simulation results of processing time.

Number of Cores	Processing time [sec]
1	10.89
4	7.78

As a result, it is able to perform the calculation processing at a higher speed. The compute resources is typically either a single computer with multiple processor/core computer or, of the plurality of network-connected in any number of computers.

5 Testbed Description

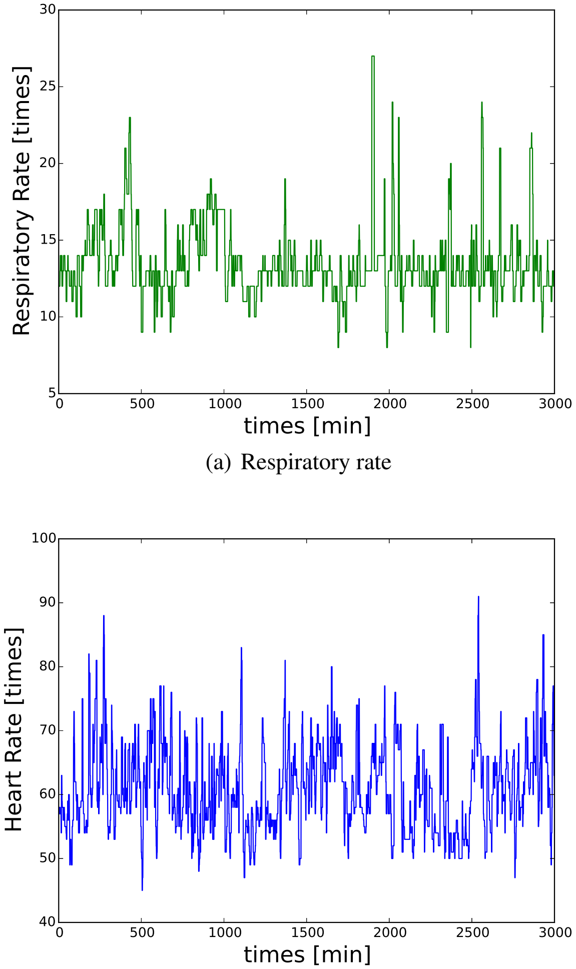
In Fig. 1 is shown the structure of AmI testbed. Our testbed is composed of five Raspberry Pi 3 Model B [16, 17, 18]. The Raspberry Pi is a credit card-sized single-board computer developed by the Raspberry Pi Foundation. The operating systems mounted on these machines are Raspbian version Debian 7.8 with kernel 3.18.11 [19].

We use Microwave Sensor Module (MSM) called DC6M4JN3000, which emits microwaves in the direction of a human or animal subject [20]. These microwaves reflect back off the surface of the subject and change slightly in accordance with movements of the subject's heart and lungs. From these changes, the DC6M4JN3000 measures biological information such as heart and respiratory rates.

The DC6M4JN3000 is capable of measuring heart rate within a margin of error of ±10 [%] when placed roughly three meters away from the target subject. The unit uses microwaves, so it can detect targets located behind obstacles such as mattresses, doors, and walls. This makes it possible to measure biological information even when the target is asleep or in situations where the targets privacy must be maintained (such as in the washroom or bathroom), thereby enabling this sensor module to boost the level of service given in elderly care or nursing care. In this way, the Quality of Life (QoL) is improved.

6 Simulation Results

The simulation parameters are shown in Table 1. We collected data for respiratory rate and heart rate.

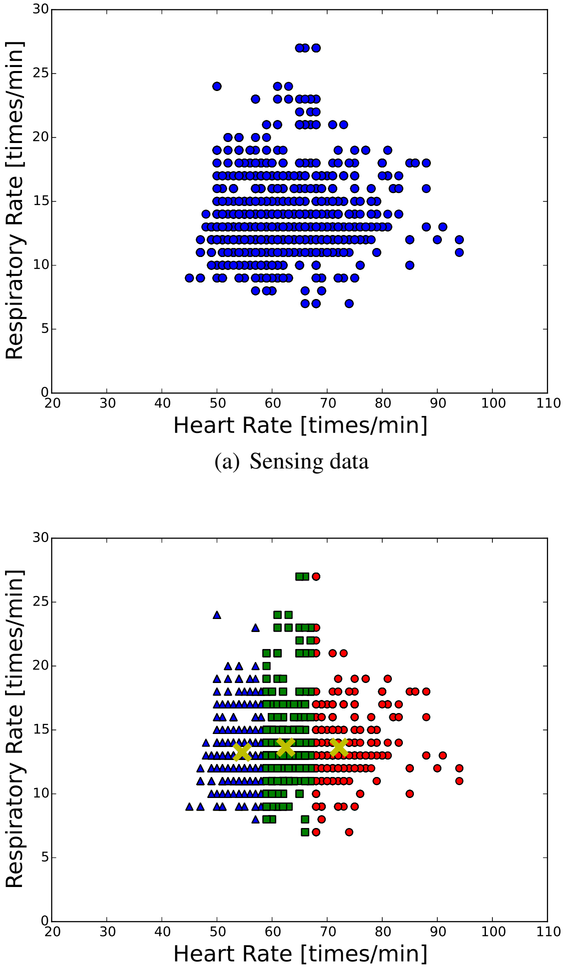


(a) Respiratory rate

(b) Heart rate

Fig. 2 Respiratory rate and heart rate.

In Fig. 2, we show the respiratory rate and heart rate. In Fig. 3, we present the result by sensing data and using *k*-means clustering algorithm. We can see 3 regions of clustering. Based on this data, the system is able to judge human health conditions. In Table 2, we show the processing time considering parallel processing. For 4 cores, the processing time is faster than 1 core.



(a) Sensing data

(b) Clustering

Fig. 3 Simulation results.

7 Conclusions

In this paper, we presented the simulation results of a AmI testbed considering *k*-means clustering algorithm. We clustered sensed data by *k*-means clustering algorithm considering parallel processing. From simulations, we found the following results.

- The *k*-means clustering algorithm can cluster sensed data.
- Using our testbed, the QoL can be improved.
- Speed up the *k*-means clustering by parallel processing.

In the future, we would like to make extensive simulations for different simulation scenarios and carry out experiments using the implemented testbed.

References

1. M. Lindwer, D. Marculescu, T. Basten, R. Zimmermann, R. Marculescu, S. Jung, E. Cantatore, “Ambient Intelligence Visions and Achievements: Linking Abstract Ideas to Real-World Concepts”, Design, Automation and Test in Europe Conference and Exhibition, pp. 10-15, 2003.
2. O. Gabel, L. Litz, M. Reif, “NCS Testbed for Ambient Intelligence”, IEEE International Conference on Systems, Man and Cybernetics, Vol. 1, pp. 115-120, 2005.
3. I. del Campo, M.V. Martinez, J. Echanobe, K. Basterretxea, “A Hardware/Software Embedded Agent for RealTime Control of Ambient-Intelligence Environments”, IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1-8, 2012.
4. Virtex 5 Family Overview, Xilinx Inc., San Jose, CA, 2009.
5. A. M. Bernardos, P. Tarrio, J. R. Casar, CASanDRA, “A framework to provide Context Acquisition Services ANd Reasoning Algorithms for Ambient Intelligence Applications”, International Conference on Parallel and Distributed Computing, Applications and Technologies, pp. 372-377, 2009.
6. G. Acampora, D. Cook, P. Rashidi and A. V. Vasilakos, “A Survey on Ambient Intelligence in Health Care”, Proc. of the IEEE, Vol.101, No.12, pp. 2470-2494, 2013.
7. E. Aarts and R. Wichert, “Ambient intelligence”, Technology Guide, pp. 244-249, 2009.
8. E. Aarts and B. de Ruyter, “New research perspectives on ambient intelligence”, Journal of Ambient Intelligence and Smart Environments, Vol. 1, No. 1, pp. 5-14, 2009.
9. A. Vasilakos and W. Pedrycz, “Ambient Intelligence, Wireless Networking, And Ubiquitous Computing. Norwood”, MA, USA: Artech House, Inc., 2006.
10. F. Sadri, “Ambient intelligence: A survey”, ACM Comput. Surv., Vol. 43, No. 4, pp. 36:1-36:66, Oct. 2011.
11. S. Na, L. Xumin, G. Yong, “Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm”, Third International Symposium on Intelligent Information Technology and Security Informatics (IITSI), pp. 63-67, 2010.
12. S. Jigui, L. Jie, Z. Lianyu, “Clustering algorithms Research”, Journal of Software, Vol. 19, No. 1, pp. 48-61, 2008.
13. A. M. Fahim, A. M. Salem, F. A. Torkey, “An efficient enhanced *k*-means clustering algorithm”, Journal of Zhejiang University Science A, Vol. 10, pp. 1626-1633, 2006.
14. K. A. Abdul Nazeer, M. P. Sebastian, “Improving the Accuracy and Efficiency of the *k*-means Clustering Algorithm”, Proc. of the World Congress on Engineering, vol. 1, 2009.
15. M.J. Gonzalez, IEEE, “Program Suitability for Parallel Processing”, 2006.
16. “Raspberry Pi Foundation.”, http://www.raspberrypi.org/.
17. T. Oda, A. Barolli, S. Sakamoto, L. Barolli, M. Ikeda, K. Uchida, “Implementation and Experimental Results of a WMN Testbed in Indoor Environment Considering LoS Scenario”, The 29-th IEEE International Conference on Advanced Information Networking and Applications (AINA-2015), pp. 37-42, 2015.
18. R. Obukata, T. Oda, L. Barolli, “Design of an Ambient Intelligence Testbed for Improving Quality of Life”, The 9-th International Symposium on Mining and Web (MAW-2016), pp. 714-719, 2016.
19. “Raspbian: FrontPage.”, https://www.raspbian.org/.
20. “Sharp to Release Microwave Sensor Module.”, http://www.sharp-world.com/corporate/news/150625.html/.

Variable Neighborhood Search Algorithms for the Node Placement Problem in Multihop Networks

Kengo Katayama, Yusuke Okamoto, Elis Kulla, Noritaka Nishihara

Abstract We consider a problem of finding an optimal node placement that minimizes the amount of traffic by reducing the weighted hop distances in multihop networks. The problem is called Node Placement Problem (NPP) and is known to be NP-hard. Therefore, several heuristic and metaheuristic algorithms have been proposed for solving NPP, such as local search, genetic algorithm, simulated annealing, tabu search, iterated local search, ant colony optimization, etc. Although Variable Neighborhood Search (VNS) is known to be one of the most promising and efficient metaheuristic algorithms for optimization problems, VNS has not been shown for NPP yet. In this paper we propose VNS algorithms for NPP. The proposed VNSs consist of two phases: local search phase to obtain a local optimum and perturbation phase to get out of the corresponding valley in the search space. We show six types of neighborhood change schemes for the perturbation phase of VNS, and through computational experiments, we compare each performance of six VNSs incorporating k-swap local search, called VNS1, VNS2,. . ., VNS6. The experimental results indicate that VNS4 outperformed the others for large problem instances particularly, which adopts a suitable perturbation size selected by exploring from the upper bound that is adaptively lower in the search.

Kengo Katayama
Department of Information and Computer Engineering, Okayama University of Science,
Okayama, Japan, e-mail: katayama@ice.ous.ac.jp

Yusuke Okamoto
Department of Information and Computer Engineering, Okayama University of Science,
Okayama, Japan, e-mail: t15jm02oy@ous.jp

Elis Kulla
Department of Information and Computer Engineering, Okayama University of Science,
Okayama, Japan, e-mail: kulla@ice.ous.ac.jp

Noritaka Nishihara
Department of Information and Computer Engineering, Okayama University of Science,
Okayama, Japan, e-mail: nisihara@ice.ous.ac.jp

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing,*
Communication and Applications, Lecture Notes on Data Engineering
and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_62

1 Introduction

Heuristics and metaheuristics [2] are widely applied to numerous combinatorial optimization problems to obtain near-optimal solutions efficiently, because no polynomial time algorithms are known for solving them exactly. In this paper, we consider a combinatorial optimization problem of finding an optimal node placement that minimizes the amount of traffic by reducing the weighted hop distances in multihop networks. The problem is called Node Placement Problem (NPP).

Since NPP is NP-hard [1], several heuristic and metaheuristic algorithms have been proposed: greedy method [6, 8], local search [6], tabu search [6], genetic algorithm [6], simulated annealing [6, 8], and ant colony optimization[10] for NPP. More recently, Katayama *et al.* [5, 4] presented an effective metaheuristic called Iterated k-swap Local Search (IKLS for short) that consists of a new local search called k-swap Local Search (KLS) based on the idea of variable depth search [7, 9] and a perturbation method. As the perturbation methods, we investigated Cross-Kick [5] and newly showed several kicks such as Rhombus-Kick and IV-Kick [4].

In this paper, we propose Variable Neighborhood Search (VNS for short) based metaheuristics for NPP. Although VNS is known to be one of the most promising and efficient metaheuristic approaches to numerous hard optimization problems, to the best of our knowledge, our VNSs are the first investigation for NPP. The proposed VNSs simply consist of two phases in each iteration: local search phase to obtain a local optimum and perturbation phase to get out of the corresponding valley in the search space. We investigate six types of neighborhood change schemes for the perturbation phase, and through computational experiments, we compare each performance of the six VNSs incorporating KLS, called VNS1, VNS2,..., VNS6. The experimental results show that VNS4 that adopts a suitable perturbation size selected by exploring from the upper threshold which is adaptively lower in the search, outperformed the other variants for large problem instances particularly.

2 Node Placement Problem (NPP)

In this section, we describe NPP as a combinatorial optimization problem. We consider the Bidirectional Manhattan Street Network (BMSN) [6, 8], which has a regular topology in multihop WDM lightwave networks. It can be represented by a grid graph on torus $G_{m,m} = (V,E)$, where V is the set of node slots that form m columns and m rows ($n = m \times m$) and E is the set of bidirectional edges.

Let (x,y) denote a slot address (coordinates) in the x-th column and the y-th row ($x,y = 0,1,\ldots,m-1$) of the graph. Each node slot has four bidirectional edges in vertical and horizontal directions so as to form a torus. Each of n nodes $(0,1,\ldots,n-1)$ can be assigned to each of the n slots on the graph without duplications, and the amount of traffic among network nodes can be given by an $n \times n$ traffic matrix T, where each element $t_{i,j}$ denotes the traffic flow from node i to node j ($i,j = 0,1,\ldots,n-1$, $i \neq j$) that has a real or integer value. We simplify the amount of traffic into two types (heavy and light, denoted by t_H and t_L, respectively) as follows:

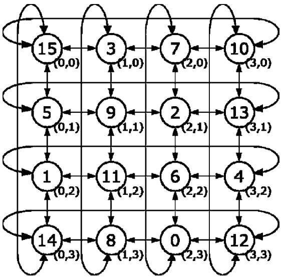


Fig. 1 A grid graph with 16 nodes (4 × 4)

$$t_{i,j} = \begin{cases} 1 & \text{if} \quad \text{nodes } i \text{ and } j \text{ exchange a lot of traffic} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The diagonal is off, $t_{i,i} = 0$ for all i.

Assuming that $t_{i,j} = 1$, the node i can communicate with the node j *directly* if the nodes are assigned in *adjacent* slots of the graph. If not adjacent, they must communicate through several node slots, i.e., the number of hops increases. A function $h(i, j)$ is provided that returns the hop distance in the shortest path between two nodes i and j on the graph.

The objective of NPP is to find a node placement σ of n nodes that minimizes the weighted hop distances:

$$f(\sigma) = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} t_{i,j} \times h(\sigma_i, \sigma_j), \tag{2}$$

where σ_i $(i = 0, 1, \ldots, n-1)$ represents i-th node in the node placement that corresponds to the node number assigned at $(\sigma_i \mod m,\ \lfloor \sigma_i/m \rfloor)$ slot coordinates in the graph of the network. Figure 1 shows an example of node placements for graph $G_{4,4}$ that consists of 4 columns and 4 rows $(n = 16)$, and the corresponding node placement is $\sigma = \{14, 8, 6, 1, 11, 4, 10, 2, 13, 5, 3, 9, 15, 7, 12, 0\}$. If the traffic matrix gives that for example, nodes 5 and 0 (that are assigned at slots (0,1) and (2,3), respectively) exchange a lot of traffic (i.e., $t_{5,0} = 1$), the number of hops in the shortest path is obviously 4. Therefore, the cost for this pair of nodes is $t_{5,0} \times h(\sigma_5, \sigma_0) = 4$.

3 Variable Neighborhood Search for NPP

Variable neighborhood search (VNS) is a metaheuristic based on systematic changes of neighborhoods both in local search phase, to find a local optimum, and in perturbation phase to emerge from the corresponding valley [3]. We focus systematic neighborhood changes in the perturbation phase. One of the most important roles in the phase is to adaptively select, in accordance with a search situation, an appropriate *perturbation strength* that is the number of solution components which are modified, in order to escape from local optima and to lead to suitable search space for the local search phase.

3.1 VNSs for NPP

We show six types of VNSs for NPP, called VNS1, VNS2,..., VNS6. The main difference between them is neighborhood change schemes that are characterized by the fluctuations—increase/decrease of the perturbation strength and upper/lower threshold values in the perturbation phase. All VNSs for NPP basically require setting of two parameters to control the change of neighborhood: *kmax* and *kmin* that are the upper and lower threshold values of the perturbation strength, respectively.

The flow of VNS1 is shown in Figure 2. VNS1 is based on the standard VNS algorithm. First, an initial solution σ is generated randomly, and the parameters *kmax* and *kmin* are set. The variable *ksize* corresponding to the perturbation strength is set to *kmin* initially in VNS1. At line 3 (and line 7), the solution σ is locally optimized by a local search method. In this phase we employ an effective k-swap local search (KLS). The detail of KLS is described in the next subsection. In the main process of VNS (line 5–13), the perturbation and local search phases are repeated until the stopping condition (line 13) is met. At line 6, a local optimal solution σ found by KLS is perturbed according to *ksize*. In this process, we employ one of the simplest perturbation operations called *Random Kick* that performs to select *ksize* nodes randomly and to perturb them, i.e., it repeats to swap two nodes from the selected nodes without duplication. At line 7, the perturbed solution σ is locally optimized by KLS. The cost of σ after KLS and σ_{best} are then compared, and the best found solution σ_{best} is saved if the condition of line 8 is satisfied. The following processes of lines from 9–11 are the neighborhood change scheme for VNS1. At line 9, if σ after KLS at current iteration *Iter* is better than σ_{prev} that is the previous solution found at $Iter-1$, perturbation strength *ksize* is reset to *kmin*, otherwise we increase *ksize* one percent of problem size n to find a suitable perturbation strength. If *ksize* is larger than *kmax* that is the upper threshold value, *ksize* is reset to the lower threshold one. At line 12, we set σ to σ_{prev} to use it at line 9. This leads to perform a random walk search for VNS, i.e., the solution σ given to the perturbation process at line 6 of iteration *Iter* is the one found at the previous iteration $Iter-1$. The process of VNS is repeated until the stopping condition is satisfied.

The remaining VNS variants for NPP are explained as follows.

- VNS2: Line 9 of Figure 2 for VNS1 is replaced as follows to be VNS2:

  ```
  9    if $f(\sigma) < f(\sigma_{prev})$ then $kmin := (ksize + kmin)\,/\,2$; $ksize := kmin$;
  ```

  Different to VNS1, the lower threshold *kmin* in VNS2 is adaptively changed to an average value of current *ksize* and *kmin* if the condition at line 9 is satisfied. It is expected to contribute to finding of the suitable strength more efficiently.
- VNS3: Lines 1, 9, 10 and 11 of Figure 2 for VNS1 are replaced as follows to be VNS3:

```
    procedure basic Variable Neighborhood Search
    input: traffic matrix $T$
    output: best solution $\sigma_{best}$
    begin
 1    set the initial values of $kmin$, $kmax$, and set $ksize = kmin$;
 2    generate a solution randomly $\sigma$;
 3    $\sigma$ := Local Search($\sigma$);
 4    $\sigma_{best}$ := $\sigma_{prev}$ := $\sigma$;
 5    repeat
 6        $\sigma$ := Perturbation($\sigma$, $ksize$);
 7        $\sigma$ := Local Search($\sigma$);
 8        if $f(\sigma) < f(\sigma_{best})$ then $\sigma_{best}$ := $\sigma$; endif
 9        if $f(\sigma) < f(\sigma_{prev})$ then $ksize$ := $kmin$;
10        else increase $ksize$; endif
11        if $kmax < ksize$ then $ksize$ := $kmin$; endif
12        $\sigma_{prev}$ := $\sigma$;
13    until terminate = true;
14    return $\sigma_{best}$;
    end;
```

Fig. 2 A basic flow of variable neighborhood search for NPP

```
 1  set the initial values of $kmin$, $kmax$, and set $ksize = kmax$;
 9      if $f(\sigma) < f(\sigma_{prev})$ then $ksize$ := $kmax$;
10      else decrease $ksize$; endif
11      if $kmin > ksize$ then $ksize$ := $kmax$; endif
```

Opposite to VNS1, a suitable perturbation strength in VNS3 is found from upper threshold value *kmax* to lower one *kmin*.

- VNS4: Like VNS3, a suitable perturbation strength in VNS4 is found from *kmax* to *kmin*. In addition, opposite to VNS2, line 9 is replaced as follows to be VNS4:

```
 9      if $f(\sigma) < f(\sigma_{prev})$ then $kmax$ := $(ksize + kmax)$ / 2; $ksize$ := $kmax$;
```

Therefore, all settings of VNS4 scheme are opposite to those of VNS2.

- VNS5: It is based on settings of VNS2. In addition, the upper threshold *kmax* is also changed adaptively to be VNS5. The upper threshold *kmax* is updated as follows: $\alpha = (ksize - kmin)/2$ and $kmax = kmax - \alpha$. Therefore, both of *kmax* and *kmin* are changed with α simultaneously if the condition at line 9 is satisfied.
- VNS6: It is based on settings of VNS4. Opposite to VNS5, the lower threshold *kmin* is also changed adaptively to be VNS6. The lower threshold *kmin* is updated as follows: $\alpha = (kmax - ksize)/2$ and $kmin = kmin + \alpha$.

3.2 Variable k-swap Local Search Algorithm

One of the most important parts in VNS is the local search phase (used at lines 3 and 7 shown in Figure 2). We employ an effective k-swap local search (KLS) based on the idea of Lin and Kernighan [9, 7].

```
    procedure Variable k-swap Local Search ($\sigma$)
    begin
 1    $P_{out} := \{0,1,\dots,n-1\}$, $g_{LastImp} := -\infty$;
 2    repeat
 3      $\sigma_{prev} := \sigma$, $g := 0$, $g_{best} := 0$, $P_{in} := \{0,1,\dots,n-1\}$;
 4      select $i \in P_{out}$ randomly;
 5      $P_{out} := P_{out} \backslash \{i\}$, $P_{in} := P_{in} \backslash \{i\}$;
 6      repeat
 7        find a node $j$ with $\max_{j \in P_{in}} \delta_{i,j} := \mathrm{SwapGain}(i,j,\sigma)$;
 8        $\sigma := \mathrm{SwapMove}(i,j,\sigma)$, $g := g + \delta_{i,j}$, $P_{in} := P_{in} \backslash \{j\}$;
 9        if $g > g_{best}$ then $\sigma_{best} := \sigma$, $g_{best} := g$;
10      until $P_{in} = \emptyset$ or $g < g_{LastImp}$;
11      if $g_{best} > 0$ then $P_{out} := \{0,1,\dots,n-1\}$, $\sigma := \sigma_{best}$, $g_{LastImp} := -g_{best}$ else $\sigma := \sigma_{prev}$;
12    until $P_{out} = \emptyset$;
13    return $\sigma$;
    end;
```

Fig. 3 A flow of variable k-swap local search for NPP

KLS determines dynamically at each iteration the value of k (the number of swaps of nodes), since it is computationally too expensive to search the *complete* k-swap neighborhood. In KLS, the (variable) k-swap neighborhood $\mathcal{N}_{k\text{-swap}}$ of a given node placement σ is defined as the set of chained neighbors σ' that can be obtained by applying a sequence of the single-swap moves to σ in feasible search space. It indicates that KLS attempts to search a small fraction of the large neighborhoods in reasonable times. The length l of the sequence is adaptively decided in the algorithm. All l chained neighbor solutions obtained by the sequence are different because we assure that *cycling* among the neighbors in the sequence is avoided.

The pseudo-code of KLS is shown in Figure 3. KLS has outer (lines 2–12) and inner (lines 6–10) loops. At lines 1 and 3 P_{out} and P_{in} are initialized. These sets ensure that no node is allowed to be swapped twice in one sequence in order to avoid the cycling among the chained neighbors. In the outer loop, we select a random node i from P_{out} at line 4. The selected node i is removed from each of P_{out} and P_{in}. Node i is one of the nodes to be paired. In the inner loop, the other pairing node j is selected in P_{in} such that the gain value of $\mathrm{SwapGain}(i,j,\sigma)$ is maximal even if the gain value is smaller than zero (line 7). At line 8, the pair of nodes i and j is swapped to move to a neighbor solution from σ. Moreover, the gain g is updated and the node j is removed from P_{in}. If the gain g is larger than the gain of the best solution found so far in the search, the best solution σ_{best} and the corresponding best gain value g_{best} are saved. The search of inner loop is repeated until $P_{in} = \emptyset$. If the best gain value is better than zero at line 11, P_{out} is initialized, and the k-swap neighborhood search is repeated until $P_{out} = \emptyset$. Finally, the best solution σ found in the search is returned.

To reduce the computation time of KLS, the termination condition of the inner loop is modified so that the loop is terminated if the current gain value g is smaller than the best gain value $g_{LastImp}$ recorded at the last iteration as shown at line 10 in Figure 3. The related processing can be found at lines 1 and 11. This modification

Table 1 Comparison Results of VNSs ($kmin = 4$ and $kmax = n \times 0.3$)

solver	n	opt	best	avg	err(%)	stdev	time[s]
VNS1	64	76	**80**	83.5	9.9	2.2	0.07
	256	307	355	376.4	22.6	10.3	1.38
	1024	1228	1866	2017.9	51.1	57.5	41.49
VNS2	64	76	**80**	83.3	9.6	2.2	0.08
	256	307	366	381.8	24.3	8.1	1.76
	1024	1228	1675	1760.8	36.4	44.1	75.14
VNS3	64	76	81	84.1	10.7	2.3	0.08
	256	307	367	381.8	24.3	8.3	2.03
	1024	1228	1682	1737.2	37.0	36.7	98.46
VNS4	64	76	**80**	**83.1**	9.2	2.6	0.08
	256	307	**354**	**367.1**	19.6	8.2	1.83
	1024	1228	**1534**	**1622.3**	24.9	39.8	87.44
VNS5	64	76	**80**	**83.1**	9.2	2.0	0.08
	256	307	361	374.3	21.9	9.3	1.65
	1024	1228	1602	1722.8	30.5	58.0	69.49
VNS6	64	76	**80**	83.7	10.1	2.5	0.08
	256	307	363	377.8	23.0	8.3	1.97
	1024	1228	1621	1681.5	32.0	34.5	95.26

is quite useful to considerably increase the efficiency of KLS without large loss of solution qualities. Note that no parameter setting by user is required in KLS.

4 Experimental Results

To evaluate each performance of the VNS algorithms shown above, we performed computational experiments on the benchmark instances of NPP. The instances are standard ones for NPP provided in [5] which are identified by the following problem sizes: n = 16 (=4×4), 64 (=8×8), 256 (=16×16), 1024 (=32×32). Each set consists of 20 instances. The optimal solution values are known as 19, 76, 307 and 1228 for all 20 instances in each set.

All experiments were performed on a Linux computer with Ubuntu, Intel Core i7 3.6GHz and 15.6GiB RAM. All algorithm codes were written in C, and the codes were compiled with the gcc 4.8 with '-O3' option. We performed each computation with VNS1, VNS2,…, VNS6 in a single run, for each of 20 instances in each of the problem sets. The number of iterations that is the termination condition of VNSs for the single run were set to 100 iterations. The parameters in the neighborhood change process are set as follows: $kmin = 4$ and to investigate suitable values, $kmax$ are set to $n \times \beta$, where $\beta = \{0.1, 0.2, 0.3, 0.4, 0.5\}$, respectively.

Table 1 shows the results of six variants of VNSs with $kmax = n \times 0.3$ (which are the best one in the results obtained by setting each value of $kmax$). In the table, the first two columns are the problem size n of each problem set and the optimal solution cost value corresponded to the problem size. In the following columns we show the best cost value "best", average one "avg", the quality of the average one "err" in the 20 solutions obtained, standard deviation "stdev", and computational time "time" in seconds. The values in bold in each columns of "best" and "avg" indicate the best results for each problem size in all VNSs.

It can be observed in Table 1 that VNS4 has better results in the best and average columns than the other VNSs on the middle and large-sized instances particularly. Furthermore we observed that VNS4 is robust in the parameter setting values *kmax*, because VNS4 obtained better results than those of the other VNSs in wide range of *kmax* values. It means that the performances of VNS4 do not affect strongly the setting values of *kmax* in the experiments.

5 Conclusion

In this paper, we presented variable neighborhood search (VNS) metaheuristic algorithms and compared each performance of six variants of VNS for NPP. Experimental results showed that VNS4 outperformed the others. We also observed that settings of parameter values *kmax* in VNS4 are robust in comparison to the others.

More work remains to be carried out in investigating the VNS performances in a *longer* running time. In our initial experiments, we performed the VNSs on the same setting of longer running time permitted in [4]. The initial results showed that the basic VNS, i.e., VNS1, obtained slightly better average solutions than the ones obtained by the effective metaheuristic presented in our previous work [4].

References

1. Garey, M.R., Johnson, D.S.: Computers and Intractability: A Guide to the Theory of NP-Completeness. Freeman, New York (1979)
2. Gendreau, M., Potvin, J.Y.: Handbook of Metaheuristics. International Series in Operations Research & Management Science. Springer (2010)
3. Hansen, P., Mladenović, N., Moreno-Pérez, J.A.: Variable neighbourhood search: Methods and applications. Annals of Operations Research **175**(1), 367–407 (2010)
4. Katayama, K., Akagi, Y., Kulla, E., Minamihara, H., Nishihara, N.: New kick operators in iterated local search based metaheuristic for solving the node placement problem in multihop networks. In: Proceedings of the 17th International Conference on Network-Based Information Systems (NBiS-2014), pp. 141–148 (2014)
5. Katayama, K., Yamashita, H., Narihisa, H.: Variable depth search and iterated local search for the node placement problem in multihop WDM lightwave networks. In: Proceedings of IEEE Congress on Evolutionary Computation, pp. 3508–3515 (2007)
6. Kato, M., Oie, Y.: Reconfiguration algorithms based on meta-heuristics for multihop WDM lightwave networks. In: Proc. IEEE International Conference on Communications, pp. 1638–1644 (2000)
7. Kernighan, B., Lin, S.: An efficient heuristic procedure for partitioning graphs. Bell System Technical Journal **49**, 291–307 (1970)
8. Kitani, T., Yonedu, M., Funabiki, N., Nakanishi, T., Okayama, K., Higashino, T.: A two-stage hierarchical algorithm for wavelength assignment in WDM-based bidirectional manhattan street networks. In: Proc. the 11th IEEE International Conf. on Networks, pp. 419–424 (2003)
9. Lin, S., Kernighan, B.: An effective heuristic algorithm for the traveling salesman problem. Operations Research **21**, 498–516 (1973)
10. Melo, L.A., Pereira, F.B., Costa, E.: MC-ANT: A multi-colony ant algorithm. In: Proc. 9th International Conference on Artificial Evolution, EA 2009, (LNCS 5975), pp. 25–36 (2010)

Performance Evaluation of VANETs in Different Real Map Scenarios

Ryuji Ono, Elis Kulla and Evjola Spaho

Abstract With the advancement of Vehicular technologies, Vehicular Adhoc Networks will soon be a reality in our daily lives. However in order to fully take advantage of the new applications, we need to consider different parameters, while implementing the network. Road patterns, types of data, vehicle density and so on are some parameters that can affect the performance. In this paper we conduct simulations for VANETs in the Japanese city of Soja, Okayama Prefecture and one of the busiest parts of Tokyo metropolitan area, Shibuya. In order to create the road map we used real data from Open Source Map (OSM). Then, we generated mobility traces from eWorld into SUMO-compatible format. The data traffic is generated and evaluated by NS3 and its related tools. We investigate the effect of road patterns, types of traffic data and vehicle density.

1 Introduction

Vehicular Ad-hoc Networks (VANETs) are a special type of Ad-hoc networks and are an important component of the Intelligent Transportation Systems (ITS). They can been utilized to guarantee road safety, to avoid potential accidents by creating

Ryuji Ono
Okayama University of Science, Department of Information ad Computer Engineering,
1-1 Ridai-cho, Kita-ku, 700-0005 Okayama, Japan
e-mail: t16jm02or@ous.jp

Elis Kulla
Okayama University of Science, Department of Information ad Computer Engineering,
1-1 Ridai-cho, Kita-ku, 700-0005 Okayama, Japan
e-mail: kulla@ice.ous.ac.jp

Evjola Spaho
Polytechnic University of Tirana, Department of Electronics and Communications,
Mother Teresa Square, No. 4, Tirana, Albania
e-mail: evjolaspaho@hotmail.com

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering
and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_63

new forms of inter vehicle communications and applications. Due to the high cost of deploying and implementing VANET systems in a real environment, most of research is concentrated on simulations. In the recent years, a lot of simulators for VANETs have been developed [1]. For example, the IMPORTANT framework has been one of the first attempt to understand the dependence between vehicular traffic and communication performance [2], [3]. The authors analyzed the impact of the node mobility on the duration of communication paths. In [4], the authors present a simulator written in Java, which can generate mobility traces in several formats. There are also other powerful traffic simulators, like TranSim [5], which makes use of a cellular automaton for simulating the interaction of vehicles. Cellular Automaton based VEhicular NETwork (CAVENET) [7] is a lightweight simulator which can be used to understand the properties of the mobility models of vehicular traffic and their impact on the performance of VANETs. SUMO is another powerful traffic simulator, intended for traffic planning and road design optimization. There is an attempt to interface SUMO with NS2 [6]. Since VANETs are a specific class of ad-hoc networks, the commonly used ad-hoc routing protocols initially implemented for MANETs have been tested and evaluated for VANET environments. VANETs share some common characteristics with MANETs. They are both characterized by the movement and self organization of the nodes. We consider the possibility of using ad-hoc and MANET protocols for VANET scenarios. In other previous work [9], the evaluated the performance of MANET routing protocols, using as network simulator NS2 and CAVENET vehicular mobility model.

In this paper we conduct simulations for VANETs in the Japanese city of Soja, Okayama Prefecture and one of the busiest parts of Tokyo metropolitan area, Shibuya. In order to create the road map we used real data from Open Source Map (OSM)[10]. Then, we generated mobility traces from eWorld [11] into SUMO-compatible [12] format. The data traffic injected in the network is generated and evaluated by NS3 [13] and its related tools. We analyze and compare three routing protocols: Ad-hoc On-demand Distance Vector (AODV) [14], Optimized Link State Routing (OLSR) [15] and Destination-Sequenced Distance-Vector [**?**].

This paper is organized as follows. In Section 2, the three routing protocols are summarized. The simulation system design and implementation is presented in Section 3. In Section 4, we show the simulation results. Finally, the conclusions and future work are presented in Section 5.

2 Routing Protocols

2.1 Optimized Link-State Routing (OLSR) Protocol

The OLSR protocol [11] is a pro-active routing protocol, which builds up a route for data transmission by maintaining a routing table inside every node of the network. The routing table is computed upon the knowledge of topology information, which is exchanged by means of Topology Control (TC) packets. OLSR makes use of HELLO messages to find its one hop neighbors and its two hop neighbors through their responses. The sender can then select its Multi Point Relays (MPR) based on the one hop node which offer the best routes to the two hop nodes. By using this MPR-based flooding mechanism, the amount of control traffic can be reduced. Each node has also an MPR selector set which enumerates nodes that have selected it as an MPR node. OLSR uses TC messages along with MPR forwarding to disseminate neighbor information throughout the network. Host Network Address (HNA) messages are used by OLSR to disseminate network route advertisements in the same way TC messages advertise host routes.

2.2 Ad-hoc On-Demand Distance Vector (AODV) Routing Protocol

AODV is an on-demand routing protocol. It performs Route Discovery using control messages: Route Request (RREQ) and Route Reply (RREP). In AODV, routes are set up by flooding the network with RREQ packets. As a RREQ traverses the network, the traversed nodes store information about the source, the destination, and the node from which they received the RREQ. The later information is used to set up the reverse path back to the source. When the RREQ reaches a node, that knows a route to the destination or the destination itself, the node responds to the source with a RREP packet which is routed through the reverse path set up by the RREQ. This sets the forward route from the source to the destination. To avoid overburdening the nodes with information about routes which are no longer (if ever) used, nodes discard this information after a timeout. When either destination or intermediate node moves, a Route Error (RERR) is sent to the affected source nodes. When source node receives the RERR, it can reinitiate route discovery if the route is still needed. Neighborhood information is obtained by periodically broadcasting Hello packets [14]. For the maintenance of the routes, two methods can be used: a) ACK messages in MAC level or b) HELLO messages in network layer.

2.3 Dynamic Source Distance Vector

Destination Sequenced Distance Vector Routing (DSDV) [16] uses distance vectors to continuously maintain routes throughout a network. Unlike RIP, DSDV uses per-node sequence numbers to provide a total ordering on route information age in order to prevent loops. In DSDV, each node maintains a route to each other node.

3 Simulation System Design and Implementation

In order to generate our scenarios and conduct the simulations in this paper, we used different systems that can be interconnected with each other. We used eWorld, OSM, SUMO and NS3.

3.1 Simulation Environment

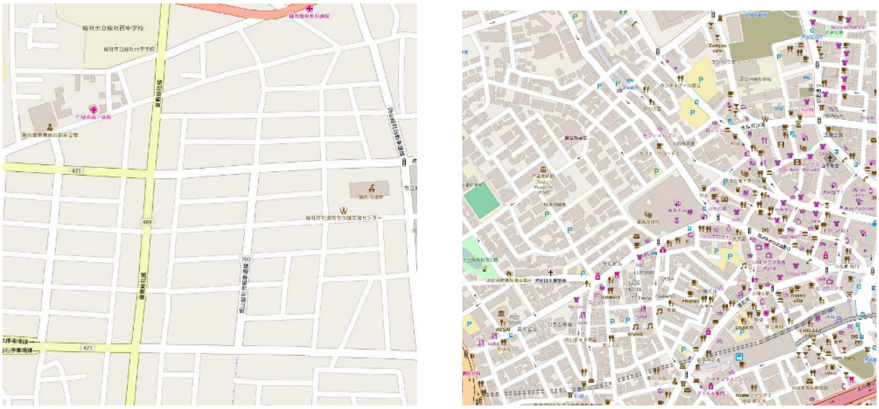
First, we use eWorld in order to get real from data from JOSM, a widely used open source map of the whole world. After getting the map data for the two parts of Japan, where we want to conduct simulations, we load it into eWorld and make certain adjustment to roads and streets. Then we add vehicle movement. eWorld allows users to setup starting and ending points of vehicle movements and the number of vehicles generated in a given time. You can also define different type of cars, with different maximum moving speed and physical length. All this data is exported to SUMO, where it is edited and compiled in order to get NS3 mobility data format.

After defining the mobility pattern, the network simulation model is based on NS-3. The NS3 simulator is developed and distributed completely in the C++ programming language, because it better facilitated the inclusion of C-based implementation code. The NS3 architecture is similar to Linux computers, with internal interface and application interfaces such as network interfaces, device drivers and sockets. The goals of NS3 are set very high: to create a new network simulator aligned with modern research needs and develop it in an open source community. Users of NS3 are free to write their simulation scripts as either C++ main() programs or Python programs. The NS3s low-level API is oriented towards the power-user but more accessible helper APIs are overlaid on top of the low-level API.

In order to achieve scalability of a very large number of simulated network elements, the NS3 simulation tools also support distributed simulation. The NS3 support standardized output formats for trace data, such as the pcap format used by network packet analyzing, tools such as tcpdump, and a standardized input format such as importing mobility trace files from different simulators.

The NS3 simulator has models for all network elements that comprise a computer network. For example, network devices represent the physical device that connects a node to the communication channel. This might be a simple Ethernet network interface card, or a more complex wireless IEEE 802.11 device. In our simulations we used IEEE 802.11p standard and TwoRayGroundPropagationLossModel.

IEEE 802.11p: Is an approved amendment to the IEEE 802.11 standard to add wireless access in vehicular environments (WAVE). It defines enhancements to 802.11 required to support Intelligent Transportation Systems (ITS) applications. The 802.11p standard is based on the 802.11 architecture, but version p is aimed at communications between vehicles and between them and fixed infrastructure. This new technology uses the 5.9 GHz band in various propagation environments: vehicle, open, urban, and so on. This standard defines the WAVE as the signaling



(a) Soja City Map

(b) Shibuya Map

Fig. 1 JOSM Maps of Soja city and Shibuya.

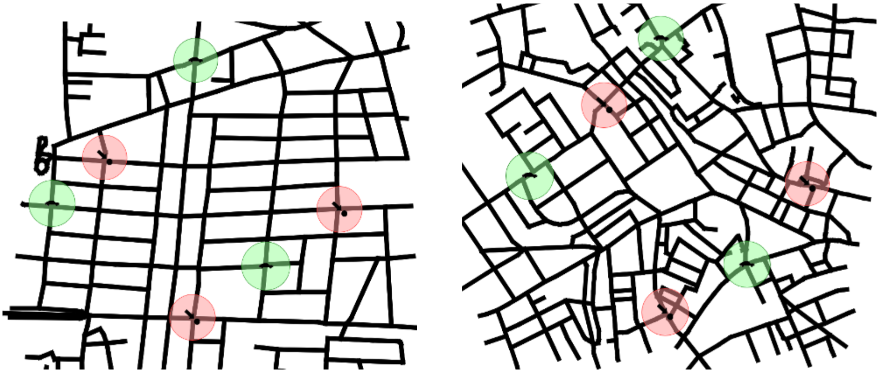
technique and interface functions that are controlled by the physical layer (MAC) devices where the physical layer properties change rapidly and where the exchanges of information have a short duration. The purpose of this standard is to provide a set of specifications to ensure interoperability between wireless devices trying to communicate in rapidly changing environments and in particular time periods. TwoRayGroundPropagationLossModel: It considers the direct path and a ground reflection path. The received power at distance t is calculated with the following equation:

3.2 Simulation Settings

In order to analyze VANETs and the behavior of the routing protocols, we setup different scenarios for different maps (Soja and Shibuya), different number of vehicles in the simulation area and different routing protocols: OLSR, AODV and DSDV.

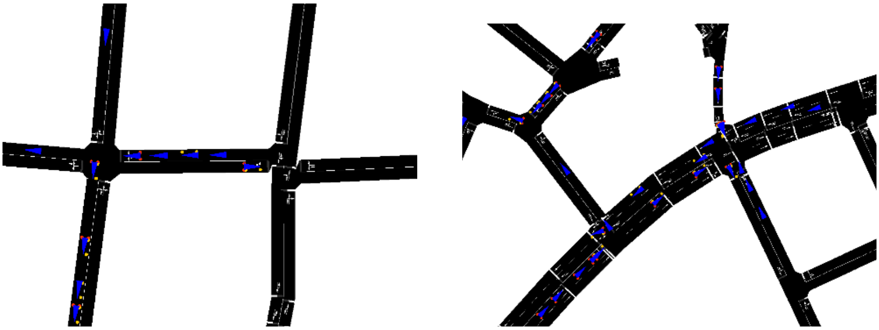
Different maps are used in order to create two environments where the density of the roads varies, considering Soja Town near our University and the traffic-packed region of Shibuya in Tokyo. The maps can be seen in Fig. 1. Data from the maps in then converted to e World and SUMO format, where we are able to run the vehicles in different patterns. Converted maps and a detailed road view of Soja town and Shibuya from SUMO Simulator can be seen in Fig. 2 and Fig. 3, respectively.

In order to analyze the behavior of VANETs, we setup three different densities in each map. In Fig. 2, we show three starting points and three ending points of car movements. For each pair of start-end points, vehicles with maximum moving speed of 60kmh, are generated in the area. They find their way to the end point in a random fashion, which is not the scope of this paper. Number of generated vehicles is: 50, 100 and 150. For a summary of parameter settings see Table 1.



(a) Soja City Map (b) Shibuya Map

Fig. 2 Maps of Soja city and Shibuya from SUMO Simulator.



(a) Soja City Map (b) Shibuya Map

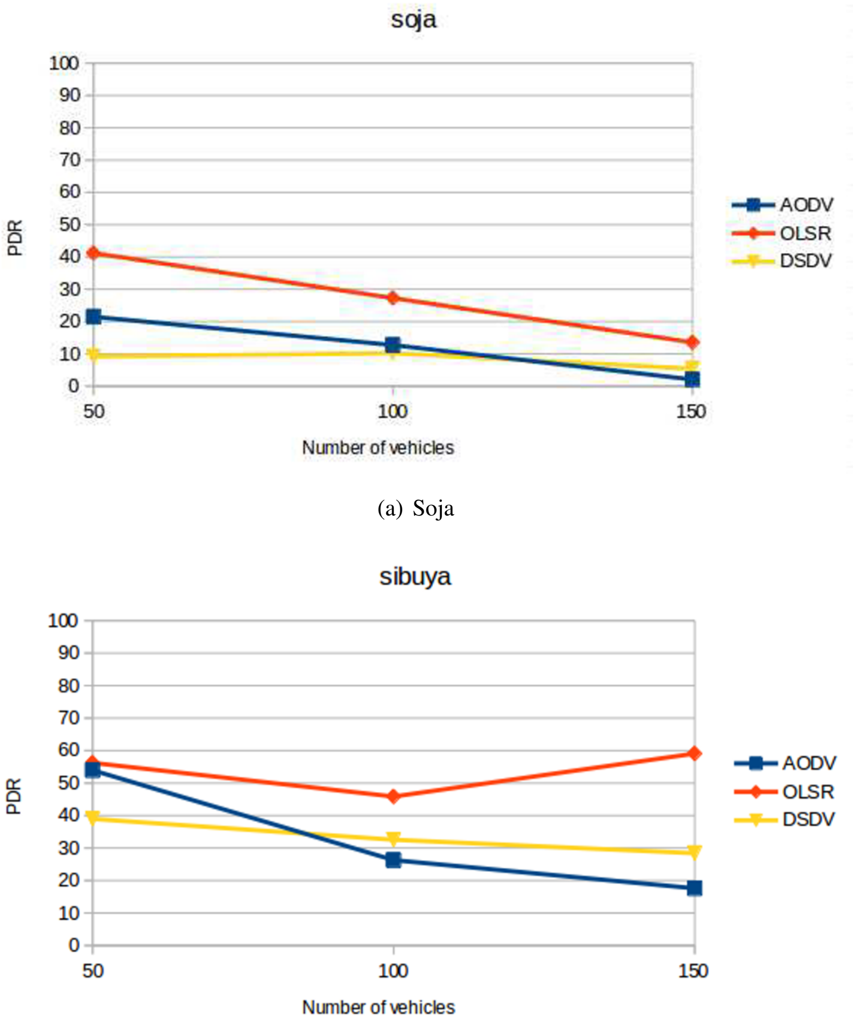
Fig. 3 Detailed road view of Soja town and Shibuya from SUMO Simulator.

Table 1 Simulation Parameters.

Parameter	Values
Area Size	$1000m \times 1000m$
Communication Distance	$250m$
Simulation Time	$300s$
Number of Vehicles	$50, 100, 150$
Maximum Moving Speed	$60km/h$
Routing Protocol	AODV, OLSR, DSDV

4 Simulation Results

In order to evaluate the performance of each scenario we used Packet Delivery Ratio (PDR) as a metric. In each of the start-end pairs of vehicle generation points, instead of choosing source and destination randomly among the moving vehicles, we put static nodes and sent three flows of CBR data transported over UDP. The reason we



(a) Soja

(b) Shibuya

Fig. 4 PDR Results for different number of vehicles and different routing protocols.

setup static nodes is because, when simulation starts, many random functions will decide the movement of vehicles. But, we wanted that at least the distance between source and destination that a packet should traverse in the network was fixed for all cases. If source and destination nodes were moving randomly, the performance would be affected considerably by the random functions of the simulator rather than the behavior of routing protocols. The results are shown in Fig. 4 as a graph and Table 2 and Table 3 as average values.

From results we can see that all protocols in all scenarios show a better performance when the vehicles move in Shibuya, compared to the case of Soja. An interesting finding concerns the performance decrease in both maps, when the number of vehicles increase. We expected an increase in performance, because by increasing the number of nodes there would be more routes available. But this was not the case in most of the scenarios, because the increase in number of vehicles created traffic and the vehicles were grouped in certain locations in the map. This was also affected by the difference is the road pattern difference between Soja and Shibuya. Shibuya has more roads and connections so the vehicles are more spread in the are and we see a PDR of $5-10\%$ in Soja and $30-40\%$ in Shibuya.

Table 2 Simulation Results (PDR Soja).

	AODV	OLSR	DSDV
50	21.48	41.19	9.16
100	12.71	27.26	10.20
150	2.05	13.53	5.39

Moreover, an increase in the number of routes and a high moving speed of vehicles, creates difficult and sometimes wrong route decisions especially from AODV, which is an on-demand routing protocol. In fact, OLSR shows an improvement for 150 vehicles in Shibuya, because OLSR MPR-based flooding mechanism makes the network respond faster to dynamic route changes.

Table 3 Simulation Results (PDR Shibuya).

	AODV	OLSR	DSDV
50	53.94	56.20	38.95
100	26.28	45.85	32.54
150	17.59	59.07	28.39

5 Conclusion and Future Works

In this paper we conducted simulations for VANETs in two Japanese cities with different road patterns. We analyze and compare three routing protocols: AODV, OLSR and DSDV. As evaluation metric, we used PDR and compared data for different scenarios. From the simulation results we found the following:

- When the number of vehicles increased, the PDR performance decreased in general, because of increased dynamism or routes.

- OLSR was able to handle dynamic routes better, because of the MPR-based flooding mechanism.
- In suburban areas like Soja, the lack of streets and connections creates traffic congestion, and brings performance decrease.

In our future works, we would like to investigate the behaviour of VANETs and VANET-specific routing protocols and VANET-specific applications. Moreover, we would like to test VANET in delay-tolerant applications.

References

1. J. Harri, F. Filali and C. Bonnet, Mobility Models for Vehicular Ad Hoc Networks: A Survey and Taxonomy, IEEE Communications Surveys & Tutorials, Vol. 11, No. 4, pp. 19-41, 2009.
2. F. Bai, N. Sadagopan, and A. Helmy, IMPORTANT: A Framework to Systematically Analyze the Impact of Mobility on Performance of Routing Protocols for Ad-hoc Networks, In Proc. of IEEE INFOCOM-2003, pp. 825-835, March-April 2003.
3. N. Sadagopan, F. Bai, B. Krishnamachari, and A. Helmy, PATHS: Analysis of Path Duration Statistics and Their Impact on Reactive MANET Routing Protocols, In Proc. of the 4-th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc-2003), pp. 245- 256, 2003.
4. M. Fiore, J. Harri, F. Filali, and C. Bonnet, Vehicular Mobility Simulation for VANETs, Proc. of the 40-th Annual Simulation Symposium (ANSS-2007), pp. 301-309, 2007.
5. L. Smith, R. Beckan, R. Anson, K. Nagel, and M. Williams, TRANSIMS: Transportation Analysis and Simulation System, In Proc. of the 5-th National Transportation Planning Methods Applications Conference, LA-UR 95-1664, April 1995.
6. M. Piorkowski, M. Raya, A. L. Lugo, M. Grossglauser, and J. P. Hubaux, Joint Traffic and Network Simulator for VANETs, In Proc. of Mobile and Information Communication Systems Conference (MICS-2006), October 2006, Available on line at: http://www.mics.ch/.
7. G. De Marco, M. Tadauchi and L. Barolli, Description and Analysis of a Toolbox for Vehicular Networks Simulation, In Proc. of IEEE ICPADS/PMAC-2WN-2007, Vol. 2, pp. 1-6, 2007.
8. K. Nagel and M. Schreckenberg, A Cellular Automaton Model for Freeway Traffic, Journal of Physics I France, Vol. 2, pp. 2221-2229, 1992.
9. E. Spaho, Gj. Mino, L. Barolli and F. Xhafa, Goodput and PDR Analysis of AODV, OLSR and DYMO Protocols for Vehicular Networks using CAVENET, International Journal of Grid and Utility Computing (IJGUC), Vol. 2, No. 2, pp. 130-138, 2011.
10. Open Street Map. Online at https://www.openstreetmap.org/.
11. eWorld framework. Online at http://eworld.sourceforge.net/.
12. SUMO User Documentation. Online at http://www.sumo.dlr.de/userdoc/.
13. Network Simulator version 3. Online at https://www.nsnam.org/.
14. C. Perkins, E. Belding-Royer and S. Das, ”Ad hoc On- Demand Distance Vector (AODV) Routing”, IETF RFC 3561 (Experimental), 2003.
15. T. Clausen and P. Jacquet, ”Optimized Link State Routing Protocol (OLSR)”, IETF RFC 3626, October 2003.
16. C. E. Perkins and P. Bhagwat, ”Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers”, In Proc. of ACM SIGCOMM Conference on Communications Architectures, Protocols and Applications, pp. 234-244, 1994.

Error Performance of Successive Interference Cancellation Schemes in MIMO Systems

Sangjoon Park[1], Kangbin Yim[2], Byeong-Gwon Kang[3]
[1] Department of Information & Communication Engineering, Wonkwang University, Iksan, Jeon buk 54538, Republic of Korea
sjpark24@wku.ac.kr
[2] Department of Information Security Engineering, Soonchunhyang University, Asan, Choongnam 31538, Korea
yim@sch.ac.kr
[3] Department of Information and Communication Engineering, Soonchunhyang University, Asan, Choongnam 31538, Republic of Korea (Corresponding author)
bgkang@sch.ac.kr

Abstract. In this paper, the error performance of various successive interference cancellation (IC) schemes is analyzed in spatially multiplexed multiple-input multiple-output (MIMO) systems. First, we classify the successive IC scheme by their ordering strategy. In addition, considering the channel coding is usually applied for the conventional wireless communication systems, we apply the channel coding to the spatially multiplexed MIMO systems and evaluate the block error rate (BLER) as the error performance of the systems. Using numerical simulations, the error performance of IC schemes is shown according to various system parameters, e.g., the number of transmit and receive antennas, the utilized linear filter, and the MIMO fading channel characteristics. In this way, this paper analyses and compares the error performances of the successive IC in various perspectives.

1 Introduction

In multiplexed multiple-input multiple-output (MIMO) systems with spatial multiplexing, a number of symbols can be jointly transmitted and received by using multiple antennas at the transmitter and the receiver [1], [2]. Since a number of symbols are jointly utilized, a detection scheme that can accurately estimate each of the transmitted symbols is required for MIMO systems. Interference cancellation (IC) based detection schemes are usually known as the suboptimal detection schemes for MIMO systems [1]-[3]. The successive IC scheme, e.g., the vertical Bell laboratories layered space-time architecture [1], [2], is one of the most typical IC schemes designed for MIMO systems which sequentially eliminates the interference from transmitted symbols. At each detection stage in a successive IC scheme, each

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_64

transmitted symbol is selected for detection and cancellation of the current stage, which is repeatedly performed for all the transmitted symbols. That is, a successive IC scheme repeatedly performs the detection stages until all the transmitted symbols are detected.

The successive IC schemes can be classified by the ordering strategy used for to select the symbol in each detection stage [1]-[3]. In specific, the symbol selection at each detection stage of the successive IC scheme is determined by the ordering of the transmitted symbols with a certain strategy. The most well-known ordering strategy for the successive IC scheme is based on the signal-to-interference-plus-noise ratio (SINR) of the symbol after the linear filtering, which is called the post-processing SINR (PSINR) ordering in the sequel. In addition to the PSINR based ordering, the ordering can also be done by using the initially received SINR (RSINR), which is called the RSINR ordering in the sequel. Finally, the symbol selection can be done by a random order, which is called the random ordering in the sequel.

The error performance of the successive IC schemes can be affected by many system properties and related parameters. For examples, the use of the channel coding scheme [4], e.g., the convolutional coding, can greatly impact the cancellation accuracy of the successive IC scheme. In addition, the system environments such as the number of transmit and receive antennas, the utilized linear filter, and the MIMO fading channel characteristics can also impact the error performance of the successive IC schemes. However, there have been little efforts to investigate the error performance of various successive IC schemes with various system configurations and parameters.

In this paper, the error performance of various successive IC schemes is analyzed and compared in spatially multiplexed MIMO systems. Three types of successive IC schemes are considered in this paper, i.e., the successive IC scheme with the post-processing SINR ordering, received SNR ordering, and random ordering. Since we consider the use of the channel coding for the spatially multiplex MIMO systems, the block error rate (BLER) is utilized for the error performance of the systems. Using numerical simulations, the error performance of IC schemes is shown according to various system parameters, i.e., the number of transmit and receive antennas, the utilized linear filter, and the MIMO fading channel characteristics.

2 System Model

We consider a spatially multiplexed MIMO system with N transmit and M receive antennas. For simplicity, we assume the system model that uses only one transmit signal vector for each transmission block, although the transmission block in the following numerical simulations can have a number of transmit signal vectors. Let $\mathbf{s} = [s(1), \ldots, s(N)]^T$ denote the N x 1 transmit signal vector that satisfies $E[\mathbf{s}\mathbf{s}^H] = \mathbf{I}_N$ with the N x N identity matrix $\mathbf{I}_N$, where the superscripts T and H denote the transpose and the conjugate-and-transpose operators, respectively. Also, let $\mathbf{r} = [r(1), \ldots, r(M)]^T$ denote the M x 1 receive signal vector. Then, the input-output relationship, i.e., the relationship between the transmit signal vector $\mathbf{s}$ and the receive signal vector $\mathbf{r}$ can be written as

$$\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{n}. \tag{1}$$

In (1), **H** is the $M \times N$ MIMO channel matrix between the transmitter and the receiver for the transmit signal vector **s** and the receive signal vector **r**. Also, **n** is the $M \times 1$ zero-mean additive white Gaussian noise (AWGN) vector with the covariance matrix $E[\mathbf{n}\mathbf{n}^H] = \sigma^2\mathbf{I}_M$.

3 Successive IC Operations

Based on the input-output relationship in (1), the reception procedures for the successive IC scheme are performed for a transmit signal vector. Since each transmit signal vector includes N transmit symbols, we assume that the number of the detection stages for the successive IC scheme is equal to N.

Let $\mathbf{r}_k$ and $\mathbf{H}_k$ denote the receive signal vector and channel matrix for the kth detection stage of the successive IC scheme, respectively. At the initial stage, $\mathbf{r}_1$ and $\mathbf{H}_1$ are set to **r** and **H**. respectively. Also, let $a(k)$ and $b(k)$ denote the indices of the same transmit symbol in **H** and $\mathbf{H}_k$ (indices of the corresponding columns in **H** and $\mathbf{H}_k$) that will be detected and cancelled at the kth detection stage, respectively. Then, at the beginning of the kth stage, the estimate of $s(a(k))$, $f(a(k))$, can be obtained as

$$f(a(k)) = [\mathbf{G}_k\mathbf{r}_k]_{b(k)}. \tag{2}$$

In (2), $\mathbf{G}_k$ denotes the $N \times M$ zero-forcing (ZF) or minimum mean-squared-error (MMSE) linear filtering matrix and $[\mathbf{G}_k\mathbf{r}_k]_{b(k)}$ denotes the $b(k)$th element of the vector $\mathbf{G}_k\mathbf{r}_k$.

Let $d(a(k))$ denote the detected transmit symbol by (2) from the modulation constellation set using $f(a(k))$. After $d(a(k))$ is detected at the kth stage, if $k < N$, then the IC operation is executed for the next detection stage. This operation can be written as

$$\mathbf{r}_{k+1} = \mathbf{r}_k - [\mathbf{H}]_{a(k)}d(a(k))_). \tag{3}$$

In (3), $[\mathbf{H}]_{a(k)}$ is the $a(k)$th column of **H**. After (3) is executed, the submatrix of $\mathbf{H}_k$ including all columns of $\mathbf{H}_k$ except the $b(k)$th column is set to $\mathbf{H}_{k+1}$. This is the end of the kth detection stage.

As shown above, the reception procedures for the successive IC scheme can be changed by using different $a(k)$ and $b(k)$, which are determined by the applied ordering strategy. Therefore, it is predicted that the ordering strategy has a great impact on the error performance of the successive IC scheme in spatially multiplexed MIMO system. This will be verified in detail in the following section.

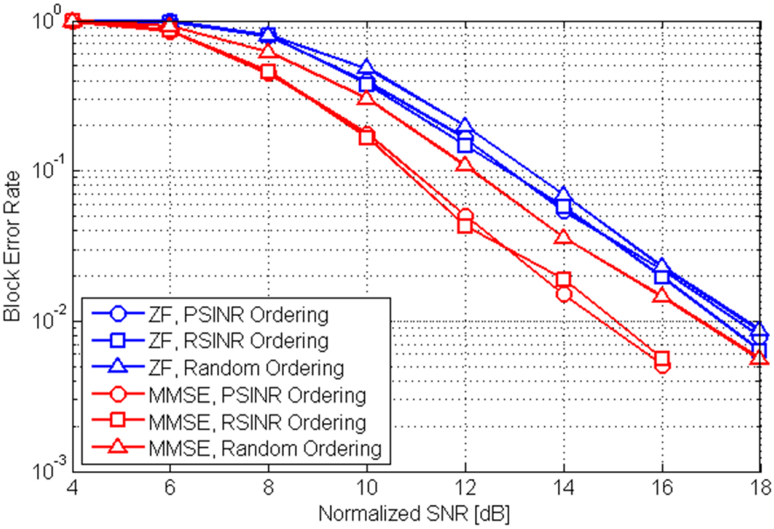


Fig. 1. Average BLER performance of the successive IC schemes in 2x2 MIMO systems under independent Rayleigh fading channel.

4 Numerical Simulation Results

In this section, the numerical simulation results of the successive IC schemes for the various system configurations are shown. As explained, we consider three successive IC schemes: PSINR ordering based successive IC scheme, RSINR ordering based successive IC scheme, and random ordering based success IC scheme. The numbers of data bits and coded bits in each codeword for a transmission block are 200 and 400, respectively, and a rate 1/2 convolutional code with a constraint length of 7 and code generator polynomial of 177 and 133 (in octal numbers) is considered at the transmitter. Quadrature phase shift keying (QPSK) modulation is considered throughout the simulations, and a hard decision Viterbi decoder is considered at the receiver.

As the channel model, we consider the MIMO independent Rayleigh fading channel and the MIMO block Rayleigh fading channel. In the MIMO independent Rayleigh fading channel, each transmit signal vector has the channel response independent with those for the other channel response. Meanwhile, in the MIMO block Rayleigh fading channel, every transmit signal vector in a transmission block has the same channel response, which is independent with those for transmit signal vectors in a different transmission block.

Fig. 1 shows the average BLER performance of the successive IC schemes in 2x2 MIMO systems under the independent Rayleigh fading channel. It is shown in Fig. 1 that the PSINR ordering based successive IC scheme and the RSINR ordering based successive IC scheme achieve almost the same average BLER performance in 2x2

MIMO system under the independent Rayleigh fading channel, while the BLER performance of the random ordering based successive IC scheme is inferior to that of the other schemes. It is worthwhile to mention that this phenomenon is observed for both ZF and MMSE linear filters. Since there are only two transmit symbols per transmit signal vector in 2x2 MIMO systems, the effects of the ordering on the error performance are not significantly observed in Fig. 1.

Fig. 2 shows the average BLER performance of the successive IC schemes in 2x2 MIMO systems under the independent Rayleigh fading channels. Unlike the BLER performances observed in Fig. 1, it is shown in Fig. 2 that the PSINR ordering based successive IC scheme outperforms the RSINR ordering based successive IC scheme. In addition, the SNR gain of the PSINR ordering based successive IC scheme over the RSINR ordering based successive IC scheme is more dominant with the MMSE linear filter than the ZF linear filter. In a spatially multiplexed MIMO system, the number of symbols used to perform the ordering is increased with the number of transmit antennas. Therefore, as the number of transmit antennas increases, the effects of the ordering on the error performance can be significantly observed. Furthermore, as the number of receive antennas increases, the attainable diversity and array gains are also increased, which enables the accurate ordering for the successive IC scheme. Therefore, as shown in Figs. 1 and 2, the performance gap between the successive IC schemes become more significant in 4x4 MIMO system than 2x2 MIMO system. Meanwhile, the random ordering based successive IC scheme shows the worst error performance for both 2x2 and 4x4 MIMO systems regardless of the utilized linear filter.

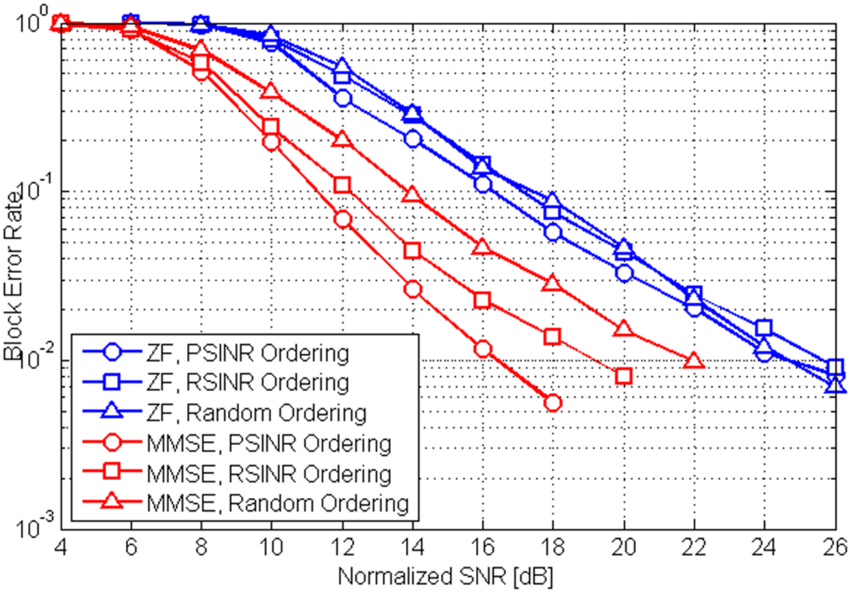


Fig. 2. Average BLER performance of the successive IC schemes in 4x4 MIMO systems under independent Rayleigh fading channel

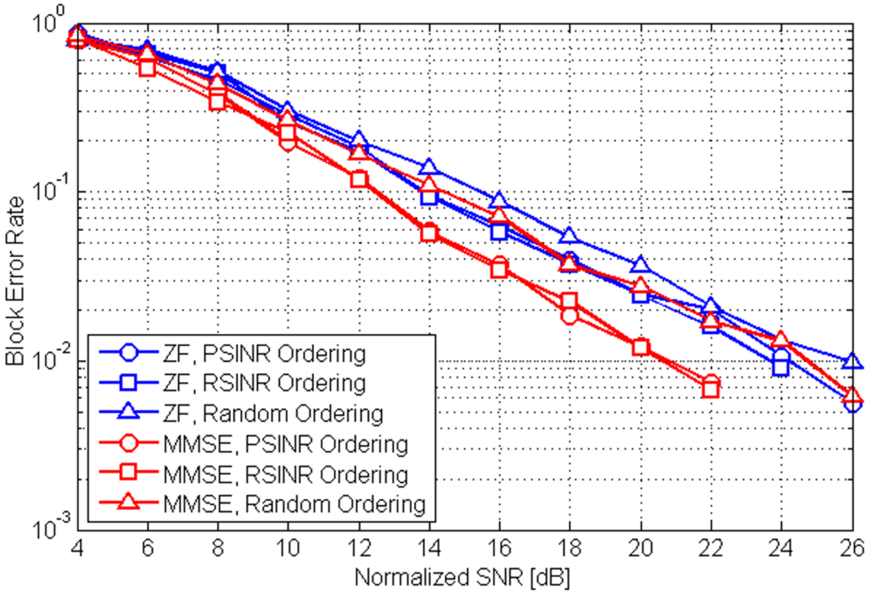


Fig. 3. Average BLER performance of the successive IC schemes in 2x2 MIMO systems under block Rayleigh fading channel

Finally, Fig. 3 shows the average BLER performance of the successive IC schemes in 2x2 MIMO systems under the block Rayleigh fading channels. Similar to the BLER performances under the independent Rayleigh fading channels observed in Fig. 1, the PSINR ordering based successive IC scheme and the RSINR ordering based successive IC scheme show the similar BLER performance, while the random ordering based successive IC scheme shows the worst error performance. That is, due to the limited number of transmit and receive antennas, the well-designed successive IC schemes, e.g., PSINR & RSINR ordering based successive IC schemes, achieve the similar detection accuracy, although the poorly-designed successive IC scheme, e.g., random ordering based success IC scheme, shows the degraded performance in spite of the limited number of antennas. Although the relative performance characteristics of the successive IC schemes are similar in both Figs. 1 (independent Rayleigh fading channel) and 3 (block Rayleigh fading channel), due to the lack of time diversity, the successive IC schemes under the block Rayleigh fading channel show the degraded performance compared with the successive IC scheme under the independent Rayleigh fading channel.

5 Conclusions

In this paper, the error performance of various successive IC schemes was numerically analyzed in spatially multiplexed MIMO systems. First, we classify the successive IC scheme by their ordering strategy. Numerical results showed that the effects of the ordering on the BLER performance becomes significant as the number of transmit and receive antennas increases. Also, it was shown that the performance gap between the successive IC schemes becomes significant with the MMSE linear filter than with the ZF linear filter. Finally, it was observed that the successive IC schemes under the block Rayleigh fading channel show the degraded BLER performance than those under the independent Rayleigh fading channel. In this way, this paper compared the error performances of the successive IC schemes in various perspectives.

In addition to the system parameters tested in this paper, the other system parameters, e.g., the block length, the coding rate, the modulation order, etc., can impact the error performance of the successive IC schemes in a spatially multiplexed MIMO system. The effects of the other system parameters can be investigated in a future work.

References

1. Paulraj, A., Nabar, R., and Gore, D.: Introduction to Space-Time Wireless Communications. Cambridge University Press (2008)
2. Cho, Y., Kim, Jaekwon., Yang, W., and Kang, C.: MIMO-OFDM Wireless Communications with MATLAB. John Wiley and Sons (2010)
3. Wu, J., Zhong, J., Cai, Y.. Zhao, M., and Zhang, W.:New Detection Algorithms Based on The Jointly Gaussian Approach and Successive Interference Cancelation for Iterative MIMO Systems. International Journal of Communication Systems, Vol. 27. (2014) 1964-1983
4. Moon, T.K.: Error Correction Coding: Mathematical Methods and Algorithms. John Wiely and Sons (2005)

A Study on the Classification of Common Vulnerabilities and Exposures using Naïve Bayes

Sarang Na, Taeeun Kim, and Hwankuk Kim

Security R&D Team 2
Korea Internet & Security Agency
Seoul, Republic of Korea
{no.1.nasa,tekim31,rinyfeel}@kisa.or.kr

Abstract. National Vulnerability Database (NVD) provides publicly known security vulnerabilities called Common Vulnerabilities and Exposures (CVE). There are a number of CVE entries, although, some of them cannot provide sufficient information, such as vulnerability type. In this paper, we propose a classification method of categorizing CVE entries into vulnerability type using naïve Bayes classifier. The classification ability of the method is evaluated by a set of testing data. We can analyze CVE entries that are not yet classified as well as uncategorized vulnerability documents.

Keywords: Vulnerability analysis, Common Vulnerabilities and Exposures (CVE), Common Weakness Enumeration (CWE), naïve Bayes classifier, document classification.

1 Introduction

Security vulnerabilities inherent in software packages can be easily exploited for conducting malicious manipulations. Attackers can identify vulnerable Web services by using an Internet-wide scanning tool and conduct malicious behavior [1]. Thus, security experts must be aware of known vulnerabilities and be able to quickly cope with threats.

National Vulnerability Database (NVD) provides Common Vulnerabilities and Exposures (CVE) entries to easily share publicly known security vulnerabilities [2]. CVE system provides a reference-method for the security vulnerabilities of released software packages. A CVE entry is composed of vulnerability overview, Common Vulnerability Scoring System (CVSS), references, Common Platform Enumeration (CPE), and Common Weakness Enumeration (CWE).

There are over 77,000 CVE entries, but they cannot provide satisfactory vulnerability information that is available in the vulnerability overview or reference sites. In particular, the CWEs that identify types of vulnerabilities are provided for only 57.6% of all CVE entries (Figure 1).

To find out which type of vulnerability is explained by a CVE entry, it is possible to use the vulnerability overview of each CVE entry and thus insufficient information may be supplemented. The overview text is structuralized in a certain form, but as the

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_65

structure is not perfectly the same, we need to convert this text into an appropriate form through data preprocessing.

In this paper, we propose a classification method to categorize the CVE entries that predicts the vulnerability types explained by text documents. We collect CVE entries from NVD and generate a vulnerability classification model based on naïve Bayes. By using this method, we are able to classify CVE entries into vulnerability category, i.e., CWE.

The remainder of the paper is organized as follows. In Section II, we review the related work. In Section III, we explain the proposed method. Experimental results are described in Section IV and this paper concludes in Section V.

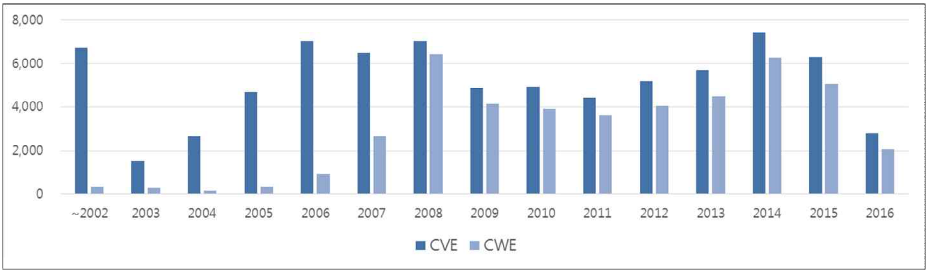


Fig. 1. Number of CVE entries and CWEs by year.

2 Related Work

Genge and Enăchescu [3] proposed a vulnerability assessment tool for devices connected to the Internet identified by Shodan [4]. This tool simply matched CVE entries to the corresponding devices without additional processing of the CVE entries.

Chang et al. [5] analyzed vulnerability trends using CVE entries from 2007 to 2010. They showed the vulnerability trends through vulnerability frequency and severity by using the CVEs and CVSS scores, respectively. As vulnerabilities that occurred in that year were additionally discovered and registered until now, it is different than the security trends that were analyzed in the past.

Neuhaus and Zimmermann [6] used topic models to analyze vulnerability trends, such as vulnerability types of CVE entries until 2009. The authors found 28 topics in CVE entries by using Latent Dirichlet Allocation (LDA) and assigned LDA topics to CWEs. The precision and recall of LDA is good at some CWEs, such as CWE-79 and CWE-89, but is poor at other categories, such as CWE-310 and CWE-94.

Guo and Wang [7] modeled CVE vulnerabilities based on ontology and used it to analyze similar vulnerabilities. We refer the structure of CVE vulnerabilities to be used in the classification model in this research.

Li et al. [8] analyzed the characteristics of bugs and classified the bugs through text classification and information retrieval techniques. In this paper, we use naïve Bayes classifier to categorize vulnerabilities.

3 Methodology

We propose a vulnerability classification method using the overview texts of CVE entries. Figure 2 shows the conceptual map of our method. We collect CVE xml files from NVD and parse each CVE entry. Next, we conduct the preprocessing that removes useless words, such as stop words and software product names in the selected overview in order to improve the accuracy of the classification model. Finally, we generate a vulnerability classification model and categorize CVE entries.

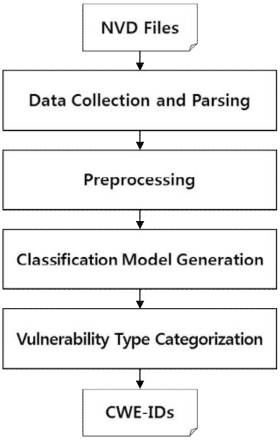
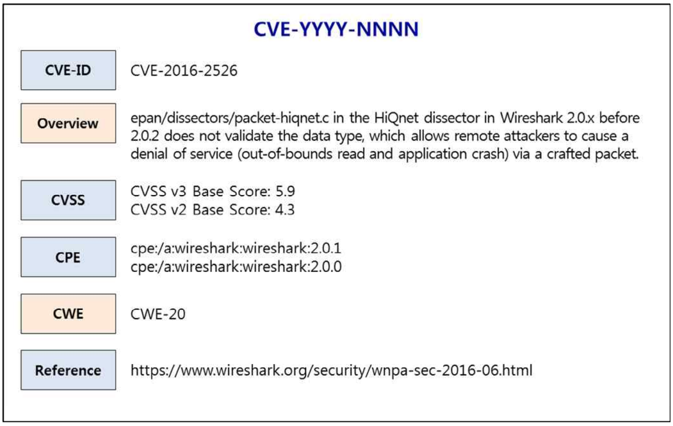

NVD Files
Data Collection and Parsing
Preprocessing
Classification Model Generation
Vulnerability Type Categorization
CWE-IDs

Fig. 2. Conceptual map of the proposed method.

3.1 Preprocessing of the overview text

A CVE entry consists of an identifier number (CVE-ID), overview, CVSS, CPE names and CWE as shown in Figure 3. The overview text is composed of the following in general:

- ‘place where a vulnerability was discovered’
- (in) ‘related software product names’
- (when) ‘conditions of the vulnerability occurrence’
- (allow) ‘attacker type’
- (to) ‘results of attack’
- (via) ‘means of attack’
- (aka) ‘vulnerability title in the reference site’
- (a different vulnerability than) ‘other CVE-IDs’



CVE-YYYY-NNNN

CVE-ID	CVE-2016-2526
Overview	epan/dissectors/packet-hiqnet.c in the HiQnet dissector in Wireshark 2.0.x before 2.0.2 does not validate the data type, which allows remote attackers to cause a denial of service (out-of-bounds read and application crash) via a crafted packet.
CVSS	CVSS v3 Base Score: 5.9 CVSS v2 Base Score: 4.3
CPE	cpe:/a:wireshark:wireshark:2.0.1 cpe:/a:wireshark:wireshark:2.0.0
CWE	CWE-20
Reference	https://www.wireshark.org/security/wnpa-sec-2016-06.html

Fig. 3. Example of a CVE entry.

We use a part of the overview text about the 'results of attack,' 'means of attack' and 'vulnerability title in the reference site,' for which vulnerability types can be identified. To this end, the character string after 'to' is selected from the 'allow ~ to' phrase and the character string is split before the 'a different vulnerability than' phrase. For example, in Figure 3, the "cause a denial of service (out-of-bounds read and application crash) via a crafted packet." character string is selected. All words in the modified texts are converted to lower case.

In the next step, we remove some of the words regardless of vulnerability classification, such as stop words like 'because' and 'with,' and related product information. Through this phase, the common words unrelated to the vulnerability category can be removed.

3.2 Generating a vulnerability classification model

The training/testing dataset consists of the vulnerability overview text and vulnerability type identifier (CWE-ID). We generate a vulnerability classification model with released CVE entries by using naïve Bayes classifier and evaluate the model using other CVE entries that were not used for the classification.

4 Results

4.1 Experimental data

We collected 77,885 CVE entries between 1999 and 2016 from NVD for the classification and its evaluation. Among them, CVE entries with more than 1,000 identified CWEs were used as experimental data (Table I). In this paper, we classified CWE-119 and CWE-79, which have the greatest number of identified CWEs, and the top 10 CWEs in terms of CWE frequency. These CWEs are described in Table II. We used 500 CVE entries for each type of CWE in the classification and evaluation, respectively. In this experiment, thus, a total of 10,000 CVE entries randomly selected regardless of publish date were used.

Table 1. Number of identified CWEs.

CWE	Frequency	CWE	Frequency
119	7,048	362	390
79	6,559	284	345
264	4,762	16	295
89	4,189	254	217
20	3,919	78	203
200	2,790	17	168
399	2,710	134	164
310	2,270	19	117
94	2,078	77	67
22	1,888	345	25
189	1,364	74	23
352	1,166	18	5
287	1,002	199	3
255	633	21	2
59	424	361	1

Table 2. Top 10 CWEs used in CVE entries.

CWE	CWE Name
119	Buffer Errors
79	Cross-Site-Scripting
264	Permissions, Privileges, and Access Control
89	SQL Injection
20	Input Validation
200	Information Leak / Disclosure
399	Resource Management Errors
310	Cryptographic Issues
94	Code Injection
22	Path Traversal

4.2 Experimental results

In the first experiment, we classified the experimental dataset into two categories, CWE-119 and CWE-79 with the greatest number of CWEs. The accuracy of the classification model was 99.8%. In the next experiment classifying the top 3 CWEs and top 5 CWEs, the accuracy was 95.1% and 84.5%, respectively. In the last experiment classifying the top 10 CWEs, the accuracy was 75.5%. The precision and recall values for each experiment are shown in Table III. In the top 5 CWEs and top 10 CWEs experiments, as there are different CWEs with a similar vulnerability overview, some CVE entries were wrongly classified as similar vulnerabilities.

Table 3. Results of the classification experiment.

Type of the Experiment	Precision (%)	Recall (%)
Top 3 CWEs	95.2	95.3
Top 5 CWEs	84.2	84.5
Top 10 CWEs	75.0	75.0

5 Conclusion

We proposed a classification method using naïve Bayes to categorize CVE entries into vulnerability type and evaluated the classification ability of this method. We are planning to enhance the accuracy of the vulnerability classification model by conducting an in-depth study of different vulnerabilities made up of similar texts and advanced feature engineering. Eventually, we will analyze CVE entries that are not yet to be identified.

Acknowledgments. This work was supported by Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korea government (MSIP). (No. B0717-16-0099, IoT Security Vulnerabilities Search, Sharing, and Testing Technology Development)

References

1. Durumeric, Z., Adrian, D., Mirian, A., Bailey, M., Halderman, J.A.: A search engine backed by Internet-wide scanning. In: SIGSAC Conference on Computer and Communications Security (CCS'15), pp. 542-553. (2015)
2. National Vulnerability Database, https://nvd.nist.gov/.
3. Genge, B., Enăchescu, C.: ShoVAT, Shodan-based vulnerability assessment tool for Internet-facing services. In: Security and Communication Networks, pp. 1-19. (2015)
4. Shodan, https://www.shodan.io/.
5. Chang, Y.Y., Zavarsky, P., Ruhl, R., Lindskog, D.: Trend analysis of the CVE for software vulnerability management. In: IEEE International Conference on Privacy, Security, Risk, and Trust (PASSAT) and IEEE International Conference on Social Computing (SocialCom), pp. 1290-1293. (2011)
6. Neuhaus, S., Zimmermann, T.: Security trend analysis with CVE topic models. In: IEEE International Symposium on Software Reliability Engineering, pp. 111-120. (2010)
7. Guo, M., Wang, J.A.: An ontology-based approach to model common vulnerabilities and exposures in information security. In: American Society for Engineering Education (ASEE) Southeastern Section Conference. (2009)
8. Li, Z., Tan, L., Wang, X., Lu, S., Zhou, Y., Zhai, C.: Have things changed now?: An empirical study of bug characteristics in modern open source software. In: Workshop on Architectural and System Support for Improving Software Dependability (ASID), pp. 25-33. (2006)

A Study on The behavior-based Malware Detection Signature

Sungtaek OH, Woong Go, and Taejin Lee

KISA, R&D Security 1 Team,
Seoul, Korea
{angelrick,wgo,tjlee}@kisa.or.kr

Abstract. As smartphone are becoming more common, services using smartphones are becoming more pervasive too. Among them, as mobile banking transactions are increasing, payment fraud is also rapidly increasing. These services handle sensitive information, such as users' personal information and payment information, but as they have several security vulnerabilities, they are attacked by malicious apps. This paper proposes a method of deriving malicious app detection signatures based on the behavior information, obtained by analyzing malicious apps collected through several application distribution channels, and these signatures will be used for analysis of variants of malicious apps and development of rule-based malicious app detection systems.

Keywords: android, malware, static analysis, dynamic analysis, detection, malicious signature.

1 Introduction

Recently the domestic penetration rate of smartphones sharply increased from 65% to 86.4% (2012 to 2015). Also, the number of mobile banking transactions using smart devices like smartphones and tablets was 41.01 million a day on average in the second quarter of 2015. Mobile customers account for all Internet banking transactions. Mobile banking transactions, such as mobile micropayment and banking, are quite pervasive. Along with the growth of the mobile banking market, however, mobile payment fraud like Smishing is sharply increasing, and 68% of all mobile malicious apps circulated around the world include payment-related malicious behavior. To prevent damages due to these payment fraud malicious apps, it is necessary to analyze the malicious apps in circulation, check whether they are malicious apps or not, and block them.

Currently, these malicious apps can be analyzed either by statically analyzing the manifest information and the DEX file, which are obtained by deconstructing the APK file, or by installing and executing the APK file in the analysis device and dynamically analyzing. In case of static analysis, however, if code obfuscation was applied to the source code, normal analysis is difficult. In case of dynamic analysis,

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_66

there are several analysis methods, including activity-based analysis and user interaction-based analysis.

This paper proposes a method of analyzing characteristics that can be used for detecting malicious apps by analyzing Android APIs, system calls and internal strings, which can be collected through analysis of the behavior of apps to classify behavior information that occurs in malicious apps and normal apps, and the author is planning to use such characteristics to develop technologies for analyzing similarity to detect variants of malicious apps and rule-based technologies for detecting malicious apps.

2 Related Work

2.1 Static analysis technology

Static analysis refers to analyzing the manifest information and DEX file obtained by deconstructing the APK file, and Figure 1 illustrates the analysis flow chart for the static analysis system developed in this study. The manifest information in the APK file contains a lot of information about apps, and can decompile the DEX file for static analysis at the source code level.

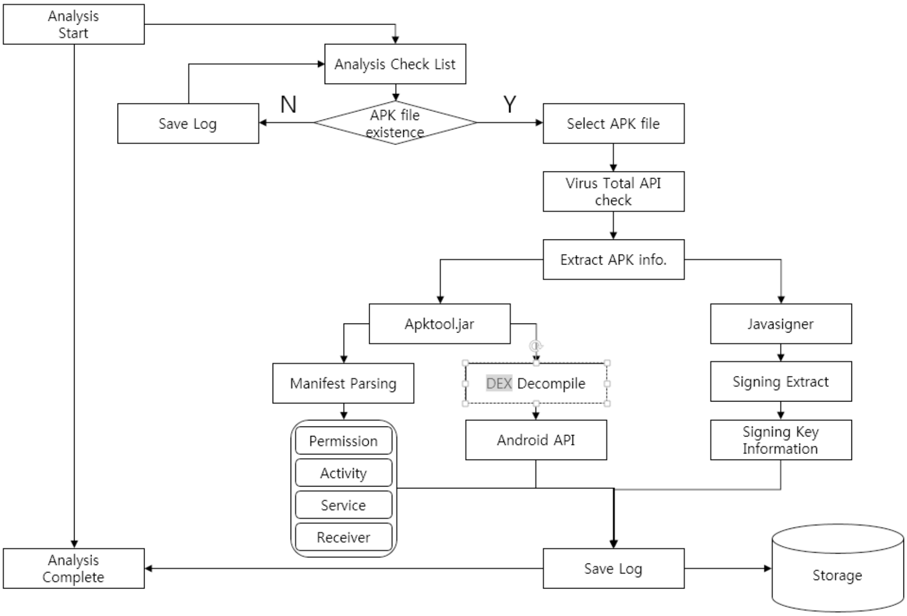


Fig. 1. Static Analysis Flow chary

2.2 Dynamic analysis technology

Dynamic analysis refers to installing and executing the APK file in the actual device or emulator and tracking and recording how it behaves as it is difficult to accurately know whether it is malicious or not simply by statically analyzing the APK file, and Figure 2 illustrates the analysis flow chart for the dynamic analysis system developed in this study. It tracks Android APIs, System Calls and Network information.

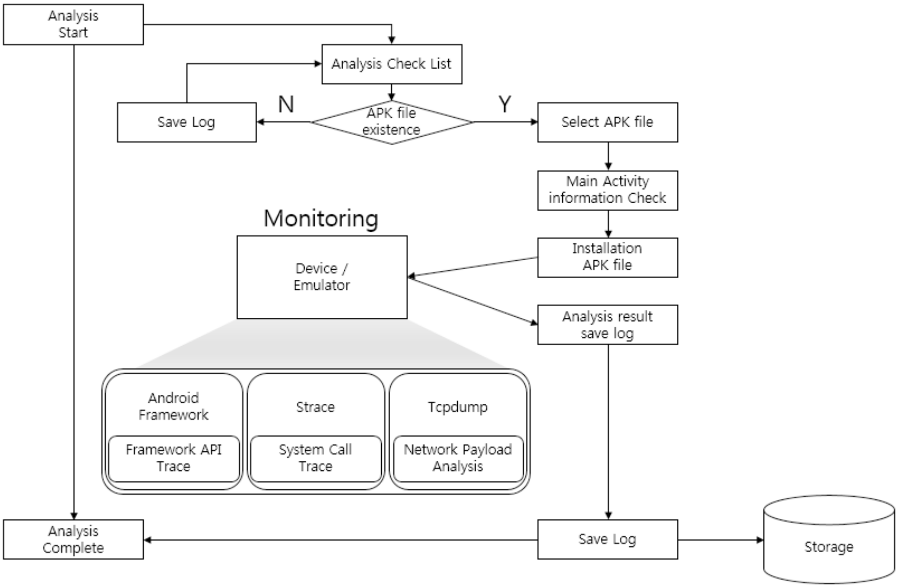


Fig. 2. Dynamic Analysis Flow chart

3 Analysis of malicious behavior based on data

This section used existing researches and studies, and malicious app analysis data to analyze the malicious behavior of malicious apps related to payment fraud. Android APIs and System Calls are analyzed, and characteristics that can distinguish malicious apps from normal apps can be extracted and used for detecting malicious apps.

3.1 Analysis of the characteristics of malicious behaviors based on research and investigation

Android APIs [Table 1] and System Calls [Table 2], used in existing malicious apps related to payment fraud are as shown below, and the malicious behaviors analyzed in this paper and their roles are summarized below.

Table 1. Common Android API List

API	Analysis	Role	Malicious
sendTextMessage	Send Smishing messages / Send acquired information	Send messages	High
getMessageBody	Take text messages	Get the message body	Low
getOriginatingAddress	Get originating numbers	Get originating numbers	High
createFromPdu	Create malicious text messages / Take raw text messages	Convert Raw PDU	High
abortBroadcast	Abort text message receive broadcast	sequential delivery broadcast	High
setRingerMode	Hide text message receive	Switch the ringer mode	High
setComponent EnabledSetting	Hide Smishing app icons	Switch the package component status	High
android.app.action. ADD_DEVICE_ADMIN	Limit/interrupt deletion of Smishing apps	Register the Admin mode (system app)	High
ContactsContract$Contacts	Take user contacts	Contacts management provider	Low
lockNow	Hide/lock by force the current status	Switch the lock screen	High
wipeData	Delete Smishing app evidence	Delete user data	High
getLatitude	Take user location information	Latitude	Low
getLongitude	Take user location information	Longitude	Low
getLastKnownLocation	Take user location information	Saved user’s last location	Low
getAccounts	Take user account information	Provide information on all accounts registered in the device	Low

Table 2. Common System Call List

Command	Analysis	Role	Malicious
killProcess	End certain apps by force	N/A	Low
getAsciiBytes	Convert character strings	N/A	Low
shell	Execute shell commands	N/A	High
copyclassdex	Copy binary files executed externally	N/A	Low
copyfile	Copy tool execution files	N/A	High
copylib	Copy obfuscation / malicious libraries	N/A	High
chmod	Give authority to execute malicious files	File authority change tool	Low
shield	Obfuscation of malicious codes	Obfuscation tool	Average
classloader	Load binary files executed externally by force	N/A	Low

3.2 Analysis of the characteristics of malicious behaviors based on malicious app analysis data

This paper analyzed collected malicious apps, characterized them, and checked their differences from general apps. General apps are those apps downloaded from the Google Play Store.

Actually, the APIs used by malicious apps were compared with those used by general apps, and the result of comparing the APIs of malicious apps and general apps related to malicious behavior is shown in Figure 3. As illustrated in the figure, malicious apps and general apps showed difference in rate of API used. Also, the comparison of malicious behaviors other than Android APIs are as shown in the figure 4. The key malicious behaviors of malicious apps, such as the taking of NPKI information and the aborting of processes, were analyzed.

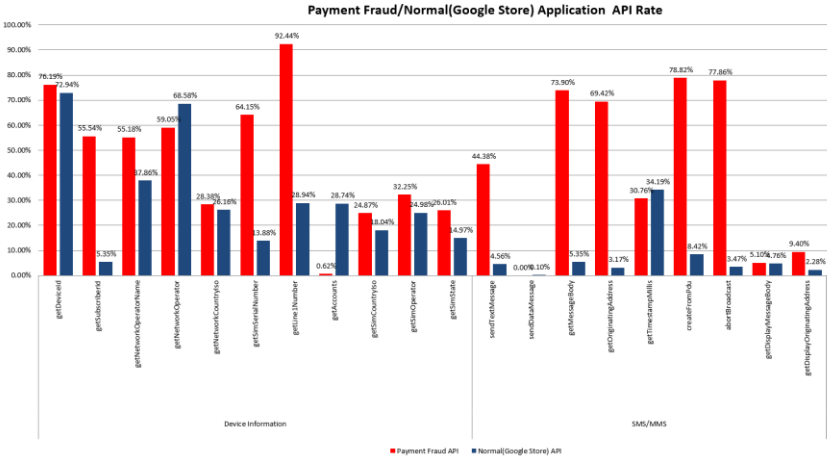


Fig. 3-1. Result of Fraud/Normal API rate

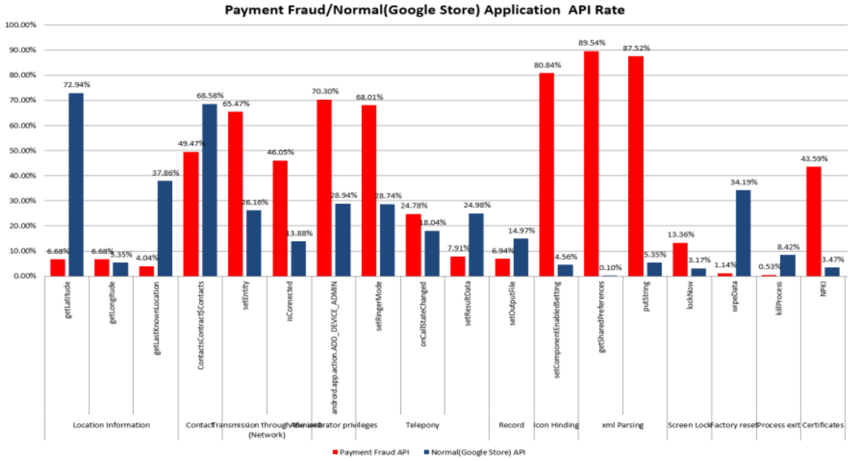


Fig. 3-2. Result of Fraud/Normal API rate

4 Result

The characteristics of malicious behaviors extracted based on the analysis data are as shown in [Table 3]. The Android APIs, System Calls and Strings that can be used to calculate the risks of malicious behaviors are summarized, and with weight given to the information frequently appearing in malicious apps and non-malicious apps, it seems that they can be used to calculate the risks for detection of malicious apps.

Table 3. Extract malicious behavior

API/String/function	Analysis	Role	Malicious
application/vnd.android.package-archive	Remote installation of apps	in-app installer parameter	High
getOriginatingAddress	Take originating numbers	Get originating numbers	High
ContactsContract$Contacts	Take user contacts	Contacts management provider	Low
NPKI	Take certificates	Certificate name	High
getLatitude	Take user location information	Latitude	Low
getLongitude	Take user location information	Longitude	Low
getLastKnownLocation	Take user location information	Saved user's last location	Low
getDeviceId	Take device information	Get device ID	Low
getSubscriberId	Take device information	Get IMSI codes	High
getNetworkOperator	Take device information	Get network operator codes	Low
getNetworkCountryIso	Take device information	Get country code	Low
getSimSerialNumber	Take device information	Get SIM card serial number	High
getLine1Number	Take device information	Get device phone number	High
ContactsContract$CommonDataKinds $Phone.CONTENT_URI	Take user information	Contacts access constant	Average
abortBroadcast	Abort text message receive broadcast	sequential delivery broadcast	High
setRingerMode	Hide text message receive	Switch the ringer mode	High
setComponentEnabledSetting	Hide Smishing app icons	Switch the package component status	High

android.app.action.ADD_DEVICE_ADMIN	Limit/interrupt deletion of Smishing apps	Register the Admin mode (system app)	High
lockNow	Hide/lock by force the current status	Switch the lock screen	High
killProcess	Abort certain apps	N/A	Low
setResultData	Execute certain behavior	Broadcast parameter	Average
getPackageInfo	Access installed app information	Installed app manifest information	Low
android.intent.action.NEW_OUTGOING_CALL	Execute certain behavior	Broadcast parameter	Low
incoming_number	Malicious codes Characteristics Character string	Get incoming numbers	Average
createFromPdu	Create malicious texts/ take raw texts	Convert the Raw PDU	High
content://sms/	Take text messages	Get and query text messages	High
getMessageBody	Take text messages	Get the message body	Low
getDisplayMessageBody	Take text messages	Get text message/ email body	Low
getDisplayOriginatingAddress	Take text messages	Get text message originating numbers	High
sendTextMessage	Send smishing messages / Send acquired information	Send messages	High

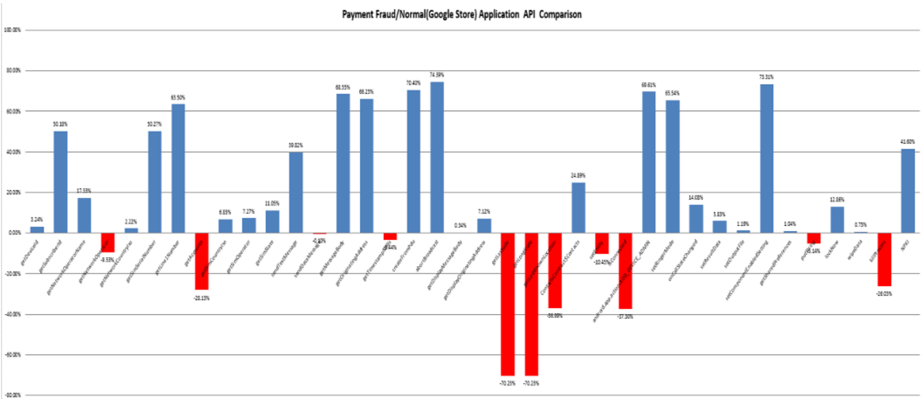


Fig. 4. Comparison API rate

5 Conclusion

This paper conducted an association analysis of the malicious behaviors of malicious apps, which were studied previously to generate the basic data for implementing the automated malicious app detection system, and the malicious behaviors of malicious apps that were actually collected by the system, and derived malicious app signatures that can be used for detecting malicious apps. Currently, the author is using this signature information to develop and test an algorithm for detecting malicious apps. Currently available security systems related to malicious apps use the app installation APK file to check if there is any malicious code, or use authority information, and the static analysis data of the source codes to detect malicious apps, or conduct a dynamic analysis of network packets when they are executed to detect malicious apps. This paper analyzed the domestic and overseas research and investigation data that has been confirmed so far, and the malicious apps that were actually collected, and conducted an association analysis of the derived malicious behaviors to derive the characteristics of the malicious behaviors of malicious apps (malicious app signatures), and if this information is utilized, it will be possible to detect malicious apps more accurately. In the future, the author is planning to use the malicious app signatures, derived in this paper, to develop technologies and systems for detecting malicious apps.

Acknowledgments. This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIP) (No.R0132-16-1004, Development of Profiling-based Techniques for Detecting and Preventing Mobile Billing Fraud Attacks).

References

1. Kwang-hwi Ahn: A Design and Implementation of Detection System by using Pattern based Behavior Analysis in Android Environment. Graduate School of Soongsil University (2013)
2. Jin-Sik Yun: Malware Detection Technique of Android-based Smartphone using Static Analysis. Graduate School of Korea Maritime University (2010)
3. A-Young Lee: A Study on Realtime Detecting Smishing on Cloud Computing Environments. Graduate School of Soongsil University (2014)
4. Patrick P.K., CHAN, Wen-Kai Song: Static Detection of Android Malware by using permissions and API Calls. Proceedings of the 2014 International Conference on Machine Learning and Cybernetics (2014)

Part IV
The 7-th International Workshop on Methods, Analysis and Protocols for Wireless Communication (MAPWC-2016)

Design and Implementation of a Simulation System Based on Genetic Algorithm for Node Placement in Wireless Sensor and Actor Networks

Kosuke Ozera, Tetsuya Oda, Donald Elmazi and Leonard Barolli

Abstract A Wireless Sensor and Actor Network (WSAN) is a group of wireless devices with the ability to sense physical events (sensors) or/and to perform relatively complicated actions (actors), based on the sensed data shared by sensors. In order to provide effective sensing and acting, a coordination mechanism is necessary among sensors and actors. This coordination can be distributed-local coordination among the actors or centralized coordination from a remote management unit, usually known as sink in Wireless Sensor Networks (WSNs). In this work, we propose a simulating system based on Rust for actor node placement problem in WSAN, while considering different aspects of WSANs including coordination, connectivity and coverage. We describe the implementation and show the interface of the simulation system.

1 Introduction

Wireless Sensor Networks (WSNs) can be defined as a collection of wireless self-configuring programmable multihop tiny devices, which can bind to each other in an arbitrary manner, without the aid of any centralized administration, thereby dynamically sending the sensed data to the intended recipient about the monitored phenomenon [1].

Kosuke Ozera, Tetsuya Oda, Leonard Barolli
Department of Information and Communication Engineering, Fukuoka Institute of Technology (FIT), 3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811–0295, Japan, e-mail: kosuke.o.fit@gmail.com, oda.tetsuya.fit@gmail.com, barolli@fit.ac.jp

Donald Elmazi
Department of Information and Communication Engineering, Fukuoka Institute of Technology (FIT), 3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811–0295, Japan, e-mail: donald.elmazi@gmail.com

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_67

Wireless Sensor and Actor Networks (WSANs), have emerged as a variation of WSNs. WSANs are capable of monitoring physical phenomenons, processing sensed data, making decisions based on the sensed data and completing appropriate tasks when needed. WSAN devices deployed in the environment are sensors able to sense environmental data, actors able to react by affecting the environment or have both functions integrated [2]. For example, in the case of a fire, sensors relay the exact origin and intensity of the fire to actors so that they can extinguish it before spreading in the whole building or in a more complex scenario, to save people who may be trapped by fire.

Unlike WSNs, where the sensor nodes tend to communicate all the sensed data to the sink1 by sensor-sensor communication, in WSANs, two new communication types may take place. They are called sensor-actor and actoractor communications. Sensed data is sent to the actors in the network through sensor-actor communication. After the actors analyse the data, they communicate with each other in order to assign and complete tasks. To provide effective operation of WSAN, is very important that sensors and actors coordinate in what are called sensor-actor and actor-actor coordination. Coordination is not only important during task conduction, but also during networkfs selfimprovement operations, i.e. connectivity restoration [3, 4], reliable service [5], Quality of Service (QoS) [6, 7] and so on.

Actor-Actor (AA) coordination helps actors to choose which actor will lead performing the task (actor selection), how many actors should perform and how they will perform. Actor selection is not a trivial task, because it needs to be solved in real time, considering different factors. It becomes more complicated when the actors are moving, due to dynamic topology of the network.

In this paper, we propose and implement a simulation system for actor node placement in WSAN. The system is based on Genetic Algorithm (GA). We describe the implementation and show the interface of the simulation system.

The remainder of the paper is organized as follows. In Section 2, we describe the basics of WSANs including architecture and research challenges. In Section 3, we present the overview of GA. In Section 4, we show the description and design of the simulation system. Simulation results are shown in Section 5. Finally, conclusions and future work are given in Section 6.

2 WSAN

2.1 WSAN Architectures

The main functionality of WSANs is to make actors perform appropriate actions in the environment, based on the data sensed from sensors and actors. When important data has to be transmitted (an event occurred), sensors may transmit their data back to the sink, which will control the actorsf tasks from distance, or transmit their data to actors, which can perform actions independently from the sink node.

Here, the former scheme is called Semi-Automated Architecture and the latter one Fully-Automated Architecture. Obviously, both architectures can be used in different applications. In the Fully-Automated Architecture are needed new sophisticated algorithms in order to provide appropriate coordination between nodes of WSAN. On the other hand, it has advantages, such as low latency, low energy consumption, long network lifetime [2], higher local position accuracy, higher reliability and so on.

2.2 WSAN Challenges

Some of the key challenges in WSAN are related to the presence of actors and their functionalities.

- Deployment and Positioning: WSAN are heterogeneous networks [8], where actors and sensors have different processing powers, mobility abilities and functionalities. Thus, at the moment of node deployment, algorithms must consider to optimize the number of sensors and actors and their initial positions based on application [9, 10].
- Architecture: The main functionality of WSANs is to make actors perform appropriate actions in the environment, based on the data sensed from sensors and actors [11, 12]. When important data has to be transmitted (an event occurred), sensors may transmit their data back to the sink, which will control the actorsf tasks from distance or transmit their data to actors, which can perform actions independently from the sink node.
- Real-Time: The purpose of using WSANs in most of the applications is mainly related to their ability to react independently to situations where human intervention is physically difficult or time-restricted [6, 13]. In other words, there are a lot of applications that have strict real-time requirements. In order to fulfill them, real-time limitations must be clearly defined for each application and system.
- Coordination: Unlike WSN, where sensors coordinate with each-other to send data to the sink [14], in WSAN, sensor-actor coordination occurs as well, because all sensed data controls actorfs behavior. Also, actor-actor coordination is important in cases when actors collaborate on performing tasks together. In order to provide effective sensing and acting, a distributed local coordination mechanism is necessary among sensors and actors [12, 15].
- Power Management: Similar to energy-constrained WSNs [16], in WSANs sensors have limited power supplies, which limits the network lifetime. Actors have more powerful power supplies but their functionalities are more sophisticated, so they spend more energy when completing complicated tasks. Thus, WSAN protocols should be designed with minimized energy consumption for both sensors and actors [7, 17]. It should be also kept in mind, that energy consumption requirements differ, depending on application of WSAN.
- Mobility: In WSANs, nodes, especially actors can be mobile [18]. For example, robots used in industrial monitoring sites or flying drones aver a disaster recovery

area. Therefore, protocols developed for WSANs should support the mobility of nodes, [4, 19, 20], where dynamic topology changes, unstable routes and network isolations are present.
- Self Healing: One of the main problems in mobile SelfOrganizing Networks (SON) is the high probability of node isolations during network runtime. An actor failure may lead to partitioning the network and thus hinder the fulfillment of the application requirements. Many works have been done on connectivity restoration, by using actors ability to move without using much energy [3, 4]. Actors may also be specialized to carry extra energy supplies, in order to charge sensors or other actors in the network in cases of emergency.
- Scalability: Smart Cities are emerging fast and WSAN, with its practical functions of simultaneous sensing and acting, are a key technology. The heterogeneity is not limited and most of the systems will continue to grow together with cities. In order to keep the functionality of WSAN applicable, scalability should be considered when designing WSAN protocols and algorithms. Data replication, clustering and so on, can be used in order to support growing networks [10, 20].

2.3 Node Placement Problems and Their Applicability to WSANs

Node placement problems have been long investigated in the optimization field due to numerous applications in location science (facility location, logistics, services, etc.) and classification (clustering). In such problems, we are given a number of potential facilities to serve to costumers connected to facilities aiming to find locations such that the cost of serving to all customers is minimized [21]. In traditional versions of the problem, facilities could be hospitals, polling centers, fire stations serving to a number of clients and aiming to minimize some distance function in a metric space between clients and such facilities. One classical version of the problem is that of p-median problem, defined as follows.

Definition1 : Given a set $\mathscr{F}$ of m potential facilities, a set $\mathscr{U}$ of n users, a distance function $d : \mathscr{U} \to \mathscr{F}$, and a constant $p \leq m$, determine which p facilities to open so as to minimize the sum of the distances from each user to its closest open facility.

The problem, which is known for its intractability, has many application not only in location science but also in communication networks, where facilities could be servers, routers, etc., offering connectivity services to clients. In WSANs node provide network connectivity services to events. The good performance and operability of WSANs largely depends on placement of nodes in the geographical deployment area to achieve network connectivity, stability and user coverage. The objective is to find an optimal and robust topology of the nodes network to support connectivity services to events.

Facility location problems are thus showing their usefulness to communication networks, and more especially from WSANs field. In a general setting, location models in the literature have been defined as follows. We are given:

(a) a universe $\mathscr{U}$, from which a set $\mathscr{E}$ of event input positions is selected;

(b) an integer, $\mathscr{N} \geq 1$, denoting the number of facilities to be deployed;
(c) one or more metrics of the type $d : \mathscr{U} \times \mathscr{U} \rightarrow \mathscr{R}_{+}$, which measure the quality of the location; and,
(d) an optimization model.

The optimization model takes in input the universe where facilities are to be deployed, a set of client positions and returns a set of positions for facilities that optimize the considered metrics. It should be noted that different models can be established depending on whether the universe is considered: (a) continuous (universe is a region, where clients and facilities may be placed anywhere within the continuum leading to an uncountably infinite number of possible locations); (b) discrete (universe is a discrete set of predefined positions); and, (c) network (universe is given by an undirected weighted graph; in the graph, client positions are given by the vertices and facilities may be located anywhere on the graph). For most formulations, node placement problems are shown to be computationally hard to solve to optimality and therefore heuristic and meta-heuristic approaches are useful approaches to solve the problem for practical purposes.

3 Overview of GA

As an approach to global optimization, GA have been found to be applicable to optimization problems that are intractable for exact solutions by conventional methods [22, 23]. It is a set-based search algorithm, where at each iteration it simultaneously generates a number of solutions. In each iteration, a subset of the current set of solutions is selected based on their performance and these solutions are combined into new solutions. The operators used to create the new solutions are survival, where a solution is carried to the next iteration without change, crossover, where the properties of two solutions are combined into one, and mutation, where a solution is modified slightly. The same process is then repeated with the new set of solutions. The crossover and mutation operators depend on the representation of the solution, but not on the evaluation of its performance. They are thus the same even though the performance is estimated using simulation. The selection of solutions, however, does depend on the performance. The general principle is that high performing solutions (which in genetic algorithms are referred to as fit individuals) should have a better change of both surviving and being allowed to create new solutions through crossover. The simplest approach is to order the solutions $J(\theta_{[1]}) \leq J(\theta_{[2]}) \leq \ldots \leq J(\theta_{[n]})$, and only operate on the best solutions. If a strict selection of the top k solutions were required, this would complicate the issue significantly in the simulation optimization context, and considerable simulation effort would have to be spent to obtain an accurate ordering of the solutions.

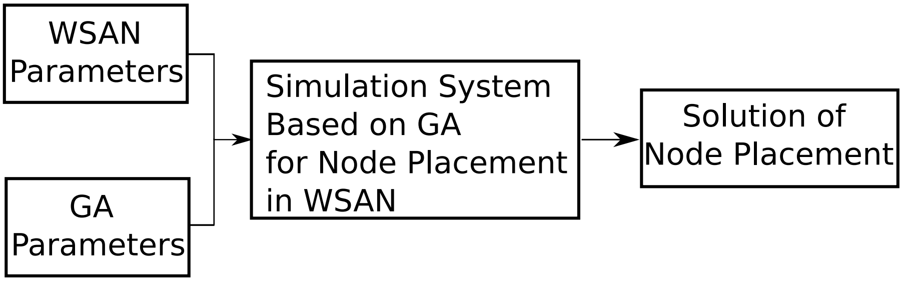

WSAN Parameters
GA Parameters
Simulation System Based on GA for Node Placement in WSAN
Solution of Node Placement

Fig. 1 Simulation system structure.

4 Design and Implementation of Actor Node Placement Simulation System

In this section, we present design and implementation of a simulation system based on GA for node placement in WSANs. The simulation system structure is shown in Fig. 1. The proposed simulating system is based on Rust [24, 25]. Rust is a system programming language focused on three goals: safety, speed, and concurrency [26]. Rust supports a mixture of programming styles: imperative procedural, concurrent actor, object-oriented and functional.

Our system can generate instances of the problem using different distributions of events, sensor nodes and actor nodes. For the network configuration, we use: distribution of events, number of events, number of sensor nodes, number of actor nodes, area size, radius of communication range and radius of sensing range. For the GA parameter configuration, we use: number of independent runs, GA evolution steps, population size, crossover probability, mutation probability, initial placement methods, selection methods.

We explain in details the GA operations in following.

Selection Operator

As selection operator, we use roulette-wheel selection [22, 23, 27]. In roulette-wheel selection, each individual in the population is assigned a roulette wheel slot sized in proportion to its fitness. That is, in the biased roulette wheel, good solutions have a larger slot size than the less fit solutions. The roulette wheel can obtain a reproduction candidate.

Crossover Operator

The crossover operators are the most important ingredient of GAs. Indeed, by selecting individuals from the parental generation and interchanging their *genes*, new individuals (descendants) are obtained. The aim is to obtain descendants of better quality that will feed the next generation and enable the search to explore new regions of solution space not explored yet.

There exist many types of crossover operators explored in the evolutionary computing literature. It is very important to stress that crossover operators depend on the chromosome representation. This observation is especially important for the WSAN nodes problem, since in our case, instead of having strings we have a area of nodes located in a certain positions. The crossover operator should thus take into account the specifics of WSAN nodes encoding. We have considered the following crossover operator, called *intersection operators* (denoted `CrossRegion`, hereafter), which take in input two individuals and produce in output two new individuals.

Mutation Operator

Mutation operator is one of the GA ingredients. Unlike crossover operators, which achieve to transmit genetic information from parents to offsprings, mutation operators usually make some small local perturbation of the individuals, having thus less impact on newly generated individuals.

Crossover is "a must" operator in GA and is usually applied with high probability, while mutation operators when implemented are applied with small probability. The rationale is that a large mutation rate would make the GA search to resemble a random search. Due to this, mutation operator is usually considered as a secondary operator.

In the case of WSAN node placement, the matrix representation is chosen for the individuals of the population, in order to keep the information on WSAN nodes positions, events positions, links among nodes and links among nodes and events. The definition of the mutation operators is therefore specific to matrix-based encoding of the individuals of the population. We consider *SingleMutate* mutation operator which is a move-based operator It selects a WSAN node in the problem area and moves it to another cell of the problem area.

5 Visualization Interface

In Fig. 2 is shown visualization interface of implemented simulation system. We show a simulation scenario where the number of actor nodes is 4, the number of sensor nodes is 16, and the number of events is 48. For simulation, we also consider the communication range of sensor and actor nodes, and sensing range of sensor and actor nodes.

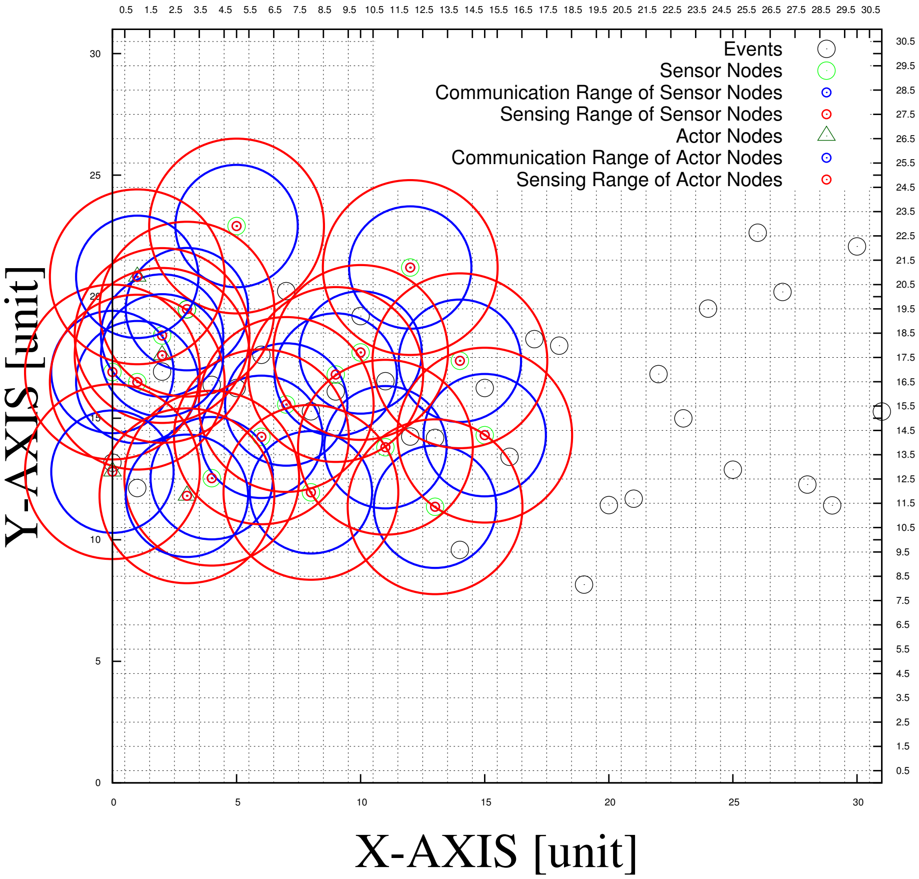


Fig. 2 Visualization interface.

6 Conclusions

In this work, we designed and implemented a simulation system based on GA for actor node placement in WSANs. We presented the implementation of the proposed simulation system and have shown also the interface and a simulation scenario. In the future, we would like to make extensive simulations for different simulation scenarios.

References

1. I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey", Computer Networks (Elsevier), vol. 38, no. 4, pp. 393-422, 2002.
2. I. F. Akyildiz and I. H. Kasimoglu, "Wireless sensor and actor networks: Research challenges", Ad Hoc Networks Journal (Elsevier), vol. 2, no. 4, pp. 351-367, October 2004.

3. N. Haider, M. Imran, N. Saad, and M. Zakariya, “Performance analysis of reactive connectivity restoration algorithms for wireless sensor and actor networks”, IEEE Malaysia International Conference on Communications (MICC-2013), Nov 2013, pp. 490-495.
4. A. Abbasi, M. Younis, and K. Akkaya, “Movement-assisted connectivity restoration in wireless sensor and actor networks”, IEEE Transactions on Parallel and Distributed Systems, vol. 20, no. 9, pp. 1366-1379, Sept 2009.
5. X. Li, X. Liang, R. Lu, S. He, J. Chen, and X. Shen, “Toward reliable actor services in wireless sensor and actor networks”, IEEE 8th International Conference on Mobile Adhoc and Sensor Systems (MASS), Oct 2011, pp. 351-360.
6. K. Akkaya and M. Younis, “Cola: A coverage and latency aware actor placement for wireless sensor and actor networks”, IEEE 64th Conference on Vehicular Technology (VTC-2006), pp. 1-5, Sept 2006.
7. J. Kakarla and B. Majhi, “A new optimal delay and energy efficient coordination algorithm for wsan”, IEEE International Conference on Advanced Networks and Telecommuncations Systems (ANTS), Dec 2013, pp. 1-6.
8. J. Kruger, D. Polajnar, and J. Polajnar, “An open simulator architecture for heterogeneous self-organizing networks”, Canadian Conference on Electrical and Computer Engineering 2006 (CCECE f06), May 2006, pp. 754-757.
9. M. Akbas and D. Turgut, “Apawsan: Actor positioning for aerial wireless sensor and actor networks”, IEEE 36th Conference on Local Computer Networks (LCN), Oct 2011, pp. 563-570.
10. M. Akbas, M. Brust, and D. Turgut, “Local positioning for environmental monitoring in wireless sensor and actor networks”, IEEE 35th Conference on Local Computer Networks (LCN), Oct 2010, pp. 806-813.
11. P. Lameski, E. Zdravevski, A. Kulakov, and D. Davcev, “Architecture for wireless sensor and actor networks control and data acquisition”, in Distributed Computing in Sensor Systems and Workshops (DCOSS), 2011 International Conference on, June 2011, pp. 1-3.
12. T. Melodia, D. Pompili, V. Gungor, and I. Akyildiz, “Communication and coordination in wireless sensor and actor networks”, EEE Transactions on Mobile Computing, vol. 6, no. 10, pp. 1126-1129, October 2007.
13. V. Gungor, O. Akan, and I. Akyildiz, “A real-time and reliable transport (rt2) protocol for wireless sensor and actor networks”, IEEE/ACM Transactions on Networking, vol. 16, no. 2, pp. 359-370, April 2008.
14. D. Estrin, R. Govindan, and S. K. J. Heidemann, “Next century challenges: scalable coordination in sensor networks”, in Proceedings of the ACM/IEEE International Conference on Mobile Computing and Networking (Mobicom-1999), Seattle, Washington, August 1999, pp. 263-270.
15. L. Mo and B. Xu, “Node coordination mechanism based on distributed estimation and control in wireless sensor and actuator networks”, Journal of Control Theory and Applications, vol. 11, no. 4, pp. 570-578, 2013. [Online]. Available: http://dx.doi.org/10.1007/s11768-013-2266-9
16. A. Goldsmith and S. Wicker, “Design challenges for energyconstrained ad hoc wireless networks”, IEEE Wireless Communications, vol. 9, no. 4, pp. 8-27, 2002.
17. K. Selvaradjou, N. Handigol, A. Franklin, and C. Murthy, “Energy-efficient directional routing between partitioned actors in wireless sensor and actor networks”, Communications, IET, vol. 4, no. 1, pp. 102-115, January 2010.
18. M. Haenggi, “Mobile sensor-actuator networks: opportunities and challenges”, in Proceedings of 7th IEEE International Workshop on Cellular Neural Networks and Their Applications (CNNA-2002), 2002, pp. 283-290.
19. T. Melodia, D. Pompili, and I. Akyldiz, “Handling mobility in wireless sensor and actor networks”, IEEE Transactions on Mobile Computing, vol. 9, no. 2, pp. 160-173, Feb 2010.
20. H. Nakayama, Z. Fadlullah, N. Ansari, and N. Kato, “A novel scheme for wsan sink mobility based on clustering and set packing techniques”, IEEE Transactions on Automatic Control, vol. 56, no. 10, pp. 2381-2389, Oct 2011.

21. T. Oda, A. Barolli, F. Xhafa, L. Barolli, M. Ikeda, M. Takizawa, “WMN-GA: A Simulation System for WMNs and Its Evaluation Considering Selection Operators”, In Press, Journal of Ambient Intelligence and Humanized Computing (JAIHC), Springer, Vol. 4, No. 3, pp. 323-330, June 2013.
22. J. H. Holland, “Adaptation in Natural and Artificial Systems”, University of Michigan Press, 1975.
23. D.E. Goldberg, “Genetic Algorithm in Search, Optimization and Machine Learning”, Addison-Wesley, Reading, MA, 1989.
24. “The Rust Programming Language”, https://www.rust-lang.org/.
25. “GitHub - rust-lang/rust: A safe, concurrent, practical language”, https://github.com/rust-lang/.
26. “’rust’ tag wiki - Stack Overflow”, http://stackoverflow.com/tags/rust/info/.
27. K. Sastry, D. Goldberg, G. Kendall, “Genetic Algorithms”, Search Methodologies - Introductory Tutorials in Optimization and Decision Support Techniques, pp. 97-125, 2005.

VegeShop Tool: A Tool for Vegetable Recognition Using DNN

Yuki Sakai, Tetsuya Oda, Makoto Ikeda and Leonard Barolli

Abstract Deep Learning also called Deep Neural Network (DNN) has a deep hierarchy that connect multiple internal layers for feature detection and recognition learning. In our previous work, we proposed vegetable recognition system which was based on Convolutional Neural Network (CNN). In this paper, we propose a tool called VegeShop for vegetable category recognition which is based on CNN. The user interface serves as e-commerce system for sellers and buyers using Android mobile device. The system can be accessed ubiquitously from any where. Moreover, our system can be applied also for other category recognition.

Key words: Deep Neural Network, Category Recognition, CNN.

1 Introduction

In recent years, Internet of Things (IoT) has attracted increased attention within the advanced technology industry in an effort to modernize and develop a more intelligent and reliable-based information system [30, 31]. IoT has been rapidly bringing a sea of technological changes in our daily lives to improve our life and more comfortable [32].

Yuki Sakai
Graduate School of Engineering, Fukuoka Institute of Technology (FIT),
3-30-1 Wajiro-higashi, Higashi-ku, Fukuoka 811-0295, Japan
e-mail: mgm15004@bene.fit.ac.jp

Tetsuya Oda, Makoto Ikeda and Leonard Barolli
Department of Information and Communication Engineering,
Fukuoka Institute of Technology,
3-30-1 Wajiro-higashi, Higashi-ku, Fukuoka 811-0295, Japan
e-mail: oda.tetsuya.fit@gmail.com, makoto.ikd@acm.org, barolli@fit.ac.jp

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_68

At Apple's annual Worldwide Developers Conference (WWDC-2016), they announced that next iOS 10 considers intelligent suggestions system based on deep learning using current location, calendar availability, contact information, recent addresses, and more. *Siri* will quickly grow into the role of an AI or a bot. Technologies to detect a specific object in images are expected to further expand to wide range of applications [8, 28, 15, 22, 34, 2, 4].

Wireless sensor networks are a wireless networks consisting of spatially distributed autonomous devices using sensors to cooperatively monitor physical or environmental condition [3, 13, 5, 35, 29].

In previous work, we proposed an object detection and tracking system which considers the mobility of objects [24]. The system was based on local feature extraction methods, such as Scale Invariant Feature Transform (SIFT) [18, 19, 20] and Speed Up Robust Features (SURF) [6, 21] methods for extracting the feature points. But, for recognizing the vegetable objects, these methods were not suitable.

In [25], we proposed a vegetable category recognition system considering Deep Neural Network (DNN). The image data used for Convolutional Neural Network (CNN) was set to eight kinds of vegetables. For evaluation, three different learning iterations was used.

In this paper, we propose we propose a tool for vegetable category recognition system which is based on CNN. The user interface serves as e-commerce system for sellers and buyers in Android mobile device.

The structure of the paper is as follows. In Section 2, Neural Networks are introduced. The proposed system design is shown in Section 3. In Section 4, we show the evaluation results. Finally, conclusions and future work are given in Section 5.

2 Neural Networks

Brain is a collection of a large number of nerve cells called neurons. Neurons receive signals through synapses located on the dendrites or membrane of the neuron. When the signals received are strong enough, the neuron is activated and emits a signal though the axon. For learning and identification of pattern, it is determined by the intensity of signal changes.

The complexity of real neurons is highly abstracted, but Artificial Neural Network (ANN) [11, 33, 26] has a biologically-inspired programming paradigm, which enables a computer to learn from observational data [7, 10]. The models inspired by ANN are:

- Convolutional Neural Network,
- Recurrent Neural Network,
- Deep Belief Network,
- Deep Boltzmann Machine.

2.1 Deep Neural Network (DNN)

Deep Learning also called Deep Neural Network (DNN) has a deep hierarchy that connect multiple internal layers for feature detection and representation learning. Representation learning is to learn how to express the extracting essential information from observation data in the real world. Feature extraction so far needs trial and error by artificially operations, however, Deep Learning uses a pixel level of the image as input value, and acquire the characteristic that is most suitable by learning, and identify it [16, 9]. The simplest kind of neural network is a single-layer perceptron network, which consists of a single layer of output, the inputs are fed directly to the outputs. In this way it can be considered the simplest kind of feed-forward network. It has become easy to learn by adopting the back propagation in a multilayer neural network. In this work, we use CNN to learn for a vegetable category recognition system.

2.2 Convolutional Neural Network (CNN)

Learning method in a CNN uses the back propagation model like an conventional multi-layer perceptron. Then, in order to update the weighting filter and coupling coefficient, CNN uses stochastic gradient descent. In this way, CNN recognize the optimized feature by using the convolutional and pooling operations [17, 27, 23, 14]. For the task of category recognition, Rectified Linear Units (ReLU) is used in CNN to speed up training.

Caffe [12] provide a complete toolkit for training, testing, fine-tuning, and deploying models, with well-documented examples for all of these tasks.

CNN have been successfully applied to object recognition. The network consists of a set of layers each of which contains one or more planes. Each unit in a plane receives input from a small neighborhood in the planes of the previous layer.

3 Object tracking system design

The structure of our proposed object detection and tracking system is shown in Fig. 1. The image processing part of the system runs on Windows7 (CPU: Intel Core i3 3.3GHz, GPU: GeForce GTX750 Ti 2GB, RAM: 12GB) equipped with Caffe. Caffe is a deep learning framework made with expression, speed, and modularity in mind. It is developed by the Berkeley Vision and Learning Center (BVLC), as well as community contributors and is popular for computer vision [12]. Monitoring system is composed of sensors and wireless camera. These devices are connected with Raspberry Pi. The monitoring system runs Linux Raspbian [1] with kernel 2.6. All experiments have been performed in indoor environment, within our departmental floor of size roughly 20m.

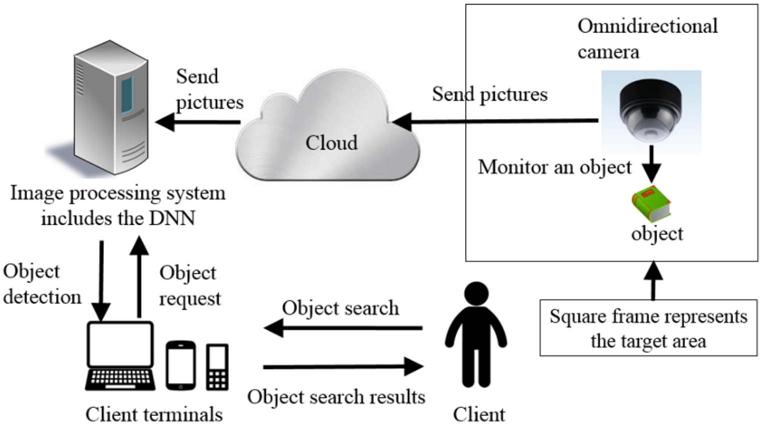


Fig. 1 Enhanced object detection and tracking system.

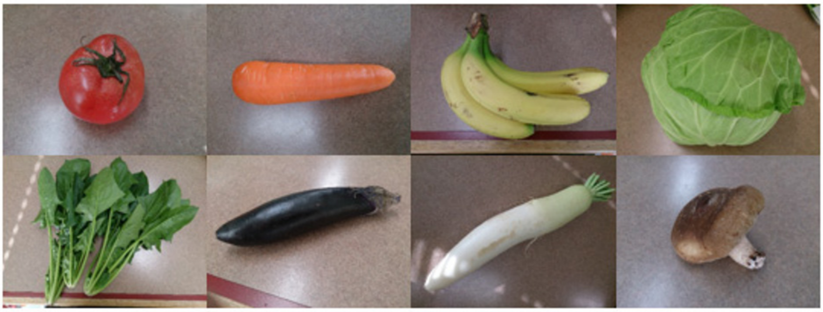
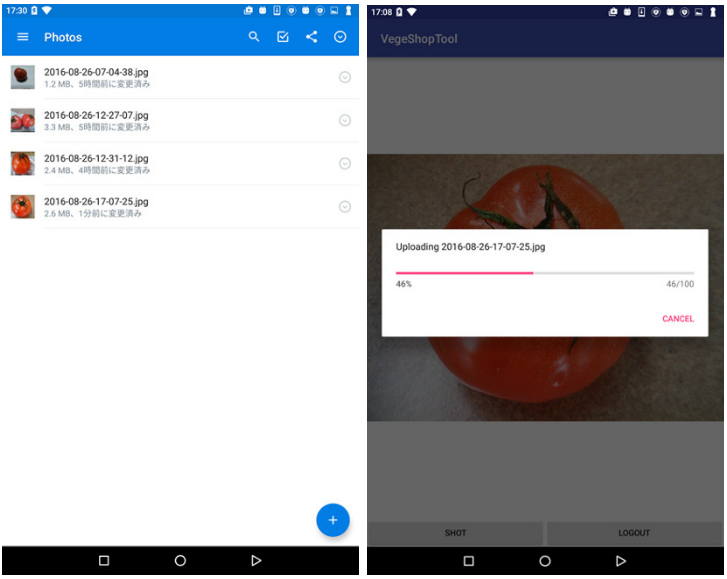


Fig. 2 Vegetables used for testing.

In order to recognize the objects from the database of image processing system includes the DNN, the images in the test set are compared to all images in the reference set by matching the respective keypoints. The object shown on the reference image with the highest number of matches with respect to the test image is chosen as the recognized object. Based on object detection characteristics and challenges, we consider the recognition ratio which is computed by the CNN algorithm. In [25], we have proposed an object category recognition within the DNN framework Caffe. For evaluation, we used image data of eight categories of vegetables (see Fig. 2).

4 Evaluation Results

Both seller and buyer use a mobile application called *VegeShop tool* to sale and purchase the vegetables. The tool offers a simple and powerful set of functions to



2016-08-26-07-04-38.jpg
1.2 MB、5時間前に変更済み

2016-08-26-12-27-07.jpg
3.3 MB、5時間前に変更済み

2016-08-26-12-31-12.jpg
2.4 MB、4時間前に変更済み

2016-08-26-17-07-25.jpg
2.6 MB、1分前に変更済み

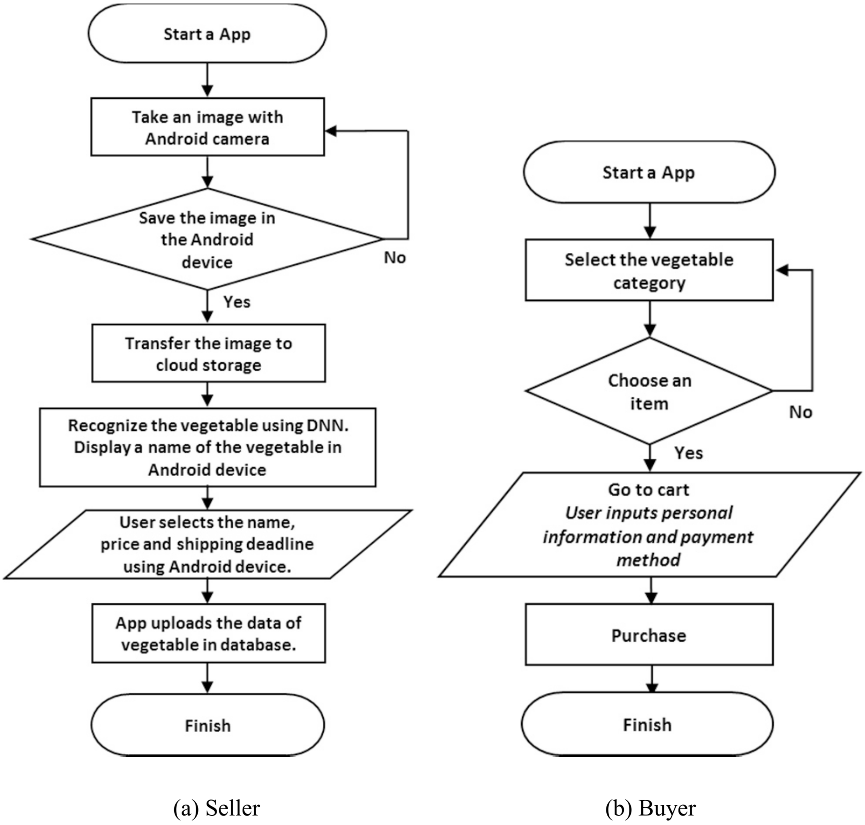
Uploading 2016-08-26-17-07-25.jpg
46% 46/100

(a) List of images (b) Upload function

Fig. 3 Snapshot of Android application.

manage the necessary features and data. It provides dynamic detection of camera features and controls (see Fig. 3). After seller takes an image of vegetables using VegeShop tool, they can select an image in Android mobile device to transfer the image in cloud storage (see Fig. 4). We use Dropbox as cloud storage. We can manage the files from any Android mobile devices or computer that’s connected to the Internet. Then, our proposed object detection and tracking system using DNN is proceeded. After recognition procedure, we see a list of vegetables in Android mobile device. In this way, seller can easily selects the vegetable name, price and other details using the mobile application. Finally, VegeShop tool uploads the data of vegetable on database in our system.

We implemented VegeShop also in website. Buyer uses VegeShop tool or VegeShop website to purchase a vegetable. The snapshot of VegeShop website is shown in Fig. 5. Our system provides the detail of vegetables for a buyer as follows: product name, seller name, shipping deadline, price and quantity. If user submits a favorite vegetable category list using VegeShop tool, they can receive a notification message via Twitter or E-mail.


Start a App
Take an image with Android camera
Save the image in the Android device
No
Yes
Transfer the image to cloud storage
Recognize the vegetable using DNN. Display a name of the vegetable in Android device
User selects the name, price and shipping deadline using Android device.
App uploads the data of vegetable in database.
Finish
Start a App
Select the vegetable category
Choose an item
No
Yes
Go to cart
User inputs personal information and payment method
Purchase
Finish

(a) Seller (b) Buyer

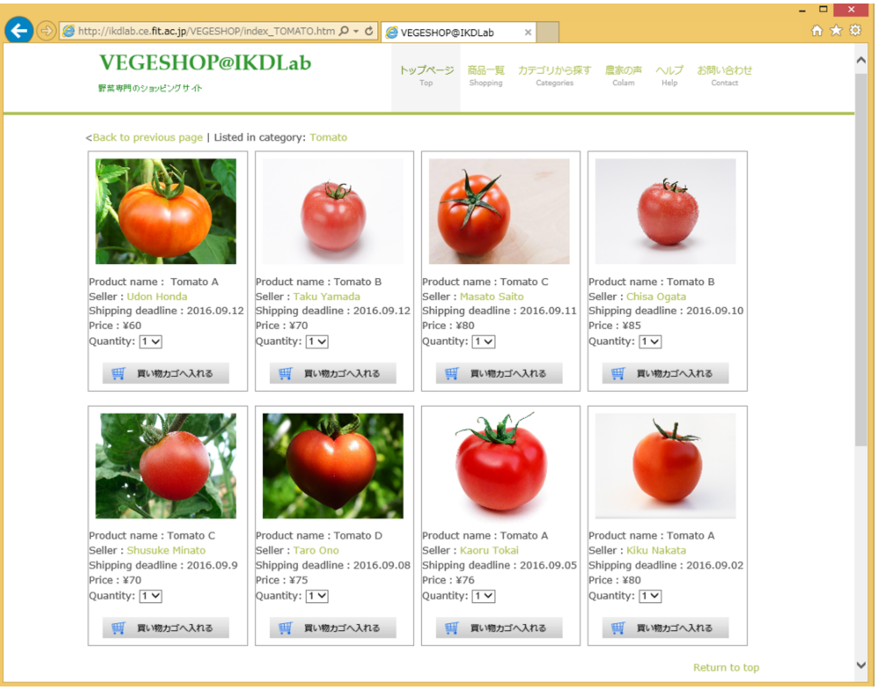
Fig. 4 Flowchart of our management system.

5 Conclusions

In this paper, we proposed a tool for vegetable category recognition system which is based on CNN. The user interface serves as e-commerce system for buyers and sellers using Android mobile device. The system can be accessed ubiquitously from any where. Moreover, our system can be applied also for other category recognition. In the future work, we will add another functions into the proposed VegeShop tool.

References

1. Raspbian website, `https://www.raspbian.org/`
2. Aapo, H.: Fast and robust fixed-point algorithms for independent component analysis. IEEE Transactions on Neural Networks 10(3), 626–634 (May 1999)



<Back to previous page | Listed in category: Tomato

Product name : Tomato A
Seller : Udon Honda
Shipping deadline : 2016.09.12
Price : ¥60
Quantity: 1

Product name : Tomato B
Seller : Taku Yamada
Shipping deadline : 2016.09.12
Price : ¥70
Quantity: 1

Product name : Tomato C
Seller : Masato Saito
Shipping deadline : 2016.09.11
Price : ¥80
Quantity: 1

Product name : Tomato B
Seller : Chisa Ogata
Shipping deadline : 2016.09.10
Price : ¥85
Quantity: 1

Product name : Tomato C
Seller : Shusuke Minato
Shipping deadline : 2016.09.9
Price : ¥70
Quantity: 1

Product name : Tomato D
Seller : Taro Ono
Shipping deadline : 2016.09.08
Price : ¥75
Quantity: 1

Product name : Tomato A
Seller : Kaoru Tokai
Shipping deadline : 2016.09.05
Price : ¥76
Quantity: 1

Product name : Tomato A
Seller : Kiku Nakata
Shipping deadline : 2016.09.02
Price : ¥80
Quantity: 1

Fig. 5 Snapshot of VegeShop website.

3. Akyildiz, I.F., Kasimoglu, I.H.: Wireless sensor and actor networks: Research challenges. Ad Hoc Networks Journal (Elsevier) 2(4), 351–367 (October 2004)
4. Azad, P., Asfour, T., Dillmann, R.: Combining harris interest points and the sift descriptor for fast scale-invariant object recognition. In: Proceeding of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2009). pp. 4275–4280 (October 2009)
5. Özgür B. Akan, Akyildiz, I.F.: Event-to-sink reliable transport in wireless sensor networks. IEEE/ACM Transactions on Networking 13(5), 1003–1016 (October 2005)
6. Bay, H., Tuytelaars, T., Gool, L.V.: SURF: Speeded up robust features. Lecture Notes in Computer Science 3951, 404–417 (October 2006)
7. Cun, Y.L.: Generalization and network design strategies. Tech. Rep. CRG-TR-89-4, Department of Computer Science, University of Toronto (June 1989)
8. Fujiyoshi, H.: Gradient-based feature extraction: Sift and hog. Tech. rep., IEICE (August 2007)
9. Hinton, G.E., Osindero, S., Teh, Y.W.: A fast learning algorithm for deep belief nets. Neural Computation 18(7), 1527–1554 (July 2006)
10. Hinton, G.E., Salakhutdinov, R.: Reducing the dimensionality of data with neural networks. Sciense 313(5786), 504–507 (July 2006)
11. Jain, A.K., Mao, J., Mohiuddin, K.M.: Artificial neural networks: a tutorial. Computer 29(3), 31–44 (March 1996)
12. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. arXiv preprint arXiv:1408.5093 (2014)
13. Jiang, X., Dawson-Haggerty, S., Dutta, P., Culler, D.: Design and implementation of a high-fidelity ac metering network. In: Proceeding of the 8-th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN-2009). pp. 253–264. San Francisco, US (April 2009)

14. Kang, L., Kumar, J., Ye, P., Li, Y., Doermann, D.: Convolutional neural networks for document image classification. In: Proceedings of 22nd International Conference on Pattern Recognition 2014 (ICPR-2014). pp. 3168–3172 (August 2014)
15. Karahan, S., Karaoz, A., Ozdemir, O.F., Gul, A.G., Uludag, U.: On identification from periocular region utilizing sift and surf. In: Proceedings of the 22-nd European Signal Processing Conference (EUSIPCO-2014). pp. 1392–1396 (September 2014)
16. Le, Q.V.: Building high-level features using large scale unsupervised learning. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing 2013 (ICASSP-2013). pp. 8595–8598 (May 2013)
17. Lee, H., Grosse, R., Ranganath, R., Ng, A.Y.: Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: Proceedings of the 26th Annual International Conference on Machine Learning. pp. 609–616 (June 2009)
18. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV-1999). pp. 1150–1157 (September 1999)
19. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (November 2004)
20. Mikolajczyk, K.: A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence 60(10), 1615–1630 (October 2005)
21. Murillo, A.C., Guerrero, J.J., Sagüës, C.: Surf features for efficient robot localization with omnidirectional images. In: Proceedings of the IEEE Robotics and Automation Roma. pp. 3901–3907 (April 2007)
22. Nakano, T., Kida, T.: Two dimensional pattern matching for jpeg images. Tech. rep., IEICE (December 2008)
23. Sainath, T.N., Kingsbury, B., Mohamed, A.R., Dahl, G.E., Saon, G., Soltau, H., Beran, T., Aravkin, A.Y., Ramabhadran, B.: Improvements to deep convolutional neural networks for LVCSR. In: Proceedings of IEEE Workshop on Automatic Speech Recognition and Understanding 2013 (ASRU-2013). pp. 315–320 (December 2013)
24. Sakai, Y., Oda, T., Ikeda, M., Barolli, L.: An object tracking system based on sift and surf feature extraction methods. In: Proceedings of the 5th International Workshop on Information Networking and Wireless Communications (INWC-2015). pp. 561–565 (September 2015)
25. Sakai, Y., Oda, T., Ikeda, M., Barolli, L.: A vegetable category recognition system using deep neural network. In: accepted, to appear in Proceedings of the 10th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS-2016) (July 2016)
26. Sikora, R., Sikora, J., Cardelli, E., Chady, T.: Artificial neural network application for material evaluation by electromagnetic methods. In: Proceedings of International Joint Conference on Neural Networks (IJCNN-1999). vol. 6, pp. 4027–4032 (July 1999)
27. Takaki, S., Yamagishi, J.: Deep auto-encoder based low-dimensional feature extraction using fft spectral envelopes in statistical parametric speech synthesis. IEICE Technical Report 2015-SLP-109(18), 1–6 (November 2015)
28. Tola, E., Lepetit, V., Fua, P.: A fast local descriptor for dense matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2008). pp. 1–8 (June 2008)
29. Tribelhorn, B., Dodds, Z.: Evaluating the roomba: A low-cost, ubiquitous platform for robotics research and education. In: Proceedings of the IEEE International Conference on Robotics and Automation (IEEE ICRA-2007). pp. 1393–1399. Roma, Italy (April 2007)
30. Tsugawa, S., Ohsaki, H.: Community structure and interaction locality in social networks. IPSJ Journal 56(6) (June 2015)
31. Ueda, K., Tamai, M., Yasumoto, K.: A system for daily living activities recognition based on multiple sensing data in a smart home. In: Proceedings of the Multimedia, Distributed, Cooperative, and Mobile Symposium (DICOMO-2014). pp. 1884–1891 (July 2014)
32. Ueki, M.: Human-centric computing to effort. Transactions of the Japan Society of Mechanical Engineers (2013)

33. Uhrig, R.E.: Introduction to artificial neural networks. In: Proceedings of the IEEE 21st International Conference on Industrial Electronics, Control, and Instrumentation (IECON-1995). vol. 1, pp. 33–37 (November 1995)
34. Uijlings, J.R.R., Smeulders, A.W.M., Scha, R.J.H.: Real-time visual concept classification. IEEE Transactions on Multimedia 12(7), 665–681 (October 2010)
35. Yu, Y., Rittle, L.J., Bhandari, V., LeBrun, J.B.: Supporting concurrent applications in wireless sensor networks. In: Proceedings of the 4-th ACM International Conference on Embedded Networked Sensor Systems (ACM SenSys-2006). pp. 139–152. Boulder, US (November 2006)

A Fuzzy-Based Wireless Sensor and Actuator Network: Simulation and Experimental Results

Keisuke Ebisu, Takaaki Inaba, Donald Elmazi, Makoto Ikeda, Leonard Barolli, Elis Kulla

Abstract Fuzzy sets and fuzzy logic have been developed to manage vagueness and uncertainty in a reasoning process of an intelligent system such as a knowledge based system, an expert system or a logic control system. In our previous work, we evaluated the performance of our proposed fuzzy-based Wireless Sensor and Actuator Network (WSAN) testbed that based on data provided by depth and RGB sensors and sink selects an appropriate actuator node. In this paper, we evaluate the performance of our proposed fuzzy-based WSAN simulation and testbed systems considering two actuators. From the performance evaluation, we observe that the difference between simulation and experiment is small when REA2 is small or medium state.

Key words: fuzzy, WSAN, testbed, simulation.

Keisuke Ebisu, Takaaki Inaba, Donald Elmazi
Graduate School of Engineering, Fukuoka Institute of Technology,
3-30-1 Wajiro-higashi, Higashi-ku, Fukuoka 811-0295, Japan
e-mail: keisuke19921001@gmail.com, g.takaaki.inaba@gmail.com, donald.elmazi@gmail.com

Makoto Ikeda, Leonard Barolli
Department of Information and Communication Engineering,
Fukuoka Institute of Technology,
3-30-1 Wajiro-higashi, Higashi-ku, Fukuoka 811-0295, Japan
e-mail: makoto.ikd@acm.org, barolli@fit.ac.jp

Elis Kulla
Department of Information and Computer Engineering, Okayama University of Science,
1-1 Ridai-cho, Kita-ku, Okayama 700-0005, Japan
e-mail: kulla@ice.ous.ac.jp

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_69

1 Introduction

The main functionality of Wireless Sensor and Actuator Networks (WSANs) is to make actuators perform appropriate actions, based on gathered information about their environment from sensors and actuators. When important data has to be transmitted, sensors may transmit their data back to the sink, which will control the actuators' tasks from distance, or transmit their data to actuators, which can perform actions independently from the sink node. WSANs are capable of monitoring physical phenomenons, processing sensed data, making decisions based on the sensed data and completing appropriate tasks when needed [4, 11, 5, 16, 8, 15]. WSANs can be established on the Internet of Things (IoT) environment [14, 13].

Technologies to detect a specific object in images are expected to further expand to wide range of applications, such as vehicle detection functions using traffic light cameras and raiders for Intelligent Transport System (ITS). In just four years, 75% of newly marketed vehicles will be connected via Wi-Fi.

In [6, 9], we proposed a fuzzy-based testbed system that based on data provided by sensor and actuator select an appropriate actuator node. In [7], we implemented a real-time control tool using Processing language to check the distance between actuator and sink.

In this paper, we evaluate the performance of our proposed Fuzzy-Based WSAN (FBWSAN) testbed system considering multiple actuators. Also, we implemented an improved real-time control tool using Processing language to consider two actuators.

The structure of the paper is as follows. In Section 2, we give an overview of WSANs. The testbed design is shown in Section 3. In Section 4, we show the evaluation results. Finally, conclusions and future work are given in Section 5.

2 Wireless Sensor and Actuator Network

WSANs are capable of monitoring physical phenomenons, processing sensed data, making decisions based on the sensed data and completing appropriate tasks when needed. The devices deployed in the environment are sensors able to sense environmental data, actuators able to react by affecting the environment or have both functions integrated.

Here, the former scheme is called semi-automated architecture and the latter on fully-automated architecture. Obviously, both architectures can be used in different applications. In the fully-automated architecture are needed new sophisticated algorithms in order to provide appropriate coordination between nodes of WSAN. On the other hand, it has advantages, such as network lifetime, energy consumption, latency, local position accuracy, reliability and so on [4].

After data has been sensed from sensors, they are collected to the sink for semi-automated architecture or sent to the actuators for fully-automated architecture. Then a task is assigned to actuators. In general, one or more actuators take responsibility and perform appropriate actions. Different actuators may be chosen for acting, depending on their characteristics and conditions.

3 FBWSAN Testbed design

3.1 Overview

Our FBWSAN testbed is composed of a laptop PC equipped with Xtion Pro Live and two Roombas (model 630) acting as sink and actuators, respectively. Our system also includes many sensors. An overview of the system is shown in Fig. 1. The experiments have been performed in indoor environment, within our departmental floor of size roughly 10 meters.

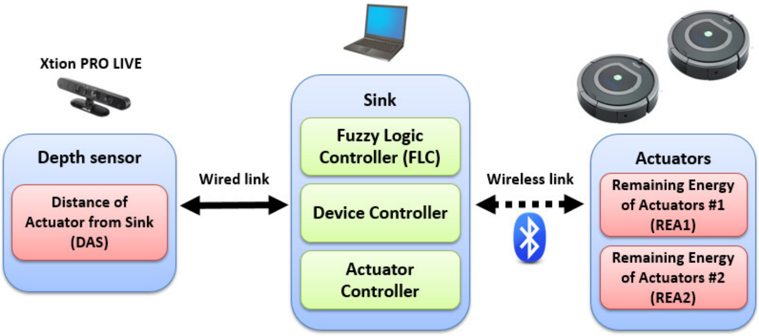

Xtion PRO LIVE
Depth sensor
Distance of Actuator from Sink (DAS)
Wired link
Sink
Fuzzy Logic Controller (FLC)
Device Controller
Actuator Controller
Wireless link
Actuators
Remaining Energy of Actuators #1 (REA1)
Remaining Energy of Actuators #2 (REA2)

Fig. 1 Overview of FBWSAN testbed.

Xtion Pro Live is motion capture device, which are mounted depth sensor, RGB sensor and stereo microphone [1]. The iRobot Corporation [2] provides Roomba Open Interface (ROI) specifications. In case of Roomba, the ROI connector is the gateway to reversible Roomba hacking. All devices can be plugged into the ROI. Our testbed uses Bluetooth module to connect the ROI. The Bluetooth module is from BlueSMiRF (WRL-12582) [3].

Based on WSAN characteristics and challenges, we consider the following parameters for implementation of our fuzzy-based testbed: Remaining Energy of Actuator #1 (REA1), Remaining Energy of Actuator #2 (REA2), and Distance of Actuator from Sink (DAS). We collect REA1 and REA2 from each Roomba via Bluetooth. DAS is measured by Xtion Pro Live equipped with depth sensor.

3.2 Scenario Settings

In this work, we build a system which measure distance from sink to actuator #1 by using Xtion Pro Live with OpenNI. OpenNI is an open source software framework that is able to read sensor data from Xtion Pro Live, among other natural user interface sensors.

In our testbed system, we implemented an improved real-time control tool using Processing language to consider two actuators. We used a *SimpleOpenNI* library in Processing to control the sensors. In our case, in order to recognize two actuators, we used color information on the object.

Snapshots of experimental environment is shown in Fig. 2. Our system has a monitored function to display the calculated distance from depth sensor.

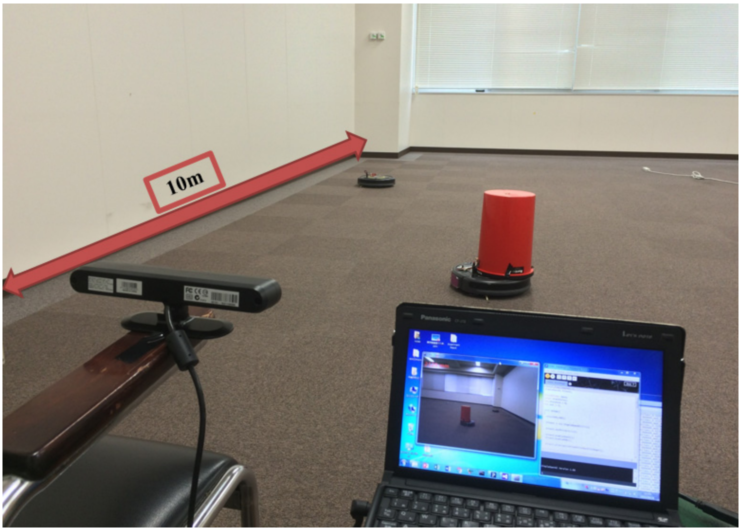


Fig. 2 Experimental environment.

3.3 Fuzzy System

In this paper, we use fuzzy logic [17] system called FuzzyC [10] to implement the proposed system in sink.

The structure of the proposed system is shown in Fig. 3. It consists of one Fuzzy Logic Controller (FLC) and its basic elements are shown in Fig. 4. They are the fuzzifier, inference engine, fuzzy rule base and defuzzifier. We use triangular and trapezoidal membership functions for FLC, because they are suitable for real-time operation [12]. The membership functions are shown in Fig. 5.

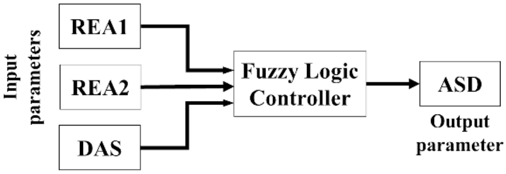


Fig. 3 Proposed system.

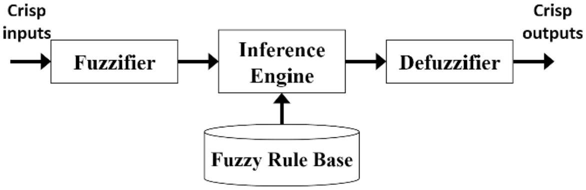


Fig. 4 FLC structure.

We use REA1, REA2 and DAS input parameters for FLC. The term sets for each input linguistic parameter are defined respectively as shown in Table 1. The output linguistic parameter is the Actuator Selection Decision (ASD).

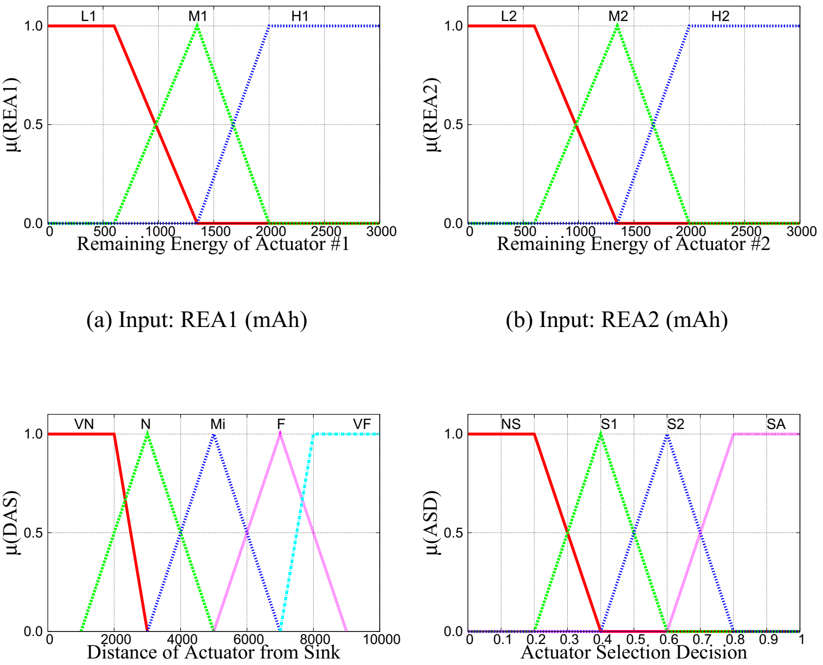
Remaining Energy of Actuator #1 (REA1): As actuators are active in the monitored field, they perform tasks and exchange data in different ways from each other. Consequently, also based on their characteristics, some actuators may have a lot of power remaining and other may have very little, when an event occurs. We consider three levels of REA1 for actuator selection. We collected the REA1 from Roomba #1 via Bluetooth.

Remaining Energy of Actuator #2 (REA2): We consider three levels of REA2 for actuator selection. We collected the REA2 from Roomba #2 via Bluetooth.

Distance of Actuator from Sink (DAS): The number of actuators in a WSAN is smaller than the number of sensors in our testbed. Thus, when an actuator is called for action near an event, the distance from the actuator to the event is different for different actuators and events. Depending on five distance levels, our system takes decisions on the availability of the actuator nodes. Our testbed uses depth sensor to measure the distance from actuator #1 to sink. We collected the DAS via Xtion Pro Live equipped with depth sensor.

Actuator Selection Decision (ASD): Our system is able to decide the willingness of an actuator to be assigned a certain task at a certain time. The actuators respond in four different levels, which can be interpreted as:

- NS: Both #1 and #2 actuators have to stay at the present position.
- S1: Actuator #1 has required information and potential to take full responsibility. On the other hand, actuator #2 must be stay at the present position.
- S2: Actuator #2 has required information and potential to take full responsibility. On the other hand, actuator #1 must be stay at the present position.
- SA: Both #1 and #2 actuators have almost all required information and potential to take full responsibility of completing the task.



(a) Input: REA1 (mAh)

(b) Input: REA2 (mAh)

(c) Input: DAS (mm)

(d) Output: ASD

Fig. 5 Membership functions.

Table 1 Parameters and their term sets for FLC.

Parameters	Term Sets
Remaining Energy of Actuator #1 (REA1)	Low1 (L1), Medium1 (M1), High1 (H1)
Remaining Energy of Actuator #2 (REA2)	Low2 (L2), Medium2 (M2), High2 (H2)
Distance of Actuator from Sink (DAS)	Very Near (VN), Near (N), Middle (Mi), Far (F), Very Far (VF)
Actuator Selection Decision (ASD)	NS (Not Select), S1 (Select actuator #1), S2 (Select actuator #2), SA (Select All actuators)

4 Evaluation Results

Here, we present the simulation and experimental results of FBWSAN system for different REA2 types. In this paper, we discuss the results when the REA1 is H1 state. In Fig. 6, when REA2 is high (H2) state, the results of ASD is higher than

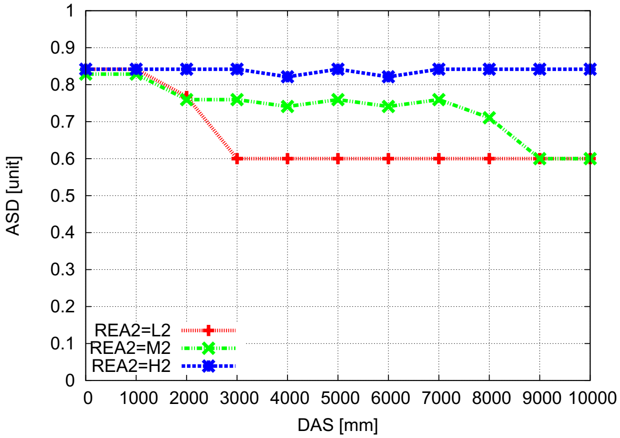


Fig. 6 Simulation results for REA1 is H1.

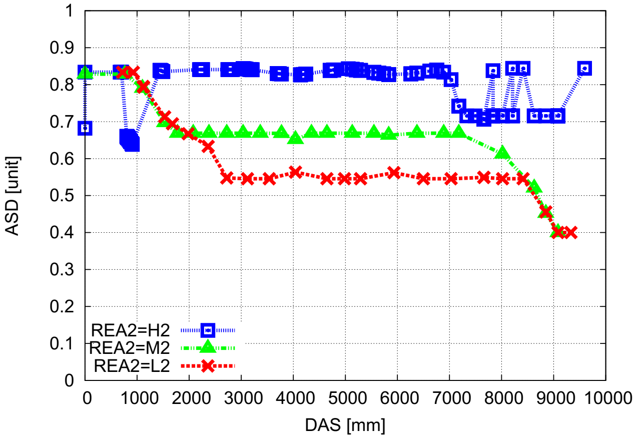


Fig. 7 Experimental results for REA1 is H1.

other states. In this case, sink selects both actuators to take full responsibility of completing the task. When REA2 is medium and low states, the simulation values of ASD are decreased with increase of the DAS. The minimum value of ASD is 0.6 unit.

For experimental results, when the REA2 is high, we can observe that there are oscillations (see Fig. 7). The fuzzy logic controller selects three levels based on sensed real data. Therefore, the REA2 of high state could have difference (from 1350 mAh to 3000 mAh).

When the REA2 is medium (M2) or low (L2) state (see Fig. 6 and Fig. 7), the difference of the ASD between simulation and experiment is small. While for DAS more than 8500mm, the difference is big.

5 Conclusions

In this paper, we evaluated the performance of our FBWSAN testbed considering two actuators. We considered three input parameters for implemented system in order to select actuators in real indoor conditions. We implemented an improved real-time control tool using Processing language to consider multiple actuators. The performance evaluation shows that the simulation and experimental results are almost the same when REA2 is small or medium state.

In the future work, we would like to make extensive experiments to evaluate different evaluation parameters and different environments.

References

1. ASUSTeK Computer Inc., `https://www.asus.com/3D-Sensor/Xtion_PRO_LIVE/`
2. iRobot Corporation, `http://www.irobot.com/`
3. SparkFun Bluetooth Modem - BlueSMiRF Gold, `https://www.sparkfun.com/products/12582`
4. Akyildiz, I.F., Kasimoglu, I.H.: Wireless sensor and actor networks: Research challenges. Ad Hoc Networks Journal (Elsevier) 2(4), 351–367 (October 2004)
5. Özgür B. Akan, Akyildiz, I.F.: Event-to-sink reliable transport in wireless sensor networks. IEEE/ACM Transactions on Networking 13(5), 1003–1016 (October 2005)
6. Ebisu, K., Inaba, T., Elmazi, D., Ikeda, M., Barolli, L., Kulla, E.: A fuzzy-based testbed design for wireless sensor and actuator networks. In: Proceedings of the 5th International Workshop on Information Networking and Wireless Communications (INWC-2015). pp. 548–553 (September 2015)
7. Ebisu, K., Inaba, T., Elmazi, D., Ikeda, M., Kulla, E., Barolli, L.: Performance evaluation of a fuzzy-basedwireless sensor and actuator network testbed considering depth and rgb sensors. In: Proceedings of the 10th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2016). pp. 69–75 (July 2016)
8. Forlizzi, J., DiSalvo, C.: Service robots in the domestic environment: A study of the roomba vacuum in the home. In: Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction (ACM HRI-2006). pp. 258–265. Utah, US (March 2006)
9. Ikeda, M., Ebisu, K., Sakai, Y., Elmazi, D., Barolli, L.: Performance evaluation of a fuzzy-based wireless sensor and actuator network testbed for object tracking. In: Proceedings of the 6th International Workshop on Methods, Analysis and Protocols for Wireless Communication (MAPWC-2015). pp. 442–447 (November 2015)
10. Inaba, T., Sakamoto, S., Oda, T., Barolli, L., Takizawa, M.: A new FACS for cellular wireless networks considering QoS: A comparison study of FuzzyC with MATLAB. In: Proceedings of the 18th International Conference on Network-Based Information Systems (NBiS-2015). pp. 338–344 (September 2015)
11. Jiang, X., Dawson-Haggerty, S., Dutta, P., Culler, D.: Design and implementation of a high-fidelity ac metering network. In: Proceedings of the International Conference on Information Processing in Sensor Networks 2009 (IPSN-2009). pp. 253–264. San Francisco, US (April 2009)
12. Mendel, J.M.: Fuzzy logic systems for engineering: A tutorial. Proceedings of the IEEE 83(3), 345–377 (1995)
13. Schmitt, S., Will, H., Aschenbrenner, B., Hillebrandt, T., Kyas, M.: A reference system for indoor localization testbeds. In: Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN-2012). pp. 1–8. Sydney, Australia (November 2012)

14. Sung, J.Y., Guo, L., Grinter, R.E., Christensen, H.I.: My Roomba is Rambo: Intimate home appliances. In: Proceedings of the 9th International Conference on Ubiquitous Computing (UbiComp-2007). pp. 145–162. Seoul, South Korea (September 2007)
15. Tribelhorn, B., Dodds, Z.: Evaluating the roomba: A low-cost, ubiquitous platform for robotics research and education. In: Proceedings of the IEEE International Conference on Robotics and Automation (IEEE ICRA-2007). pp. 1393–1399. Roma, Italy (April 2007)
16. Yu, Y., Rittle, L.J., Bhandari, V., LeBrun, J.B.: Supporting concurrent applications in wireless sensor networks. In: Proceedings of the 4th ACM International Conference on Embedded Networked Sensor Systems (ACM SenSys-2006). pp. 139–152. Boulder, US (November 2006)
17. Zadeh, L.: Fuzzy logic, neural networks, and soft computing. ACM Communications pp. 77–84 (1994)

Numerical analysis of resonance characteristics in cavities in periodic structure for WDM telecommunication system

Hiroshi Maeda, Kazuya Tomiura and Jianming Jin

Abstract Numerical analysis of frequency filtering characteristics by cavities of three different lengths situated in X-shaped photonic crystal waveguide with triangular lattice is demonstrated by constrained interpolation profile (CIP) method to solve Maxwell's equations. From fast Fourier transform (FFT) analysis of output signal by changing width of Gaussian window function, the resonant peak frequencies of filtered spectrum by microwave experiment and those by the simulation coincided each other in difference of order of 1%. Especially, relative error of the resonant frequency to the experiment was improved to the half of our previous numerical work, choosing the window function suitably.

1 Introduction

Photonic crystal structures or electromagnetic band gap structures have periodic distribution of material constants in it and are applied into practical use in optical components for signal generation, transmission and reception, because of its unique and sensitive characteristics for frequency. Those characteristics are based on photonic band gap (PBG) phenomena[1]-[2], which originates from transmission and reflection properties of periodicity. In signal transmission and processing utilizing such devices in optical integrated circuits, high density multiplexing in frequency domain is expected due to its sensitivity with respect to optical wavelength. This is important to improve capacity of information transmission in photonic network with dense multiplexing technique of signal in wavelength domain.

Hiroshi Maeda
Fukuoka Institute of Technology, 3-30-1 Wajiro-Higashi, Fukuoka 811-0295, Japan e-mail: hiroshi@fit.ac.jp

Kazuya Tomiura and Jianming Jin
Fukuoka Institute of Technology, 3-30-1 Wajiro-Higashi, Fukuoka 811-0295, Japan

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_70

The behavior of electromagnetic wave in periodic structure can be controlled by selecting material constants, designing periodic profile of the structure and the frequency spectrum range of the signal. For various kinds of materials and for various frequency ranges of purposes, PBG might be found by designing the structure with fundamental unit lattice. This means that, by setting the parameters appropriately, confinement and transmission of electromagnetic wave along line-defect in the structure is possible for desired range of frequency from radio wave to optical signal domain. In this meaning, we examined the propagation and filtering characteristics of two dimensional photonic crystal waveguide and cavities with triangular lattice of dielectric pillar in microwave frequency around 4 GHz, to be compared with experimental results[3]-[5]. In the experiment, authors have used ceramic rods as dielectric pillar. For its quite low-loss property and high dielectric constant of $\varepsilon_r = 36.0$, ceramic is suitable to confine electromagnetic field tightly when periodic structure is composed with less numbers of layers.

As a useful numerical analysis technique, finite different time domain (FDTD) method[6] is powerful and widely used, for enabling to design various boundary shape of structure with multi-dimensional problems. However, it is known that FDTD shows physically incorrect behavior for problems including large gap of material constants at the boundary. It is possible to avoid such behavior by setting smaller cells, however, it increases cell numbers for the entire analysis region with increase of memory and time for computation. This means that we should pay attention to guarantee reliable results to choose the discrete cells within reasonable computation time.

On the contrary, constrained interpolated profile (CIP) method has been proposed by Yabe et al.[7], with advantage of preventing such spurious behavior in FDTD method. Because the CIP method hires cubic polynomials to express the profile in a cell, it is possible to renew not only the profile at each discretized point but also the first order spatial derivatives of the profile. Authors have been applied the CIP method for analysis of wave propagation in periodic structures composed by ceramic pillars in air background.

In this paper, filtering characteristics of cavities, situated in the output waveguides after branching point, are numerically investigated by CIP method[8]-[9]. In the simulation, band-limited wave with time evolving envelope of sampling function is given as input. The resonant frequency peaks were obtained by fast Fourier transform (FFT) of output electric field at each output port. The results showed that obvious resonant peaks are observed in Fourier transformed domain, which coincides with experimental results in difference of order of 1%. Especially, relative error of resonant frequency to the experiment was improved to the half of our previous numerical work[9], choosing the window function suitably.

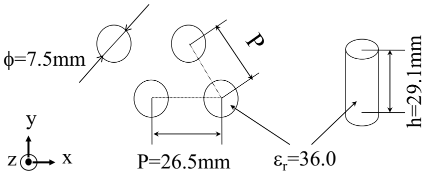
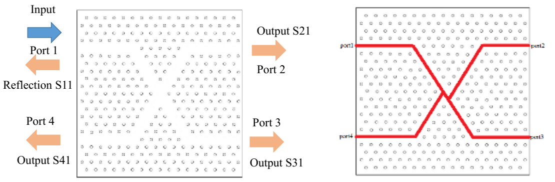


Fig. 1 Top view of fundamental triangular lattice of photonic crystal and side view of ceramic rod with parameters.

2 Numerical analysis by CIP method

From Maxwell's curl equations, we can derive formulations based on CIP method. The method hires cubic polynomials to express the field profile between discreet points. The coefficients are determined by values of profile and the spatial derivative on each discreet point, which improves numerical accuracy compared with ordinary finite difference technique based on linear approximation of the profile. The details of CIP method[7] and its application to electromagnetic wave propagation[8] are described in literatures.



(a) Symmetric X–shaped waveguide (b) Asymmetric X-shaped waveguide

Fig. 2 Waveguide by a line defect in two dimensional, pillar-type photonic crystal with triangular lattice. (a) Symmetric X-shaped waveguide. (b) Asymmetric X-shaped waveguide.

3 Two-Dimensional, Pillar-type Photonic Crystal structure, Line-defect Waveguide and Cavities

In Fig.1, top view of unit triangular lattice is shown, together with coordinate system and illustration of a circular ceramic rod and the parameters. The longitudinal axis of the cylinder corresponds to polarization direction of electric field E_z of TE mode. Material of the cylinder is ceramic with relative dielectric constant $\varepsilon_r = 36.0$ at frequency $f = 4.0GHz$. In the measurement frequency range from 3.6 to 4.2 GHz, the dielectric loss is as negligibly small as 10^{-6} and the real part of dielectric constant can be assumed to be constant. The ceramic rods are fabricated and supplied by Ky-

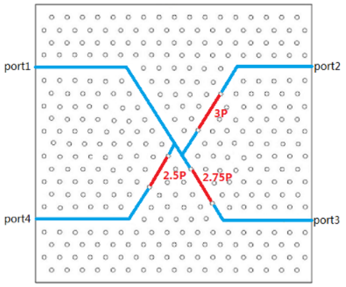


Fig. 3 Filtering circuits with pairs of rods as cavities. The waveguides and cavities are indicated by blue and red lines, respectively. The distance between each pairs are different for three kinds of filtered outputs.

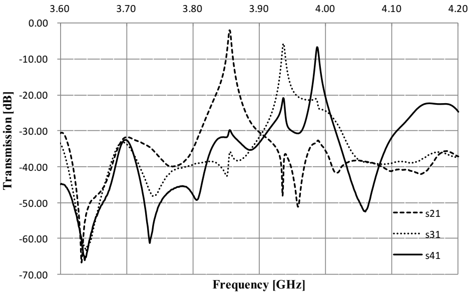
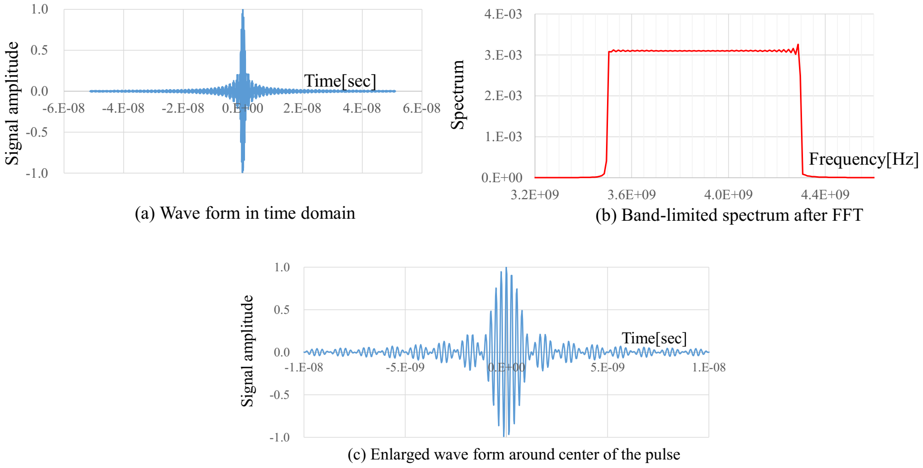


Fig. 4 Measurement of outputs to port #2, #3, and #4 by dashed, dotted, and solid line, respectively.

ocera company in Japan for general use as microwave circuit elements. The lattice period $P = 26.5mm$ was designed so that the line defect structure shows PBG for frequency range from 3.6 to 4.2 GHz in the experiment. Following the design, the incident wave is guided along with defect without penetrating into periodic structures.

In Fig.2(a), TE mode with components (H_x, H_y, E_z) is excited in port #1 at the top left. The electric field E_z has Gaussian profile along y-axis with full beam waist $w_0 = 24.8mm$. For equal dividing of power to each output port, asymmetric X-shaped branch waveguide in Fig.2(b) is employed for filtering[5].

As shown in Fig.3, three different lengths of cavity structure with a pair of dielectric rods are situated in each of output waveguide to achieve filtering circuits. Here in the figure, $P = 26.5mm$ denotes lattice period as well. As an example of measurement, output spectrum of asymmetric X-shaped waveguide in Fig.3 are shown in Fig.4. In the figure, three sharp and obvious peaks are observed in each output curve. These peaks express the center frequency of cavity filters situated in each output waveguide.



(a) Wave form in time domain

(b) Band-limited spectrum after FFT

(c) Enlarged wave form around center of the pulse

Fig. 5 Band-limited input signal for $f_L = 3.5$ to $f_U = 4.3$ GHz in (a) time domain and (b) frequency domain.

4 FFT analysis of outputs electric field in Cavity

In this section, filtered electric field in Fig.4 by measurements are compared with numerical results. All the parameters are same with our previous work[9] except variation of window functions. In numerical analysis, the discretization for space and time are set to be $\Delta x = \Delta y = 0.75mm$ and $\Delta t = 2.5 \times 10^{-13} sec$, respectively.

Supposing band-limited spectrum with stepwise square profile, real part of the time evolving input wave $f(t)$ is given by inverse Fourier transform as follows;

$$Re\{f(t)\} = -f_L \times Sa(2\pi f_L t) + f_U \times Sa(2\pi f_U t), \tag{1}$$

where

$$Sa(x) = \frac{\sin(x)}{x} \tag{2}$$

is a sampling function, f_L and f_U are lower and upper frequency [Hz] of the limited band, respectively. The input wave form is depicted as function of time in Fig.5(a). Here, $f_L = 3.5GHz$ and $f_U = 4.3GHz$ are used for obtaining the flat spectrum in frequency range from 3.6 to 4.2 GHz in the experiment. The maximum input amplitude in the simulation comes at time $t = 100/f_C$[sec], where $1/f_C$ is time period for center frequency of the range and $f_C = (f_L + f_U)/2$. Input signal in Fig.5(a) is Fourier-transformed by FFT and shown in Fig.5(b). For clarity, the input pulse in time domain is enlarged and depicted in Fig.5(c). Although Gibbs phenomena is seen near f_L and f_H, flat spectrum is obtained in frequency range from 3.6 to 4.2 GHz.

For obtaining frequency resolution to be comparable with experimental results, the time evolving data of CIP method is sampled every $\Delta t_{sample} = 100\Delta t$.

Therefore, sampled time interval $\Delta t_{sample} = 25.0psec$ with numbers of sample data $N_{sample} = 4096$ was set to obtain frequency resolution $\Delta f = (\Delta t_{sample} \times N_{sample})^{-1} \simeq 9.77MHz$. This resolution brings $600/9.77 \simeq 61$ points in the measured frequency range of 600MHz from 3.6 to 4.2 GHz.

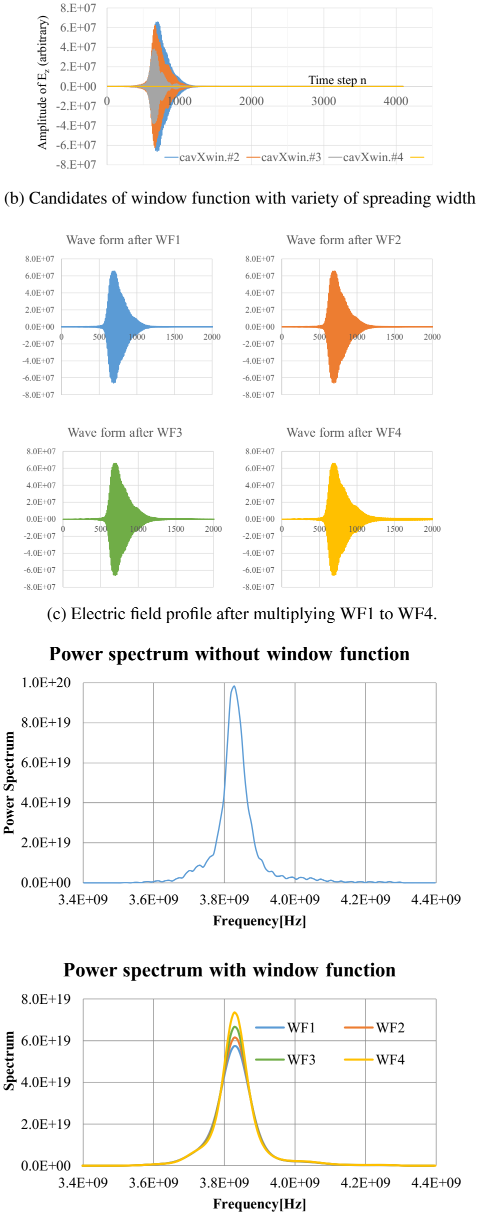
In Fig. 6 (a), electric field profile in time domain observed at center of each cavity is depicted. It is found that after $t = 1200\Delta t$, the field is diverging, because the derivative value could not be evaluated for flat and tiny electric field amplitude after the pulse wave travel through the cavity. Therefore, the diverging filed must be eliminated using suitable window function. In Fig.6(b), variety of Gaussian window function from WF1 to WF4 are illustrated. These window functions have peak at time step $n = n_0$ which shows maximum electric field amplitude. From WF1 to WF4, the width is controlled by selecting coefficient α as 1.0e-5, 8.0e-6, 6.0e-6, and 4.0e-6 for Gaussian profile $W(n) = \exp\{-\alpha(n-n_0)^2\}$, respectively. In Fig.6(c), electric field profile in cavity #2 for various window functions are depicted. In Fig.6(d), power spectrum with or without window functions are compared. Similarly for cavity #3 and #4, we obtained resonant frequency peaks.

The peak frequencies are compared with experimental results in Table 1. In the table, resonant frequency by experiment was used as reference to evaluate the difference in numerical results. In the results, the resonant frequency in each cavity shows good agreement. For resonant frequency for cavity #4, the difference is improved from our previous result[9] for hiring WF4. Also for cavity #4 with WF4 shows the highest resonant peak in Fig.6(d). In general, higher resonant peak leads to higher Q-factor. Comparison of Q-factors between experiment and numerical results are presented in the conference.

Cavity	#2	#3	#4
Experiment [GHz]	3.855	3.936	3.987
CIP after FFT with Gaussian WF1 [GHz]	3.82813	3.89648	3.95551
Difference [%]	0.697147	1.0003963	0.789867
CIP after FFT with Gaussian WF2 [GHz]	3.82813	3.89648	3.95551
Difference [%]	0.697147	1.0003963	0.789867
CIP after FFT with Gaussian WF3 [GHz]	3.82813	3.89648	3.95551
Difference [%]	0.697147	1.0003963	0.789867
CIP after FFT with Gaussian WF4 [GHz]	3.82813	3.89648	3.96484
Difference [%]	0.697147	1.003963	0.555706

Table 1 Comparison of resonant frequency

(a) Time evolving electric field in center of each cavity without window functions.



(b) Candidates of window function with variety of spreading width.

(c) Electric field profile after multiplying WF1 to WF4.

5 Conclusions

Filtering characteristics of cavities in two dimensional X-shaped pillar-type photonic crystal waveguide with triangular lattice were simulated by CIP method. As the input with limited spectrum are given and the filtered outputs were analyzed by Fourier transforms. The simulation results of resonance for 3 output ports with different cavities were compared with experimental results to show good agreement with difference of order of 1% of resonant center frequencies. This suggests that selecting window function suitably improves evaluation of resonant frequency and Q-factors by numerical analysis.

Acknowledgment

This work was financially supported by KAKENHI No. 15K06043, Grant-in-Aid for Scientific Research (C) by Japan Society for the Promotion of Science(JSPS) in 2016.

References

1. K. Yasumoto, Ed., *Electromagnetic Theory and Applications for Photonic Crystals*, CRC PRESS, 2006.
2. J. D. Joannopoulos, R. D. Meade and J. N. Winn, *Photonic Crystals*, New Jersey: Princeton University Press, 1995.
3. H. Maeda, S. Inoue, S. Nakahara, O. Hatanaka, Y. Zhang and H. Terashima, “Experimental Study on X-shaped Photonic Crystal Waveguide in 2D Triangular Lattice for Wavelength Division Multiplexing System”, *Proceedings of 26th International Conference on Advanced Information Networking and Applications (AINA-2012)*, pp.629-632, Mar. 2012.
   “An Experimental Study on X-shaped Branching Waveguide in Two-Dimensional Photonic Crystal Structure”,
4. H. Maeda, “Four-branching waveguide in 2D photonic crystal structure for WDM system”, J. of Space-Based and Situated Computing, Vol.3, No.4, pp.227-233, Dec. 2013.
5. Y. Bao, H. Maeda and N. Nakashima, “Studies on Filtering Characteristics of X-shaped Photonic Crystal Waveguide in Two-Dimensional Triangular Lattice by Microwave Model”, *Proceedings of International Symposium on Antenna and Propagation (ISAP2015)*, pp.842-845, November 2015.
6. A. Taflove, “Advances in Computational Electrodynamics – The Finite-Difference Time-Domain Method”, Artech House.
7. T. Yabe, X. Feng and T. Utsumi, “The constrained interpolation profile method for multiphase analysis, ”*Journal of Computational Physics*, Vol.169, pp.556-593, 2001.
8. H. Maeda, “Numerical Technique for Electromagnetic Field Computation Including High Contrast Composite Material”, as Chapter 3 of *Optical Communications*, pp.41-54, InTech Open Access Publisher, Oct. 2012.
9. H. Maeda, M. Cada, Y. Bao, J. Jin and K. Tomiura, “Numerical analysis of transmission spectrum of X-shaped photonic crystal waveguide for WDM system”, *Proceedings of International Conference on The Tenth International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2016)*, accepted and to be published, July 2016.

Spectrum analysis of envelope pulse after propagating in nonlinear dielectric material

Hiroshi Maeda, Jianming Jin and Kazuya Tomiura

Abstract Optical pulse wave propagation in nonlinear and dispersive dielectric medium is simulated by constrained interpolated profile (CIP) method. In one dimensional space with dispersion and nonlinearity, Gaussian envelope pulse modulated by optical carrier wave propagates successfully with physically reliable behavior for variety of electric field amplitudes, showing the spreading and the concentrating envelope wave form. By fast Fourier transform (FFT) of time domain wave form, dependencies of input pulse amplitude to the frequency spectrum after propagation is obtained and discussed. It shows that the spectra around carrier frequency is enhanced for larger input amplitude.

1 Introduction

Soliton in optical fiber communication is one of key technologies for long distance data transmission[1][2]. In the phenomena, trade-off between linear and nonlinear dispersion terms play important role. Generally, to express those linear and nonlinear dispersion terms, we need to calculate the time convolution between electric field and susceptibility of the material. For finite difference schemes, summing up the terms from initial to current state is necessary, which requires large memory space and computational procedure in the simulation.

Finite-difference time-domain (FDTD) method[3] is a powerful tool for full-wave analysis of electromagnetic field analysis from transient to steady state problems. Some authors contributed to include the dispersion into the FDTD scheme. Luebbers et al.[4][5] implied the time convolution by summation of past electric

Hiroshi Maeda
Fukuoka Institute of Technology, 3-30-1 Wajiro-Higashi, Fukuoka 811-0295, Japan e-mail: hiroshi@fit.ac.jp

Jianming Jin and Kazuya Tomiura
Fukuoka Institute of Technology, 3-30-1 Wajiro-Higashi, Fukuoka 811-0295, Japan

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_71

field, which is updated recursively since the susceptibility function is exponential. The one-dimensional equations were solved and soliton propagation was simulated in those references. Goorjian et al.[6] derived a coupled system of nonlinear ordinary differential equations which the linear and nonlinear convolutions satisfy and solved them by finite difference technique. In the reference[7], soliton propagation in two-dimensional waveguide ($110\ \mu m \times 5\ \mu m$) was simulated, however, these second order coupled differential equations must be solved at each time step. Sullivan[8]-[11] formulated linear and nonlinear dispersion terms by utilizing Z-transform and simulated one-dimensional soliton propagation for FDTD method. It can be mathematically written into closed form for most of dispersive terms, as long as the terms are expressed by analytical function in frequency domain.

The constrained interpolation profile (CIP) method has been proposed by Yabe et al.[12]. The method is based on finite difference scheme hiring cubic polynomials to express profile of the phenomena. Coefficients of the polynomials at each discrete grid are determined by values of the function and the derivative function. Time evolving field is expressed by using the polynomials with renewed coefficients. It has advantage of higher accuracy compared to the FDTD method. Authors have been applied the CIP method to analysis of wave propagation in various structures with linear and non-dispersive medium[13]-[18]. In Ref.[19], authors reported that the conventional CIP analysis is successfully expanded to the frequency dependent medium and nonlinear medium.

In this paper, we further demonstrate propagation of electromagnetic wave in one-dimensional space with dispersive and nonlinear characteristics by the CIP method[19]. The dispersion and nonlinearity are introduced by Sullivan's formulations with Z-transform technique. For incidence of Gaussian envelope pulse which is modulated by optical carrier wave, we observed variety of envelope pulse wave form from spreading to soliton-like propagation for smaller and larger input amplitude. We further investigated the frequency spectrum of wave propagated after nonlinear medium. It shows that the spectra around carrier frequency is enhanced for larger input amplitude, compared with smaller amplitude.

2 Formulation of the problem

2.1 Governing equations for electromagnetic wave propagation in nonlinear dielectric medium

Let us consider electromagnetic wave propagation in lossless, isotropic and non-conductive medium. Dielectric characteristics of the medium is assumed to be dispersive and nonlinear with respect to the electric field. Magnetic property of the medium is supposed to be constant and non-magnetic, thus the permeability is equal to μ_0 in vacuum. Then, Maxwell's curl equations for electric field vector $\mathbf{E}$, the electric displacement vector $\mathbf{D}$ and magnetic field vector $\mathbf{H}$ are given as follows:

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t}, \tag{1}$$

$$\nabla \times \mathbf{E} = -\mu_0 \frac{\partial \mathbf{H}}{\partial t}. \tag{2}$$

The constitutive relation between **D** and **E** for the dispersive and nonlinear medium is written as;

$$\mathbf{D} = \varepsilon_\infty \varepsilon_0 \mathbf{E} + \mathbf{P}_L + \mathbf{P}_{NL}, \tag{3}$$

where $\mathbf{P}_L$ and $\mathbf{P}_{NL}$ are linear and nonlinear dispersive polarizations as functions of electric field, ε_∞ is a constant relative permittivity of the medium in high frequency limit after saturation and ε_0 is permittivity in vacuum, respectively. We treat one-dimensional space, then the polarization terms and the electric field component are denoted with scalar function. P_{NL} is decomposed as $P_{NL} = P_R + P_K$, where P_R and P_K are Raman scattering and Kerr effect, respectively.

2.2 Linear Dispersive Polarization P_L

The linear dispersion is given by the following convolution between the electric field and the linear susceptibility $\chi^{(1)}(t)$ in time domain;

$$P_L(t) = \varepsilon_0 \int_0^t \chi^{(1)}(t-\tau) \cdot E(\tau) d\tau, \tag{4}$$

where $\chi^{(1)}(t) = \gamma_L \exp(\alpha_L t) \sin(\beta_L t)$, $\alpha_L = \omega_L \delta_L$, $\beta_L = \omega_L \sqrt{1-\delta_L^2}$, and $\gamma_L = \omega_L(\varepsilon_s - \varepsilon_\infty)/\sqrt{1-\delta_L^2}$. Value of the parameters are listed in the section of numerical results.

2.3 Raman Scattering P_R

The nonlinear polarization due to Raman scattering is given by;

$$P_R(t) = E(t)\chi_0^{(3)}(1-\alpha)\varepsilon_0 \int_0^t g_R(t-\tau) E^2(\tau) d\tau, \tag{5}$$

where $\chi_0^{(3)}$ is the third order nonlinear susceptibility and α is a weight factor between two kinds of nonlinear effect shown as Raman scattering P_R and Kerr effect P_K. Note that the function $g_R(t)$ has mathematically similar form with $\chi^{(1)}(t)$.

2.4 Kerr Effect P_K

Kerr effect is also third-order nonlinearity, but a spontaneous response of cubic electric field. It is given as:

$$P_K(t) = \varepsilon_0 \chi_0^{(3)} \alpha E^3(t). \tag{6}$$

2.5 Evaluation of time derivative for P_L and P_{NL}

Substituting Eq.(3) into Eq.(1), we have the following relation:

$$\nabla \times \mathbf{H} = \varepsilon_\infty \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} + \frac{\partial}{\partial t}(\mathbf{P}_L + \mathbf{P}_{NL}). \tag{7}$$

It can be well understood that all the terms in right hand side of Eq.(7) is equivalent to current density, especially the first term is called Maxwell's *'displacement current density'*. In CIP method, contributions from conducting current can be calculated separately as a procedure of split step computation, as well as the other finite difference technique solving time and space derivative terms. From the analogy, we first solve non-dispersive linear part of Maxwell's equations and later correct the solution with contribution from the equivalent current related to linear dispersive and nonlinear parts in the following section.

3 Formulation for CIP method for dispersive and nonlinear terms

Assuming one dimensional space and the wave propagates along with z axis. Electromagnetic field is independent of x and y in the one-dimensional problem, then, partial differential operators $\partial/\partial x$ and $\partial/\partial y$ can be treated formally as zero. In Maxwell's equations (2) and (7), we employ electric field component E_x, displacement component D_x, and magnetic field component H_y as followings:

$$\varepsilon_\infty \varepsilon_0 \frac{\partial E_x}{\partial t} + \frac{\partial H_y}{\partial z} + \frac{\partial}{\partial t}(P_L + P_{NL}) = 0, \tag{8}$$

$$\mu_0 \frac{\partial H_y}{\partial t} + \frac{\partial E_x}{\partial z} = 0. \tag{9}$$

In Eq.(8), third term in left hand side is treated separately in split step procedure as mentioned in previous section, we first solve a set of non-dispersive equations which are described as;

$$\varepsilon_\infty \varepsilon_0 \frac{\partial E_x}{\partial t} + \frac{\partial H_y}{\partial z} = 0, \tag{10}$$

$$\mu_0 \frac{\partial H_y}{\partial t} + \frac{\partial E_x}{\partial z} = 0. \tag{11}$$

Derivation of advection equation for linear and non-dispersive, or conventional, part for CIP method, the procedure is described in detail in Ref.[13]. The procedure to solve advection equation is explained in Ref.[12].

After solving conventional part of advection equation, we need to solve the dispersive and nonlinear part. By applying split step procedure to Eq.(8), electric field E_x and partial differential terms of P_L and P_{NL} are related as;

$$\varepsilon_\infty \varepsilon_0 \frac{\partial E_x}{\partial t} + \frac{\partial}{\partial t}(P_L + P_{NL}) = 0. \tag{12}$$

Eq.(12) can be discretized with respect to finite difference terms at each time step n, $n+1$, $*$, and $(*-1)$ as followings;

$$\varepsilon_\infty \varepsilon_0 \frac{E_x^{n+1} - E_x^*}{\Delta t} + \frac{(P_L^* + P_{NL}^*) - (P_L^{(*-1)} + P_{NL}^{(*-1)})}{\Delta t} = 0., \tag{13}$$

where $*$ and $(*-1)$ are temporal values between time steps $n-1 \rightarrow n$ and $n \rightarrow n+1$, respectively. From Eq.(13), the correction to linear solution from dispersive and nonlinear terms are involved as following:

$$E_x^{n+1} = E_x^* + \frac{\{P_L^* + P_{NL}^*\} - \{P_L^{(*-1)} + P_{NL}^{(*-1)}\}}{\varepsilon_\infty \varepsilon_0}. \tag{14}$$

Evaluation of P_L, P_R, and P_K are explained in Ref.[19], based on formulation in Ref.[11], and is omitted here.

4 Numerical examples

For examples in this section, the parameters are listed in Table 1 referring to Ref.[8].

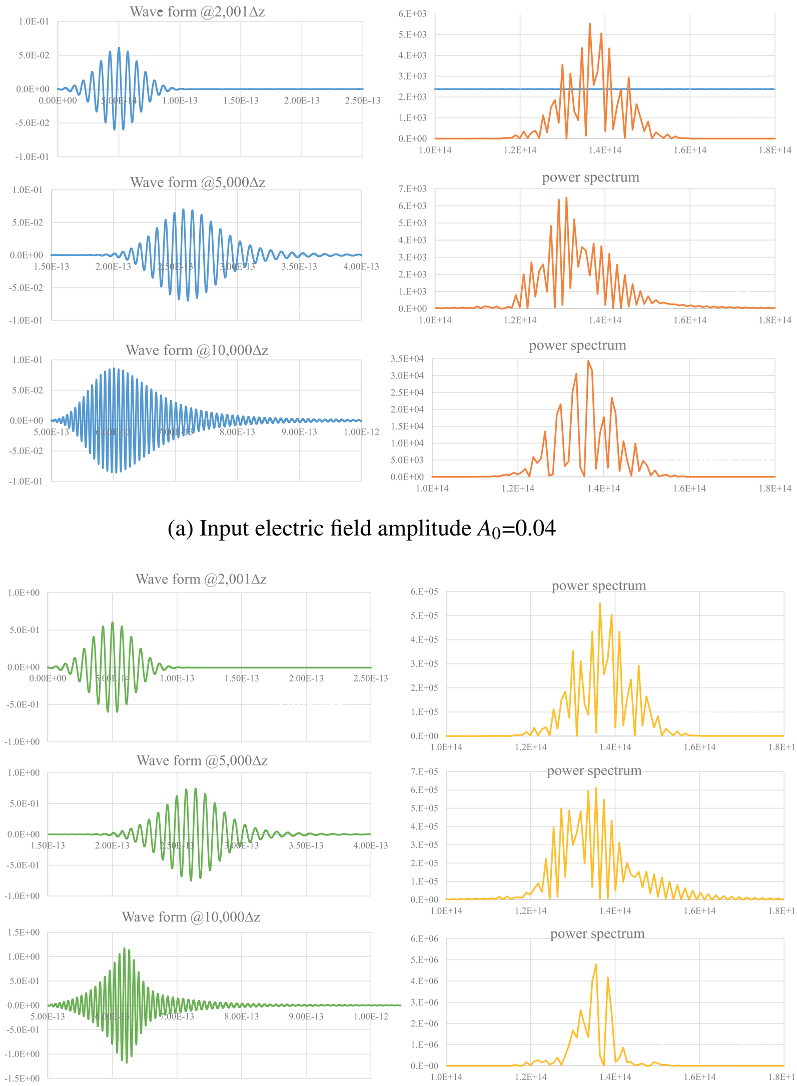
The input pulse wave is given as time-evolving Gaussian envelope pulse with carrier signal in Eqs.(15) and (16);

$$\begin{aligned} E_x(z_0, n\Delta t) &= A_0 \cos\{2\pi f_c n\Delta t\} \\ &\qquad \times \exp[-0.5\{(n - T_0)/W_0\}^2], \end{aligned} \tag{15}$$

$$H_y(z_0, n\Delta t) = E_x(z_0, n\Delta t)/Z, \tag{16}$$

where $z_0 = 2,500\Delta z$ and $Z = \sqrt{\mu_0/\varepsilon_0}$ is wave impedance of free space, respectively.

In Fig.1, propagating wave form in time domain and its power spectrum are depicted for typical time steps. In Fig.1(a), input electric field amplitude $A_0 = 0.04$ is



(a) Input electric field amplitude A_0=0.04

(b) Input electric field amplitude A_0=0.40

Fig. 1 Typical cases of envelope pulse wave forms in time domain (left column) and its power spectrum (right column).

given for case of small input amplitude to obtain dispersive behavior of the pulse. On the other hand, in Fig.1(b), larger input electric field amplitude A_0= 0.40 is given to obtain soliton-like behavior.

In both cases, power spectrum observed near excitation point are still holds Gaussian-like profiles. As the pulse wave propagates, the power spectrum shows

Discrete step size of Δz	$0.01\ \mu m$
Discrete step size of Δt	$0.167 \times 10^{-16}\ sec$
Number of discrete points NZ	25,000
Nonlinear (NL) susceptibility $\chi_0^{(3)}$	$0.07\ [V/m]^{-2}$
Relative permittivity ε_∞	2.25
Relative permittivity ε_s	5.25
Weight factor for Kerr effect α	0.7
Linear relaxation frequency f_L	$63.7\ THz$
NL relaxation frequency f_{NL}	$14.8\ THz$
Linear decaying factor δ_L	2.5×10^{-4}
NL decaying factor δ_{NL}	3.36×10^{-1}
Carrier frequency f_c	$137\ THz$
Carrier wavelength $\lambda_c = c_0/f_c$	$2.19\ \mu m$
Time to input peak T_0	$3,000\Delta t$
Width of Gaussian envelope W_0	$1,000\Delta t$

Table 1 Parameters for numerical examples

several peaks. Especially for larger input amplitude, two dominant spectrum components are observed in bottom of Fig.1(b).

5 Concluding Remarks

For numerical analysis of electromagnetic Gaussian pulse wave in dispersive and nonlinear medium, the CIP method is expanded to imply those effects by making use of Z-transform technique. From the numerical results, the simulation show physically reasonable solution for dispersive and nonlinear medium.. However, we still need to check the validity of our computation. Comparison of theoretical frequency spectrum of wave propagation in the medium with our numerical results is our future work.

Acknowledgment

This work was financially supported by KAKENHI No. 15K06043, Grant-in-Aid for Scientific Research (C) by Japan Society for the Promotion of Science(JSPS) in 2016.

References

1. A. Hasegawa and K. Kodama, "*Solitons in Optical Communications*," Oxford University Press, 1995.
2. G. P. Agrawal, "*Nonlinear Fiber Optics*," Academic Press, 1995.
3. K. S. Yee, "Numerical Solution of Initial Boundary Value Problems Involving Maxwell's Equations in Isotropic Media," IEEE Trans. on AP, Vol.14, No.3, pp.302–307, 1996.
4. R. Luebbers, F. Hunsberger, K. Kunz, R. Standler, and M Schneider, "A Frequency-Dependent Finite-Difference Time-Domain Formulation for Dispersive Materials," IEEE Trans. on EMC, Vol.32, No.3, pp.222–227, 1990.
5. R. Luebbers and F. Hunsberger, "FD-TD for Nth-order Dispersive Media," IEEE Trans. on AP, Vol.40, No.11, pp.1297–1301, 1992.
6. P. M. Goorjian, A. Taflove, R. M. Joseph and S. C. Hagness, "Computational Modeling of Femtosecond Optical Solitons from Maxwell's Equations," IEEE Jour. of QE, Vol.28, No.10, pp.2416–2422, 1992.
7. R. M. Joseph, P. M. Goorjian, and A. Taflove, "Direct Time Integration of Maxwell's Equations in Two-Dimensional Dielectric Waveguides for Propagation and Scattering of Femtosecond Electromagnetic Solitons, " Optics Letters, Vol.18, No.7, pp.491–493, 1993.
8. D. M. Sullivan, "Frequency-Dependent FD-TD Methods Using Z Transforms," IEEE Trans. on AP, Vol.40, No.10, pp.1223–1230, 1992.
9. D. M. Sullivan, "Nonlinear FD-TD Formulations Using Z Transforms," IEEE Trans. on MTT, Vol.43, No.3, pp.676–682, 1995.
10. D. M. Sullivan, "Z-Transform Theory and the FD-TD Method," IEEE Trans. on AP, Vol.44, No.1, pp.28–34, 1996.
11. D. M. Sullivan, "*Electromagnetic Simulation Using the FDTD Method (2nd Ed.)*," IEEE Press, Wiley, 2013.
12. T. Yabe, X. Feng and T. Utsumi, "The constrained interpolation profile method for multiphase analysis, "*Journal of Computational Physics*, Vol.169, pp.556-593, 2001.
13. H. Maeda, "Numerical Technique for Electromagnetic Field Computation Including High Contrast Composite Material", as Chapter 3 of *Optical Communications*, pp.41-54, InTech Open Access Publisher, Oct. 2012.
14. H. Maeda, K. Yasumoto, H. Chen, K. Tomiura and D. Ogata, "Numerical and Experimental Study on Y-shaped Branch Waveguide by PostWall", *Proceedings of 16th International Conference on Network-Based Information Systems (NBiS 2013)*, pp.508-512, Sep. 2013.
15. J. Jin, Y. Bao, H. Chen and H. Maeda, "Numerical Analysis of Y-shaped Branch Waveguide in Photonic Crystal Structures and Its Application", *Proceedings of 7th International Conference on Broadband, Wireless Computing, Communication and Applications (BWCCA-2014)*, pp.362-365, Nov. 2014.
16. H. Maeda, H. Chen, K. Tomiura and K. Yasumoto, "Numerical and experimental study on confinement in Y-shaped post wall branching waveguide,"Journal of Mobile Information Systems, Vol.10, No.2, pp.217-228, March, 2014.
17. H. Maeda, D. Ogata, N. Sakuma, N. Toyomasu, R. Nishi, "Numerical Analysis of 1×4 Branch Waveguide in Two Dimensional Photonic Crystal Structure", *Proceedings of International Conference on Advanced Information Networking and Applications (AINA2015)*, pp.366-369, March 2015.
18. H. Maeda and Y. Bao, "Numerical Analysis of Cavities in Photonic Crystal Waveguide for Filtering,"Proceedings of BWCCA-2015, pp.455-459, November, 2015.
19. H. Maeda, M. Cada, J. Jin, and K. Tomiura, "Simulation of optical pulse propagation in nonlinear and dispersive medium by constrained interpolated profile method", *Proceedings of International Conference on The Tenth International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2016)*, accepted and to be published, July 2016.

Part V
The 7-th International Workshop on Cloud, Wireless and e-Commerce Security (CWECS-2016)

Device Parameter based Secure User Authentication

Kun-Lin Tsai[1], Fang-Yie Leu[2], and King-Shing Yip[1]
[1] Department of Electrical Engineering, Tunghai University, Taichung, Taiwan
kltsai@thu.edu.tw
[2] Department of Computer Science, Tunghai University, Taichung, Taiwan
leufy@thu.edu.tw

Abstract. User authentication is a procedure used to verify a user before he/she can login a system and website. Usually, the user's account and password are employed to verify his/her identity. However, the two parameters may be hacked if the underlying network is unsafe. To provide a secure user authentication procedure, in this paper, we propose a secure user authentication scheme, named Device Parameter based User Authentication (DePUA in short), in which the hardware/software parameters of a user's trusted device, the user's password and a generated authorization code are employed to derive the device's parameter key. Without the trusted device, the adversary is unable to login the system and access user's information. The security analyses show that the DePUA can effectively verify user's identity and has the features of machine-specific device's parameter key, user anonymity, and two-factor security. Besides, it is able to resist replay attack, eavesdropping attack, and impersonation attack.

1 Introduction

To protect customers' and users' information, in the past decades, most websites or systems require users to login by using a specific user authentication procedure. Currently, inputting user's ID and password is the most popular user-authentication approach due to its convenience and simplicity. However, user's password is easy to be stolen especially when a user selects weak passwords and reuses the same passwords across different systems or websites. Typing passwords into untrusted computers may suffer password thief threat. An adversary can launch several password stealing attacks, such as phishing and some malware, to snatch passwords.

Many user authentication related studies [1-5] have improved traditional password authentication. Liao and Wang [1] proposed a robust ID-based remote user authentication scheme which involves one-time password authentication to enhance the security of user's password. Ren and Wu [2] introduced a dynamic user authentication scheme which utilizes a dynamic one-time password derived from a user's password, the authenticating time, and a unique property that the user possesses. Li *et al.* [3] utilized a smart card and Roalter *et al.* [4] used QR codes and smartphones to authorize users. Niinuma *et al.* [5] presented a framework for continuously authenticating a user's identity by employing soft biometric traits, e.g., color of user's clothing and facial skin.

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_72

Although the above schemes and systems can authorize users' identities by using different manners, some of them have been proved to be ineffective, due to their security design or additional overheads. In this paper, we present a novel approach, named Device Parameter based User Authentication (DePUA in short), in which the specific hardware/software parameters of a user's trusted device are employed for user authentication purposes, and with which the user is able to pass the user authentication procedure and then logins a protected system via a public terminal, i.e., untrusted device. During the user authentication procedure, all important parameters are encrypted by using a two-dimensional operation and Elliptic Curve Cryptography (ECC) operation. According to our security analyses, the DePUA has the features of machine-specific device's parameter key (DPK) [6], user anonymity [7], and two-factor security [8], and is able to resist replay attack, eavesdropping attack, and impersonation attack [9].

The rest of the paper is organized as follows. Section 2 introduces the related studies of this paper. Section 3 describes the system architecture and device's parameter key generation procedure. The user authentication is presented in Section 4. The security of the DePUA is analyzed in Section 5. Section 6 concludes this paper and outlines our future studies.

2 Related Studies

Many e-commerce and Internet banking services have suffered from various attacks on user authentication. Researchers have a strong desire wishing to develop and implement more secure authentication schemes to protect businesses and clients against security threats. Many user authentication studies have been proposed in the past decade. Some of them were developed based on user's password [1-2, 11-12], and some utilized smart cards, smartphones, and authorization tools [3-4].

Ren and Wu [2] employed a dynamic one-time password to provide secure user authentication. Unlike the traditional password authentication which only uses static and fixed passwords, a one-time password, utilizing a user's password, the authenticating time, and a unique property to generate the dynamic password for the user, is used only once. On next authentication, a new password is produced. Moghaddam *et al.* [10] offered a scalable user authentication scheme for cloud computing. In this scheme, two separated servers are required to store authentication and cryptography resources so as to decrease the dependency of user authentication and encryption processes from their main server. In [11], Huang *et al.* proposed the SeFEM encryption approach which presents three security schemes, including a sequential–logic style encryption/decryption mechanism, 3D operators, and a dynamic transition box, to increase the difficulty of cracking encryption keys, and effectively protect encrypted data from brute–force and cryptanalysis attacks. Shin and Kwon [12] explained the concept of remote data auditing, and described their system model and taxonomy of remote data auditing schemes. They also introduced many challenges for designing an efficient batch auditing, including high communication cost, inefficient identification protocol of corrupted data, and high computational complexity.

3 System Architecture and Device's Parameter Key

3.1 System Architecture

The DePUA architecture, as shown in Fig. 1, consists of a management server, a trusted device, a working device, and users. The management server can be the entrance of a system, a website, or any IT equipment. The trusted device can be a smartphone, notebook, or tablet computer owned by a user. When the user is using the working device and would like to login a system protected by the DePUA via the management server, he/she first keys in his/her account ID and password to the working device. The working device then sends a login request to the management server. Upon receiving the login request, the management server sends an authentication request to user's trusted device to ask the user to input his/her authentication code. Then, the trusted device generates an authentication acknowledgement message which is then sent to the management server. On receiving the authentication acknowledgement message, the management server verifies the user by comparing the device's parameter key derived from the authentication acknowledgement message with the device's parameter key currently stored in its database where the stored one is generated when the user registers with the security system. If the authentication succeeds, the management server sends back a login acknowledgement to the working device and allows the user to access the service provided by the protected system. But if the authentication fails, the management server sends an error message to the user and declines his/her request.

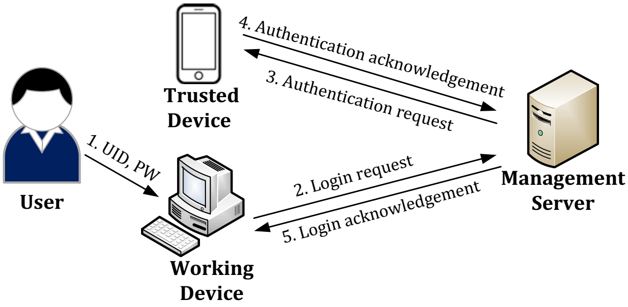


Fig. 1. System architecture of the DePUA.

3.2 Initial Phase

In the initial phase of the DePUA, the server generates system parameters and selects a large prime number $q \approx 2^r, r > 160$. After that, the server

(1) determines F_q, which is a finite field of characteristic *2* and the order of q;

(2) determines the elliptic curve $E: y^2 \equiv x^3 + ax + b\ (mod\ q)$, where $a, b \in F_q$ and $4a^3 + 27b^2 \not\equiv 0(\text{mod } q)$;

(3) selects a base point P of order n on E, and $q^k \not\equiv 1\ (mod\, n)$ for any k, $1 \leq k < 100$;
(4) generates a cyclic additive group G by P, the order of which is n;
(5) chooses a secure hash function $H_1: U \rightarrow Z_n^*$;
(6) publishes the system parameters $\{E, P, H_1\}$.

3.3 Device's Parameter Key (DPK) Generation Procedure

The DPK is similar to the environmental key introduced in [6]. A trusted device D's parameters, which can be IMEI code, CPU ID, Mac address of NIC, disk serial number, etc., are used to create a unique DPK. By utilizing D's DPK, a user can be successfully verified by the management server. Assuming that there are w parameters of D which are numbered from 1 to w, the DePUA creates a device parameter table, as shown in Table 1, in the trusted device. The DPK generation procedure is as follows.

(1) Set the user's authentication code n_{ver} and his/her password as the inputs of hash function H_1;
(2) Use $H_1(n_{ver}\|k_{PW})$ to generate a parameter sequence, e.g., {2, 2, 1, …, 6, 3};
(3) Use the parameter sequence to access the corresponding parameter numbers recorded in device parameter table, as the inputs of hash function H_2, to generate the DPK
DPK $= H_2(IMEI\ code||IMEI\ code||MAC\ address||\ ...\ ||Disk\ serial\ number)$
where H_2 is defined as $H_2: U^w \rightarrow Z_n^*$.

Table 1. Device parameter table of a trusted device.

Parameter No.	Component	ID number
1	MAC address of NIC	BCEE7BDCB8BA
2	IMEI code	355931046825508
3	Disk serial number	9VPD54XY
4	OS serial number	00359-89-9267-006-6787
…	…	…
w	CPU_ID	06F110

4 User Registration and Authentication Procedures

When a user would like to utilize a protected system's service, he/she must firstly register himself/herself with the system, so that the system can store the user's DPK in its key-management database and use it to authenticate the user when the user requests to login the system. The user registration procedure will be described in Section 4.1. After that, the registered user can login the protected system by using user authentication procedure which will be presented in Section 4.2.

4.1 User Registration Procedure

Following the procedure of the user registration, a new user and his/her trusted device's information are recorded in the management server (from now on, MS for short). The user registration procedure has four rounds. In Round 1, a user sends a random number r_A and his user ID *UID* to the MS. In Round 2, the MS generates an authentication code for the user. In Round 3, the trusted device calculates the DPK k_{DP} according to n_{ver}. Finally, the MS stores the k_{DP} in its key-management database for future authentication. The detailed descriptions for these rounds are as follows.

Round 1: by the user

The user first

(1) generates a random number r_A, $r_A \in Z_n^*$, and stores it in the trusted device;
(2) fetches the system time t_{nonce}, with which to derive k_{ct};
(3) calculates $R_B \equiv (r_B \oplus UID) \cdot P$ by using ECC, where *UID* is user's ID;
(4) calculates $r_A' = [r_A \oplus (k_{ct} +_2 k_{PW})] +_2 (k_{ct} \oplus k_{PW})$ where k_{PW} is user's password key;
(5) sends $\{UID, R_A, r_A', t_{nonce}\}$ to MS;

Round 2: **by the management server MS**

The management server MS

(1) fetches the MS's system time $t_{nonce,MS}$ and checks to see whether or not $t_{nonce,MS} - t_{nonce} \leq \Delta t$ where Δt as a predefined time threshold is the allowable maximum transmission delay from the user to MS. If not, MS sends an authentication-failure message to the user and terminates the procedure. Otherwise, it
(2) derives k_{ct} from t_{nonce};
(3) calculates $r_{A,C} = (r_A' -_2 (k_{ct} \oplus k_{PW})) \oplus (k_{ct} +_2 k_{PW})$ and $R_{A,C} \equiv (r_{A,C} || UID) \cdot P$;
(4) checks to see whether $R_{A,C} = R_A$;
If not, MS sends an authentication-failure message to the user and terminates the procedure. Otherwise, it
(5) generates a random number r_{MS}, $r_{MS} \in Z_n^*$;
(6) calculates the authentication code $n_{ver} = H_3(r_{A,C} \oplus r_{MS})$ and $N_{ver} \equiv (n_{ver} \oplus UID) \cdot P$, where H_3 is a one-way hash function;
(7) calculates $n_{ver}' = (n_{ver} \oplus r_A) +_2 k_{ct} +_2 (r_A \oplus k_{PW})$;
(8) sends $\{N_{ver}, n_{ver}'\}$ to the user;

Round 3: by the user

Upon receiving the message, the user

(1) calculates $n_{ver,C} = (n_{ver}' -_2 k_{ct} -_2 (r_A \oplus k_{PW})) \oplus r_A$;
(2) calculates $N_{ver,C} \equiv (n_{ver,C} \oplus UID) \cdot P$ and checks to see whether $N_{ver,C} = N_{ver}$ or not;
If not, the user sends an authentication-failure message to MS and terminates the procedure. Otherwise, it
(3) inputs $n_{ver,C}$ and k_{PW} to the trusted device and follows the DPK generation procedure described in Section 3.3 to generate device's parameter key k_{DP};
(4) computes $k_{DP}' = (k_{DP} \oplus r_A) +_2 n_{ver,C} \oplus (r_A +_2 k_{ct})$;

(5) computes $K_{DP} \equiv k_{DP} \cdot P$ by using ECC;
(6) sends $\{K_{DP}, k'_{DP}\}$ to MS;

Round 4: by MS

Upon receiving the message, MS

(1) calculates $k_{DP,C} = (k'_{DP} -_2 n_{ver} \oplus (r_A +_2 k_{ct})) \oplus r_A$;
(2) computes $K_{DP,C} \equiv k_{DP,C} \cdot P$ and checks to see whether $K_{DP,C} = K_{DP}$ or not; If not, it sends an authentication-failure message to the user and terminates the procedure. Otherwise, it
(3) stores *UID*, $k_{DP,C}$, r_{MS}, and $r_{A,C}$ to its key-management database.

4.2 User Authentication Procedure

When the registered user would like to login the protected system via a working device, he/she needs to pass the user authentication procedure which has five rounds. In Round 1, the user sends a request to MS. In Round 2, the MS verifies the request, generates two authentication codes, i.e., previous code and new code, and sends them to the user and his/her trusted device. In Round 3, according to the two authentication codes, the trusted device calculates two DPKs which are then sent to MS, and the MS verifies these DPKs in Round 4. If the DPKs pass the authentication, MS sends an acknowledgement message to the user. After that it allows the user to login the system via the working device; otherwise, MS rejects user's request. The detailed descriptions for all rounds are as follows.

Round 1: by the user with the working device

The user first

(1) generates a random number r_B, $r_B \in Z_n^*$;
(2) generates a login request message REQ_{login};
(3) fetches the system time t'_{nonce}, with which to derive k'_{ct};
(4) calculates and $r'_B = ((r_B || UID) \oplus (k'_{ct} +_2 k_{PW})) +_2 (k'_{ct} \oplus k_{PW})$;
(5) calculates $R_B \equiv (r_B \oplus (UID || REQ_{login})) \cdot P$;
(6) calculates $REQ'_{login} = (REQ_{login} +_2 r_B) \oplus (k'_{ct} \oplus r_B)$;
(7) sends $\{REQ'_{login}, R_B, r'_B, t'_{nonce}\}$ to MS.

Round 2: by MS

In this round, MS has two stages. The first is verifying the login request message sent by the user, and the second is sending an authentication request to the trusted device.

In Stage 1, MS

(1) fetches the system time $t'_{nonce,MS}$ and checks to see whether or not $t'_{nonce,MS} - t'_{nonce} \leq \Delta t$ where Δt as a predefined time threshold is the allowable maximum transmission delay from the user to MS;
(2) derives k'_{ct} from t'_{nonce};
(3) calculates $REQ_{login,C} = (REQ'_{login} \oplus (k'_{ct} \oplus r_B)) -_2 r_B$;
(4) calculates $r_{B,C} || UID = (r'_B -_2 (k'_{ct} \oplus k_{PW}) \oplus (k'_{ct} +_2 k_{PW})$, fetches $r_{B,C}$ and UID , and calculates $R_{B,C} \equiv (r_{B,C} \oplus (UID || REQ_{login,C})) \cdot P$;

(5) checks to see whether $R_{B,C} = R_B$;
If not, MS sends an authentication-failure message to the user and terminates the procedure. Otherwise, it continues the procedure of Stage 2.

In Stage 2, MS

(1) fetches $k_{DP,C}$, r_{MS}, and $r_{A,C}$ from its key-management database, according to *UID*;
(2) calculates $n_{ver} = H_3(r_{A,C} \oplus r_{MS})$ and $N_{ver} \equiv (n_{ver} \oplus UID) \cdot P$;
(3) calculates $n'_{ver} = (n_{ver} \oplus r_{A,C}) +_2 r_{A,C}$;
(4) generates a random number r'_{MS}, $r'_{MS} \in Z_n^*$;
(5) calculates $n_{ver\_new} = H_3(r_{A,C} \oplus r'_{MS})$ and $N_{ver\_new} \equiv (n_{ver\_new} \oplus UID) \cdot P$;
(6) calculates $n'_{ver\_new} = (n_{ver\_new} \oplus r_{A,C}) +_2 r_{A,C}$;
(7) sends $\{N_{ver}, n'_{ver}, N_{ver\_new}, n'_{ver\_new}\}$ to the trusted device.

Round 3: by the user with his/her trusted device

Upon receiving the message, the user in Round 3

(1) calculates $n_{ver} = (n'_{ver} -_2 r_A) \oplus r_A$;
(2) calculates $N_{ver,C} \equiv (n_{ver,C} \oplus UID) \cdot P$ and checks to see whether $N_{ver,C} = N_{ver}$ or not. If not, it sends an authentication-failure message to MS and terminates the procedure. Otherwise, it
(3) inputs $n_{ver,C}$ and k_{PW} to generate device's parameter key k_{DP} following the DPK generation procedure;
(4) calculates $k'_{DP} = (k_{DP} \oplus r_A) +_2 n_{ver,C} \oplus (r_A +_2 k_{PW})$ and $K_{DP} \equiv k_{DP} \cdot P$ by using ECC;
(5) calculates $n_{ver\_new} = (n'_{ver\_new} -_2 r_A) \oplus r_A$;
(6) calculates $N_{ver\_new,C} \equiv (n_{ver\_new,C} \oplus UID) \cdot P$ and checks to see whether $N_{ver\_new,C} = N_{ver\_new}$ or not.
If not, it sends an authentication-failure message to MS and terminates the procedure. Otherwise, it
(7) inputs $n_{ver\_new,C}$ and k_{PW} to the trusted device to generate device's parameter key k_{DP_new} following the DPK generation procedure;
(8) calculates $k'_{DP\_new} = (k_{DP\_new} \oplus r_A) +_2 n_{ver\_new} \oplus (r_A +_2 k_{PW})$ and $K_{DP\_new} \equiv k_{DP\_new} \cdot P$ by using ECC;
(9) sends $\{K_{DP}, k'_{DP}, K_{DP\_new}, k'_{DP\_new}\}$ to MS;

Round 4: by MS

In this round, MS has two stages. The first, i.e., Stage 1, is verifying the DPK received from the trusted device, and the second, i.e., Stage 2, is sending an acknowledgement message to the working device.

In Stage 1, MS

(1) calculates $k_{DP\_new,C} = (k'_{DP\_new} -_2 n_{ver\_new} \oplus (r_{A,C} +_2 k_{PW})) \oplus r_{A,C}$;
(2) calculates $K_{DP\_new,C} \equiv k_{DP\_new,C} \cdot P$ and checks to see whether $K_{DP\_new,C} = K_{DP\_new}$ or not;
If not, it sends an authentication-failure message to the user and terminates the procedure. Otherwise, it
(3) stores $k_{DP\_new,C}$ and r'_{MS} in its key-management database;
(4) calculates $k_{DP,C} = (k'_{DP} -_2 n_{ver} \oplus (r_{A,C} +_2 k_{PW})) \oplus r_{A,C}$;

(5) calculates $K_{DP,C} \equiv k_{DP,C} \cdot P$ and checks to see whether $K_{DP,C} = K_{DP}$ or not;
If not, it sends an authentication-failure message to the user to reject the user request and terminate the procedure. Otherwise, it continues the procedure of Stage 2.

In Stage 2, MS

(1) calculates $REQ_{login\_ACK} = (REQ_{login} \oplus r_B) -_2 k'_{ct}$;

(2) calculates $REQ'_{login\_ACK} = (REQ_{login\_ACK} +_2 k_{PW}) \oplus (r_B +_2 k'_{ct})$;

(3) calculates $K_{REQ\_ACK} \equiv REQ_{login\_ACK} \cdot P$;

(4) sends $\{REQ'_{login\_ACK}, K_{REQ\_ACK}\}$ to the working device.

Round 5: by the user with the working device

Upon receiving the message sent by MS, the user in Round 5

(1) calculates $REQ_{login\_ACK,C} = (REQ'_{login\_ACK} \oplus (r_B +_2 k'_{ct})) -_2 k_{PW}$;

(2) calculates $K_{REQ\_ACK,C} \equiv REQ_{login\_ACK,C} \cdot P$ and checks to see whether $K_{REQ\_ACK,C} = K_{REQ\_ACK}$ or not.
If not, it sends an authentication-failure message to MS and terminates the procedure. Otherwise, the user logins the system through the working device.

5 Security Analysis

(1) *Machine-specific DPK*

In the DePUA, the DPK is derived from the authentication code and the trusted device's parameters. To our knowledge, system parameters, like IMEI, MAC address of NIC, and CPU ID, are individually unique among all devices having been produced in the world. Hence, it is almost impossible for users to generate the same DPK in different devices given the same authorization code. As a result, when an adversary eavesdrops the message and decrypts the authorization code, he/she is still unable to generate correct DPK, meaning that in the DePUA, the DPK generation procedure is practically secure.

(2) *User Anonymity*

In the authorization procedure of the DePUA, the user is anonymous. The login request message sent to MS in Round 1 of the user authentication procedure, i.e., $\{REQ'_{login}, R_B, r'_B, t'_{nonce}\}$, contains no *UID* in plain text. The *UID* is hidden in r'_B and will be computed and fetched in step (4) of Round 2. Each parameter transmitted between MS and the user is secured by invoking a two-dimensional operation and ECC operation. This actually provides the DePUA with the user anonymity property.

(3) *Two-factor security*

The two security factors owned by the DePUA are the user's password and the trusted device's hardware/software parameters. It is obvious that if both security factors are known by adversary, the correct DPK also cannot be generated. Consequently, the procedure will fail. In other words, the adversary cannot launch an attack simply when he/she knows any one of the two factors, i.e., it is impossible for him/her to issue a valid message to MS without knowing the two security factors.

(4) *Replay attacks*

In this type of attack, an adversary tries to eavesdrop a valid request message sent by user to MS, and then resends this message again to MS, trying to gain all the information provided by MS. In the DePUA, the time key k_{ct} is derived in Round 1 of both user registration and user authentication procedures. When an adversary intercepts the message sent by the user to MS, there may be two cases. First, the adversary keeps the original message without modifying it and pretends the corresponding legal user to transmit this message to the server. In this case, $t_{nonce,MS} - t_{nonce} \leq \Delta t$ does not hold because the retransmission is delayed where Δt is the maximal time required by a message to be transmitted from the user to MS. Second, the adversary modifies the time t_{nonce} to make $t_{nonce,MS} - t_{nonce} \leq \Delta t$. However, the verification step (i.e., step 2) of Round 2 in both procedures would fail because the time key is different from the original one. Therefore, the DePUA is invulnerable to replay attacks.

(5) *Eavesdropping attack*

Assuming that an adversary captures a large number of messages from the underlying network, wishing to extract sensitive information, such as DPK, from these messages. In this scheme, the abovementioned parameters are encrypted by random numbers, time key, and user's password, i.e., r_A, k_{ct} and k_{PW}. In the user authentication procedure, the time key k_{ct} varies in different sessions. Even a large amount of messages are captured, the adversary is still unable to extract these parameters from these messages. Thus, the DePUA is able to thwart the eavesdropping attack.

(6) *Impersonation attack*

When wanting to impersonate a user, an adversary intercepts the message sent by the MS to the user in Round 2 of the user authentication procedure, trying to derive n_{ver} from n'_{ver}. However, due to the lack of correct r_A, the adversary cannot decrypt accurate n'_{ver} to obtain n_{ver}, and hence cannot correctly generate k_{DP}. Thus, the adversary cannot pass the verification performed in step (2) of Round 3, showing that the DePUA can effectively defend the impersonation attack.

6 Conclusions and Future Studies

To verify a user's identity with a secure method, in this study, the DePUA scheme is proposed for those systems and websites which need high level of security with the help of the trusted device. The device's parameter key derived from the user's password, MS's authentication code and the trusted device's hardware/software parameters is utilized in the user authentication procedure to check the consistence between the underlying user and a registered user. According to our security analyses, the DePUA has the features of machine-specific DPK, user anonymity, and two-factor security, and is able to resist replay attack, eavesdropping attack, and impersonation attack. As a result, the DePUA is very suitable for Internet banking and e-commerce.

In the near future, we would like to improve the user registration procedure as well as the user authentication procedure. Although the DePUA has a high level of security, the complex steps used for DPK generation and authentication need to be improved so that the trusted device and MS's processing time can be further reduced. We would also like to derive the reliability and behaviour models of the DePUA so that users can predict the system reliability and its behaviour before using it. These constitute our future studies.

References

1. Liao, Y.P., Wang, S.S.: A Robust Password-Based Remote User Authentication Scheme Using Bilinear Pairings without Using Smart Cards. In: International Computer Symposium, pp. 215–221. Taiwan (2010)
2. Ren, X., Wu, X.W.: A Novel Dynamic User Authentication Scheme. In: International Symposium on Communications and Information Technologies, Gold Coast, pp. 713–717. USA (2012)
3. Li, X., Ma, J., Wang, W., Xiong, Y., Zhang, J.: A Novel Smart Card and Dynamic ID based Remote User Authentication Scheme for Multi-server Environment. Mathematical and Computer Modelling. 58(1–2), 85–95 (2013)
4. Roalter, L., Kranz, M., Diewald, S., Möller, A.: The Smartphone as Mobile Authorization Proxy. In: International Conference on Computer Aided Systems Theory, pp. 306–307. Japan (2013)
5. Niinuma, K., Park, U., Jain, A.K.: Soft Biometric Traits for Continuous User Authentication. IEEE Transactions on Information Forensics and Security, 5(4), 771–780 (2010)
6. Tsai, K.L., Leu, F.Y., Tsai, S.H.: Data Encryption Method using Environmental Secret Key with Server Assistance. Intelligent Automation and Soft Computing, 22(3), 423–430 (2016)
7. Schreck, J.: Security and Privacy in User Modeling. Springer Science+Business Media. (2003)
8. Wang D., Wang, P.: Understanding Security Failures of Two-factor Authentication Schemes for Real-time Applications in Hierarchical Wireless Sensor Networks. Ad Hoc Networks, 20, 1–15 (2014)
9. Tsai, K.L., Huang, Y.L., Leu, F.Y., Tan, J.S., Ye, M.Y.: High-efficient Multi-Key Exchange Protocol based on Three-party Authentication. In: International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, Birmingham, pp. 487–492. UK (2014)
10. Moghaddam, F.F., Moghaddam, S.G., Rouzbeh, S., Araghi, S.K., Alibeigi, N.M., Varnosfaderani, S.D.: A Scalable and Efficient User Authentication Scheme for Cloud Computing Environments. In: IEEE Region 10 Symposium, pp. 508–513. Malaysia (2014)
11. Huang, Y.L., Dai, C.R., Leu, F.Y., You, I.: A Secure Data Encryption Method Employing a Sequential-Logic Style Mechanism for a Cloud System. International Journal of Web and Grid Services, 11(1), 102–124 (2015)
12. Shin S., Kwon, T.: A Survey of Public Provable Data Possession Schemes with Batch Verification in Cloud Storage. Journal of Internet Services and Information Security, 5(3), 37–47 (2015)

Data Preprocessing Quality Management Procedure for Improving Big Data Applications Efficiency and Practicality

Sen-Tarng Lai[1], Fang-Yie Leu[2]
[1] Dept. of Information Technology and Management, Shih Chien University,
Taipei, 10462,Taiwan
e-mail: stlai@mail.usc.edu.tw
[2] Dept. of Computer Science, Tunghai University
Taichung, 40704, Taiwan
e-mail: leufy@thu.edu.tw

Abstract. Diversification applications of network fully combined with the people's daily activities and life. All network activities generate and record the large amount of data that implies the business values of enterprises and organizations. Collecting, analyzing and visualizing the large amount of data, intelligent information may be efficiently extracted. Big data applications can help enterprises enhance market competitiveness advantages, and assist government units improve the people daily life quality. However, big data collected from network and IoT (Internet of Things) environment existed many quality defects and problems to be resolved. Data quality of big data will directly impact the analysis results of big data, and may cause wrong decisions, inaccurate predication, imperfect planning and arrangements. Data preprocessing is an important procedure of big data applications. How to ensure data preprocessing tasks quality has become a concern issue of big data applications. Based on the review activities, this paper proposes the Preprocessing Tasks Quality Measurement (PTQM) model to identify the quality defects of data preprocessing tasks. Applying Data Preprocessing Quality Management (DPQM) procedure timely modifies the preprocessing tasks quality defects to increase the big data applications efficiency and practicality.

1 Introduction

In network and IoT (Internet of Things) age, big data becomes the important and useful assets of enterprises and organizations to enhance market competitiveness advantages. The data of market analysis, E-commerce transactions and social network activities all need be effectively collected, appropriately managed and suitably stored. And using high efficiency statistic and analysis tools for speedy generating the visualization information to help enterprises and organizations make correct decision and accurate prediction [1]. Business promotion, traffic control, weather prediction, health

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_73

management all can use big data applications to assist decision making, prediction, planning and arrangement [2], [3], [4]. Consumer behavior view, big data applications technology can forecast the trend of consumer group. Assisting high level managers make the best planning and decisions to increase incomes, profits and the advantage of market competition. Daily life view, big data applications to traffic control can reduce time of traffic jam, accurate meteorological prediction may be reducing the calamities losses, suitable crop planting can improve the quality and yield, in health management can efficiently control the disease. Big data can efficiently improve people's quality of life. In network and IoT age, big data applications have become the worthy further exploration of important topics [2], [4].

Big data applications need great space to storage the large amount of data and high efficiency facilities to handle sustained rapid growth data. In addition, data collection from different web sites or devices has many formats and multi-style contents. Big data has volume, variety, velocity and veracity four characteristics [5]. These characteristics have closely related with big data applications processing. However, most of the network data has unconfirmed contents and uninspected quality [6], [7], [8]. Data bad quality of big data is bound to impact the results of big data analysis, and directly influence decisions, forecasting and planning operations. How to verify the data quality of big data has become the necessary concern issues of enterprises and organizations [6], [7], [8].

Bad quality data causes incorrect analysis results and may form the wrong decision or inaccurate prediction. Data quality will directly impact the results of the analysis of big data. In order to enhance the big data applications effectiveness and practicality, big data processing procedure must confirm the data quality in advance. In this paper, based on the task review activities, a Preprocessing Tasks Quality Measurement (PTQM) Model is proposed to identify the preprocessing tasks quality defects. Applying Data Preprocessing Quality Management (DPQM) procedure timely modifies data preprocessing tasks quality defects to increase the big data applications efficiency and practicality. In Section 2, discusses the four characteristics and the challenges of big data applications. In Section 3, describes the importance of data preprocessing. Based on data preprocessing tasks review activities, in Section 4, proposes a PTQM model and DPQM procedure. In Section 5, evaluates the efficiency of DPQM procedure. In Section 6, emphasizes the important of data preprocessing and makes a conclusion of this topic.

2 Big Data Characteristics and Challenges

2.1 Four characteristics of big data

General enterprise or organization internal data processing can base on the self-defined or industry standard data format. The internal data has controllable data amount, predictable data growth and the structured data format. Traditional relational database enough store and manage internal data, commercial statistical software sufficient process business data. However, in network and IoT environment, the big data collection from different web sites or devices has many formats and multi-style contents. Big data has a very large amount of data and sustained rapid growth data. Volume, variety, velocity and veracity are four characteristics of big data. These characteristics have many differences with the internal data of general enterprises and organizations (shown as Table 1). The following instructions discuss four key characteristics of big data [5]:

- Volume: gathering tool through the Internet can quickly collect enormous amounts of data in short time. It is several hundred thousand times of normal internal data. The large amounts data collected from different web sites or devices need perfect management and storage mechanisms than internal data.
- Variety: data format highly diversified, it may belong to structural data or unstructured multimedia data, such as sound, video, drawing, feeling, etc. Unstructured data is not able to store into the structured repository (SQL-based Data Base). In addition, data format also isn't the enterprises or organizations previously defined internal data format.
- Velocity: all kinds of online transactions and community activities are not subject to time limits. Data analysis efficiency cannot catch up the speed of data generation, then may affect the expected results of the data analysis and cause wrong decisions or inaccurate predictions. Enterprises or organizations need hire the professional data analysts and data scientists to increase data analysis efficiency.
- Veracity: This characteristic is an important feature was recently introduced. Because the data veracity will have a direct impact big data analysis results. The truthless data of web sites seriously mislead the big data analysis results and cause big data applications wrong decisions and inaccurate predictions.

Table 1. Difference analysis between big data and internal data

Features of data	Big Data	Internal Data
Data Format	Nonstructural	Structural
Data growth	Speedy	Stable
Data amount	Monumental	Manageable
Storage style	NoSQL	SQL-based
Analysis Tools	Model building	Commercial packages

2.2 Challenges of big data applications

Big data applications can obtain many business opportunities to adjust enterprises or organization business and promotion model, specifically to enhance business performance. In addition, big data applications can also assist governments to accurately predict, prevent natural disasters and develop the proper policy, improve the quality of people life. However, before successfully achieve above results, big data applications must overcome the technical problems and data quality defects. Four characteristics of big data also becomes the challenges of big data applications [6], [7], [8]:

- Data collection challenges: Data collection tools through the Internet can easy collect the data for analyzing the business pricing strategy, community discussion topics and customer transaction trends. However, in order to guard against competitors, many enterprises and organizations started to take protective measures for the anti-collecting network data. Planning the security devices to store critical data, using some special words or data to confuse data collection tools. These new data protection measures become the critical challenges of data collection of big data applications.

- Data analysis challenges: applying the cleaned and clustered data, data scientists and data analysts construct data mode and develop the appropriate analysis methods. Further, data analysis must be completed within the required time, otherwise it will affect the subsequent processing steps and lose advantages. Big data market lacks the experienced data scientists and data analysts, has become another challenge for data analysis.
- Data requirements challenges: enterprises and organizations failed to propose the complete and clear data requirements. Incomplete and unclear data requirement often cause the big data applications difficultly to reach the desired objectives. Big data applications need a mechanism to confirm the consistency between data requirements and the expected results of the big data.
- Data security challenges: data collection and analysis often involve data security issue and cause many disputes. Therefore, data collection and analysis must pay much attention to data security to protect the personal data and avoid to misuse the personal data.
- Data quality challenges: collecting a lot of repetition, no analytical value, truthless or erroneous data, not only waste data process time and human resources. The analysis results of low quality data may cause the wrong decision or inaccurate pre-diction, and even to generate unpredictable serious consequences.

The big data applications main challenges and predicaments are related to the critical issues of data quality. Data preprocessing methods are necessary to improve the processes quality of big data.

3. Importance of Big Data Preprocessing

IoT devices generation data, network transmission data and database storage data always exists several unexpected contents. The collected data contains incomplete, inconsistent, incorrect, noisy, and abnormal data often makes wrong decisions or inaccurate predictions. Causing the big data applications can not satisfy the expected objective. Low quality data causes wrong decision or inaccurate predictions, takes more analysis time, and generates unclear information. For improving the data analysis efficiency and quality, before big data applications processing, the incomplete, unsuitable and abnormal data should be indeed identified, patched and removed. Low quality data causes four unfavorable results of big data applications process:

- Imperfect data aggregate/storage: In big data processing, duplicate data or same data did not be removed or combined. Making data can’t effective and perfectly be aggregated and stored. Imperfect data aggregate/storage need take more manpower and resource for the big data processing.
- Low efficiency data statistical/analysis: In big data processing, high efficiency data statistical/analysis is a necessary condition to timely complete correct deci-

sion making and accurate prediction. Inconsistent data format or unintegrated data causes data statistical/analysis tasks cannot be completed timely.

- Data attributes unidentified: In big data processing, data attributes unidentified may cause unfavorable data clustering and management.
- Incorrect decision/inaccurate predication: The collected data contains incomplete, inconsistent, incorrect, noisy and unmoral data often makes wrong decisions or inaccurate prediction. Causing big data applications can't satisfy the expected objectives.

For getting the correct decisions, accurate predictions, perfect planning and arrangements, after data collection, big data applications must previously process low quality data. Four major tasks of data preprocessing [9], [10] for big data applications are discussed as follows:

- Data cleansing: Data cleansing is a first priority task for data preprocessing of big data applications. For assuring data quality, at first data cleaning need identify the data defects and problems. Then, according the types of data defects and problems, fill in missing values, smooth noisy data, recheck or remove abnormal data, and adjust the incomplete or inconsistent data.
- Data transformation: Big data collected from widely areas and different environments. It needs take more time and resource to handle and analyze the different formats data. Therefore, the collected data should normalize into a unified format and aggregate into the suitable clusters.
- Data integration: For increasing data analysis efficiency, data integration of multiple databases, data storages or files is an important task. High quality data integration can help reduce and avoid redundancies and inconsistencies in the stored data set. Data integration can help improve the accuracy and speed of the subsequent data analysis process.
- Data reduction: The collected data may exist the duplicate or similar contents. For reducing data volume and data analysis time, the duplicate or similar data need be identified and merged or removed. It is necessary to use the duplicate or similar data recognition tools for avoiding the incorrect data deleting.

4. Quality measurement model and management procedure

4.1 Preprocessing task quality measurement model

In order to effectively monitor and assess preprocessing task quality defects, individual measurements should make the appropriate combination [11]. In this paper, LCM (Linear Combination Model) is applied to preprocessing task quality measurement. The different level quality activities have different quality metrics be shown. Therefore, before LCM starting, the quality factors must be normalized. The normalized value is between 0 and 1. The best quality quantified value approaches to 1 and the worst quality approaches to 0. Data preprocessing tasks should consider three quantified quality that include planning quality, achievement quality, and tools using quality. Using the LCM, related quality factors can be combined into the quality metric, and then the related quality metrics can be combined into the task quality measurements:

- Task planning Quality Measurement (TPQM) is combined resource, schedule, and sequence three management metrics. Combination formula shows as Equation (1):

ReM: Resource Management Metric	*W1: Weight of ReM*
ScM: Schedule Management Metric	*W2: Weight of ScM*
SeM: Sequence Management Metric	*W3: Weight of SeM*

$$TPQM= W1* ReM + W2* ScM+ W3*SeM \qquad W1+ W2 + W3 =1 \qquad (1)$$

- Task Achievement Quality Measurement (TAQM) is combined coverage rate, complete rate and complete efficiency three quality metrics. Combination formula shows as Equation (2):

CvRM: Coverage Rate Metric	*W1: Weight of CvRM*
CmRM: Complete Rate Metric	*W2: Weight of CmRM*
CoEM: Complete Efficiency Metric	*W3: Weight of CoEM*

$$TAQM= W1* CvRM + W2* CmRM+ W3*CoEM \qquad W1+ W2 + W3 =1 \qquad (2)$$

- Tools Using Quality Measurement (UTQM) is combined suitability, existing tools and tools efficiency three quality metrics. Combination formula shows as Equation (3):

SM: Suitability Metric	*W1: Weight of SM*
ETM: Existing Tools Metric	*W2: Weight of ET*
TEM: Tools Efficiency Metric	*W3: Weight of TEM*

$$TUQM= W1* SM + W2* ETM+ W3*TEM \qquad W1+ W2 + W3 =1 \qquad (3)$$

Finally, combining TPQM, TAQM and TUQM into a Task Quality Indictor. For improving task quality, the related activities of low quality should be rigorously inspected to identify the problems or defects and propose the corrective action.

4.2 Data preprocessing quality management procedure

Major purpose of data preprocessing is to improve data quality of big data applications. Based on major tasks of data processing [9], [10], the data preprocessing quality management (DPQM) procedure is divided into data cleansing, data deduction/transformation, and data integration three phases and group to two steps (shown as Figure 1).

- Data cleansing phase: In data collection process, data cleansing should be started for identifying the defects and problems of the collected data. According to data defects and problems, fill in missing values, smooth noisy data, modify or remove outliers, and resolve inconsistencies.
- Data integration/transformation phase: For improving data storage and management efficiency, in data deduction/transformation phase, the related individual data items or components must be suitably combined into a useful data groups and special format data need be converted to the standard formats.
- Data reduction phase: In order to increase data analysis efficiency, the duplicate data must be removed or be merged.

For ensuring the phase tasks quality, before end of phase, task review activities must be indeed executed. Task review activities combine the PTQM model to monitor the phase task quality. Quantified quality measurement meets the criteria, the DPQM procedure can enter into the next phase. Otherwise, according to the quality defects, this phase tasks should be required to redo.

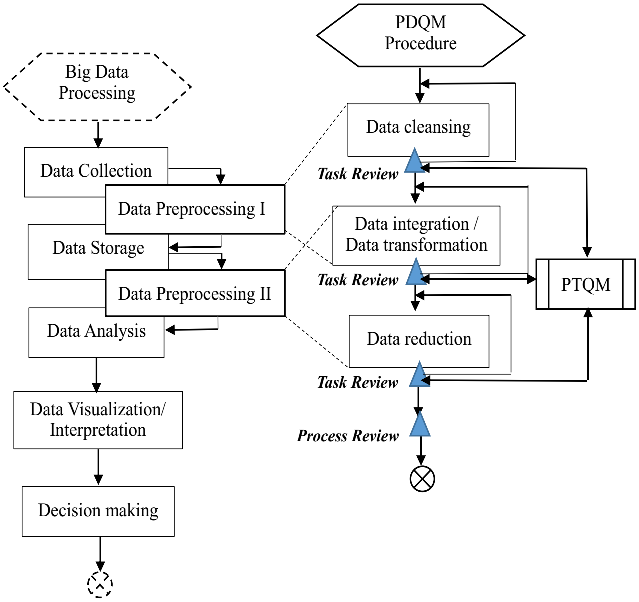


Figure 1. Flowchart of PDQM procedure

5. Evaluation of PDQM procedure

Data preprocessing is an important procedure for big data applications. How to inspect and improve the quality of data preprocessing tasks is the necessary activities to assure the quality and practicality of big data applications. Based on PTQM model, this paper proposed DPQM procedure to identify the tasks defects and improve the quality of data preprocessing tasks. DPQM procedure combines the quantified measurements to monitor and control the quality of data preprocessing tasks and improve effectively data quality of big data applications. For evaluating the advantages of DPQM procedure, four major impact items of the quality and usability of big data applications are discussed. Using four impact items include accurate predication/ correct decision, statistic/analysis efficiency, data aggregate/storage effects, and data attributes identification to evaluate the DPQM procedure (shown as Table 2).

Table 2. DPQM procedure evaluation table

Big data main tasks	Use DPQM procedure	Use data preprocessing	Omit data preprocessing
Accurate predication/ correct decision	More certain	Certain	Uncertain
Statistic/Analysis efficiency	High	Improvable	Unexpected
Data aggregate/storage effects	High	Middle	Low
Data attributes identification	High	Middle	Low

6. Conclusion

Applying low quality data or inappropriate data always produces an incorrect analysis results and may form the wrong decision or inaccurate predictions to cause unpredictable serious consequences. In advance to confirm the data quality that may enhance the effectiveness and practicality of big data analysis. For improving data quality of big data, this paper discussed the critical tasks of data preprocessing. And based on task review activities, a PTQM Model is proposed to identify data preprocessing tasks quality defects. Applying the DPQM procedure timely modifies the preprocessing tasks quality defects to increase the big data applications quality and effectiveness. Before big data statistic and analysis process, the incomplete and error data must be modified or removed to improve analysis efficiency and enhance the efficiency and practicality of big data applications. The concrete contributions of combination DPQM procedure with PTQM Model are described as follows:

- Applying quantified review activities to timely identify data preprocessing task quality defects and problems.
- Timely modify data preprocessing tasks quality defects and concretely increase big data analysis and visualization efficiency.
- Improvement data quality can reduce big data applications wrong decisions, inaccurate predication, imperfect planning and arrangements.

Acknowledgments. The research was supported by Ministry of Science and Technology research project funds (Project No.: MOST 105-2221-E-158-002)

References

1. Zikopoulos, P., Eaton, C. et al: Understanding big data: analytics for enterprise class hadoop and streaming data. McGraw-Hill Osborne Media (2011)
2. Chen, C.L. Philip, Zhang, C.-Y.: Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Sciences 275, (2014) 314–347
3. Wagner, D.: The importance of big data analytics in business. October, World of tech . http://http://www.techradar.com/news/world-of-tech/the-importance-of-big-data-analytics-in-business-1267606/2 (2014)
4. Elgendy, N., Elragal, A.: Big Data Analytics: A Literature Review Paper. Lecture Notes in Computer Science. (2014) 214-227
5. Tee, J.: Handling the four 'V's of big data: volume, velocity, variety, and veracity. TheServerSide.com. (2013)
6. Lukoianova, T., Rubin, Victoria L.: Veracity Roadmap: Is Big Data Objective, Truthful and Credible?. 24th ASIS SIG/CR Classification Research Workshop. (2014)
7. Cai, L. and Zhu, Y.: The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. Data Science Journal. 14: 2. (2015) 1-10
8. Saha, B. and Srivastava, D.: Data quality: The other face of Big Data. in 2014 IEEE 30th International Conference on Data Engineering (ICDE). (2014) 1294–1297
9. Dssouli, R., Serhani, M. A.: Big Data Pre-processing: A Quality Framework. 2015 IEEE International Congress on 2015. (2015) pp.191–198
10. Deshpande, B.: 5 situations which drive data pre-processing before data mining. 2013, http://www.simafore.com/blog/bid/116618/5-situations-which-drive-data-pre-processing-before-data-mining(2013)
11. Fenton, N. E.: Software Metrics - A Rigorous Approach. Chapman & Hall. (1991)

Inferring Smartphone User Demographics from Wi-Fi trace Logs: a Study of Users' Privacy Concerns

Cheng-Ying Hsu[1*], Shang-En Yang[2*], Hung-Yuan Chen[3a],
Fang-Yie Leu[4b], Yao-Chung Fan[5*]

[*]Dept. of Computer Science, National Chung Hsing University, Taichung, Taiwan,
[a]Industrial Technology Research Institute, Hsinchu, Taiwan,
[b]Dept. of Computer Science, Tunghai University, Taichung, Taiwan
ttconch@gmail.com[1], aberfoule@gmail.com[2]
hychen@itri.org.tw[3], leufy@thu.edu.tw[4]
yfan@nchu.edu.tw[5]

Abstract. Over the recent years mobile devices have become a ubiquitous medium supporting various forms of functionality and are widely accepted for commons. However, the privacy threats along with the intimate use of smartphone has become a primary concern. A significant number of methods for persevering privacy against smartphone usage data were also proposed in recent years. The prior research mainly focuses on the privacy leakages by motion sensors, microphones, and GPS trajectories. In this paper, we report another privacy threats by analyzing a collected trace of Wi-Fi signals (referred to as Wi-Fi logs) observed by a smartphone. Such privacy leakage is neglected in the past, as Wi-Fi log data are generally considered to be less sensitive compared with GPS or microphone data. However, in this study, we show that by analyzing the Wi-Fi logs, an adversary can readily reveal many about a smartphone holder, such as occupations, moving patterns, or even user identity. To raise the concerns on this privacy leakage, we design experiments and propose a simple scheme to analyze the Wi-Fi log traces collected by recruited participants. The goal of the scheme is not to design a perfect scheme for discovering user related information but to clearly illustrate the existence and easy identification of privacy-revealing vulnerabilities in Wi-Fi trace logs. The experiment results demonstrate that the privacy can be leaked by Wi-Fi trace logs, which can be readily collected by any app requesting innocence permissions. The experiment results are alarming and may motivate the need to improve the privacy concerns by developing better privacy preserving mechanisms.

Keywords: Wi-Fi trajectories, Privacy, User Demographics

1 Introduction

Over the recent years, mobile devices have become a ubiquitous medium supporting various forms of functionality and are widely accepted for commons.

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_74

With this trend, mobile devices can be viewed in a novel perspective: a mobile device is not just a mini computer for the device holder, but a personal behavior observer providing data around the holder or generated by the holder. As a result, mining data generated from smartphones has drawn significant attentions in recent years [3][9].

On the bright side of mining smartphone data, people gradually understand the privacy threats that come along with the intimate use of the smartphone. For instance, smartphone users all realize that the GPS data is highly sensitive to user identification privacy, as the GPS data reveals where users go, how long they stay, or where they live. A significant number of methods for persevering privacy revealed by GPS data were also proposed in recent years [5][4]. Compared with the privacy sensitive GPS data, the privacy concerns of Wi-Fi logs are often ignored, as they are considered to be less sensitive. However, from our study, we find that Wi-Fi logs are also sensitive to user privacy. We show that by analyzing the Wi-Fi logs, an adversary can readily reveal many about a smartphone holder, such as occupations, preferences, moving patterns, or even user identity. The privacy concerns of Wi-Fi logs are neglected primarily because there is no way for users to realize to what extent the privacy leakage is revealed. Aiming at this issue of unawareness, in this paper, we design experiments by collecting real traces from users and propose a scheme to analyze the collected trace. The experiment results demonstrate the privacy risk of Wi-Fi logs, which can be readily collected by any app requesting some innocence permissions. The results are alarming and clearly show that the privacy can be leaked by Wi-Fi logs without many efforts.

There are two observations for digging user information through Wi-Fi logs collected from a mobile device. First, every Wi-Fi access point is with a Service Set IDentifier (SSID), which is a 32-byte string. The SSID of a Wi-Fi access point is normally a human-readable string and thus commonly referred to as the network name of a Wi-Fi network. The SSID is typically named by the user who sets up the Wi-Fi network. Therefore, SSIDs are often with semantics. For example, the Wi-Fi access point of National Chung Hsing University is named as NCHU-WiFi, from which we can infer the place where the user stayed. Second, a Wi-Fi SSID is produced when the user is near a Wi-Fi access point. A high frequency of a consecutively observed SSID implies a long stay duration at a place. By these two observations, we can use the SSID with semantics to infer the information such as user identification and user preferences. For example, one may infer the occupation of a user from the places the user visited daily, e.g., a graduate student may go to his/her laboratory every weekday.

The contribution of this study is summarized as follows.

- In this paper, we report another privacy threats by analyzing a collected trace of Wi-Fi signals observed by a smartphone. Such privacy leakage is neglected in the past, as Wi-Fi log data are generally considered to be less sensitive compared with GPS or microphone data.
- Second, we propose a scheme for discovering user related information from collected Wi-Fi trace logs. Our goal is not to design a perfect scheme for

discovering user related information from Wi-Fi trace logs, but to clearly illustrate the existence and obvious identification of privacy-revealing vulnerabilities in Wi-Fi trace logs.

- Third, we design experiments by implementing a data collection app and inviting participants to collect real traces for the privacy leakage evaluation. The experiment results based on the real traces demonstrate the privacy leakage of the Wi-Fi trace logs.

The rest of the paper is organized as follows. In Section 2, we review the prior investigation about privacy leakage on smartphones and discuss the difference between our study and the existing works. In Section 3, we introduce the data model and propose a simple scheme to expose the information related to smartphone holder by analyzing Wi-Fi logs. In Section 4, we present the experimental evaluation results and demonstrates the alarming results on the privacy threats. Finally, Section 5 concludes the paper and provides a research roadmap for addressing the privacy threats reported in this study.

2 Related Work

The concern about privacy leaks on smartphones have drawn significant attention in recent years, and new problems are still continuing to be explored. All prior research on this parts mainly focuses on the privacy leakages by motion sensors, microphones, and GPS trajectories of smartphones [8][1][7][5][4]. As sensor-rich smartphones become more ubiquitous, sensory malware has the potential to breach the privacy of individuals. In [8], the authors report their investigation on sensory malware by presenting a stealthy Trojan with innocuous permissions that can sense the context of its audible surroundings to target and extract high-value data of smartphones. In [1], the authors study a new type of side channel attack to infer keystrokes on a smartphone without physical keyboard. While there is no physical keyboard on smartphones, motion on touch screen can be another side channel, as typing on different locations on the screen causes different vibrations, data from motion sensor can be employed to infer the keys being typed. Similar idea by using accelerometers is also reported in [7].

As smartphones are becoming popular and increasingly used in various location-based services, the location privacy of smartphone user has become a concern. In [4], a practical framework for location privacy protection is proposed and implemented for Android smartphone users. The proposed framework address the tracking, profiling, and identification threats while maintaining app functionality. In [5], the authors also propose a privacy-preserving location-based matching scheme as a basic platform primitive for exposing low-level, latitude-longitude (lat-long) coordinates to applications. In literature, we find that the privacy leakage by analyzing Wi-Fi logs remained untouched.

There are also some investigations, such as [2][6], about the privacy leakage issues between the Wi-Fi hotspot and smartphones. In [2], the authors examine the privacy leakage in public Wi-Fi hotspots from activities between hotspot and devices, such as domain name querying, search engines querying, and web browsing. This research reports that many types of user privacy such as identity privacy, financial privacy and etc, can be leaked. In [6], the privacy issues on enabling Wi-Fi fingerprint-based localization (Wi-Fi FBL) is discussed. The Wi-Fi FBL is one of the most promising techniques for indoor localization. The location is estimated by mapping a measured WiFi signal strength over a pre-constructed database owned by the service provider. Providing fine-grained indoor location information therefore becomes a privacy issue. Both research [2][6] focus on the privacy issues between a Wi-Fi side and a client side, which however orthogonal to our focus in this study.

3 The Proposed Scheme

In this section, we propose a scheme to discover information related to a user from the Wi-Fi logs collected by his/her smartphone. Our scheme consists of two components, Information Enrichment and SSID Informativeness Assessment. In Subsection 3.2, we introduce the SSID Informativeness Assessment whose goal is to serve as a data cleansing process to select informative SSIDs for user understanding, and in Subsection 3.3, we introduce the information enrichment component, which aims to enhance the information encoded in SSIDs.

3.1 Trace Attack Model

In this paper, we consider a *trace attack model*, in which an adversary studies on a sequence of Wi-Fi signals observed by a particular user and attempts to discover information about that user. In practical, the sequence of Wi-Fi signals can be readily obtained as part of an app's operation. To simulate such trace attack model, we implement an app performing scanning available Wi-Fi signal at a predefined time interval and send the obtained Wi-Fi observation to a data store server, where raw data from all participants are stored. The experiment data are collected from the smartphone of the participants installed with our app. Please refer to Table 1 as an example of the collected data.

Table 1. A portion of the collected Wi-Fi data of a User:(Time, SSID, BSSID, and Level of the Wi-Fi access point signal)

Time	SSID	Level	BSSID
2013/12/14 02:26:54 PM	TWM WiFi Auto	-72	d8:c7:c8:79:cb:d2
2013/12/14 02:26:54 PM	Jennifer's AP	-72	5c:63:bf:c9:84:9a
2013/12/14 02:26:54 PM	andrew	-74	64:66:b3:4c:6b:80
2013/12/14 02:26:54 PM	SHOYO	-73	74:d0:2b:88:6d:1c
2013/12/14 02:26:54 PM	wenshan	-94	00:13:f7:1b:c8:63
2013/12/14 02:28:44 PM	Andyhome	-90	f8:d1:11:75:54:5a
2013/12/14 02:28:44 PM	MAOWLAN	-81	00:18:e7:cb:6a:6c
2013/12/14 02:28:44 PM	unilevel	-94	90:f6:52:3a:e8:a4
2013/12/14 02:28:44 PM	Simon	-94	20:cf:30:87:dd:3b
2013/12/14 02:28:44 PM	7-11 WiFi	-95	90:f6:52:45:0c:24
2013/12/14 02:28:44 PM	Pomelo	-96	0c:82:68:34:90:22
2013/12/14 02:28:44 PM	TINASONIC	-79	00:1f:c6:27:e9:ce
2013/12/14 02:28:44 PM	MuJaHomeAP	-87	00:24:a5:34:0f:86
2013/12/14 02:28:44 PM	Starbuck-Wif	-93	78:54:2e:2f:3e:d0

3.2 SSID Informativeness Assessment

One observation for discovering user related information is that not every SSIDs are informative; some are without any semantics, and some with semantic but are not related to user information. One example is that lots of Wi-Fi access points are named by meaningless characters, such as "ZZZZ", "888", and "CHT36678", from which nothing can be inferred. Yet another example is that some SSID is named with a default SSID setting, which is a name given by equipment manufacturers, such as ZyXel, and Dlink, or a name set by hotspot infrastructure providers, such CHT and iTaiwan. For the SSID types without useful semantics, such as device default name, the only thing we can do is to eliminate them from the given SSID set, as nothing can be derived from them. Therefore, our idea is to manually enumerate SSID names obviously without useful semantics and then filter them out during the profile construction process. To this end, we manually select highly observed SSIDs that are obviously without useful information. The SSID names, such as ZyXel, DLink, and Hotspot, are examples to be included in the list.

However, one point to mention is that it is impossible for the list-based approach to be effective, as there are too many to list them all. For the purpose of judging the informativeness of an SSID, we observe the following clues. First, SSIDs are named by humans and often show linguistic features on the given strings. We observe that SSID strings often contain delimiter characters, e.g., hyphen, whitespace, and underline. The delimiters can be used to chop an SSID string into tokens. The basic idea is that an SSID with many tokens is often informative to the places where the Wi-Fi access point is installed, e.g.,"nchu-cs-udic-lab" and "Starbuck Cafe Free Wifi." In addition, some other features can be also employed. For example, if an SSID is with all upper-case letters, it's likely that the SSID is an abbreviation of something, such as a place or an affiliation, e.g. NCHU. Another clue is that if an SSID is with many digits and letters interleaved, the SSID is probably meaningless. Therefore, to leverage these characteristics, the idea is to make use of a supervised classifier on the basis of a training data set to infer if a given SSID is informative.

Therefore, for a given SSID, we compute the following features for the SSID: (1) the number of tokens, (2) the average token length, (3) the number of delimiters, (4) the number of digits, (5) the number of upper-case letters, and (6) the number of lower-case letters. As an example, for an SSID "nchu-UDIC_Lab fan23" we can extract the features from the SSID into the following form: [4, 4, 2, 2, 5, 9]. With the features, off-the-shelf classifiers can be employed to judge if the information encoded in an SSID is informative.

3.3 SSID Information Enrichment

An SSID is typically a short string without too much information. An idea to this problem is that we make use of the web search service API, such as the Google Web Search API, to enrich the information encoded in the Wi-Fi SSID. With the help of Google Web Search API, we can readily expand the meaning of a given SSID. For example, if the SSID ”nchu” is input into the API, we can obtain web documents regarding National Chung Hsing University, and if the SSID ”ITRI” is emitted, we obtain some descriptions about Industrial Technology Research Institute located at HsinChu, Taiwan. Therefore, with the employment of the web search API, an SSID string can be expanded into a set of informative documents. In Figure 1, we show an example of expanding an SSID ”Nchu-cs” to a set of terms related to the abbreviation. In the example, we observe that the terms, e.g., ”department of computer science”, ”the national Chung Hsing university”, and ”national university”, are returned, which are all about the Nchu-cs abbreviation.

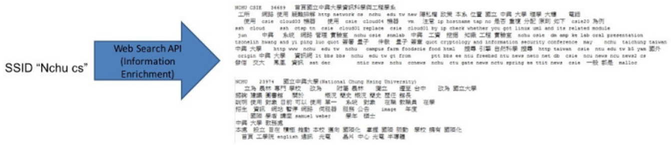


Fig. 1. SSID Information Enrichment

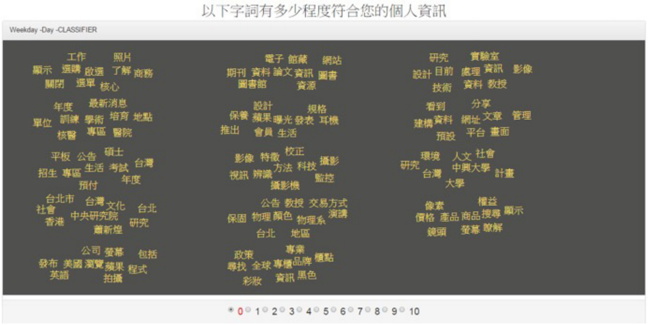
3.4 The Overview of Proposed Scheme

Given a sequence of Wi-Fi SSID logs from a smartphone, the trace attack model proceeds as follow. First, the SSIDs are sorted by the frequencies, as a high frequency of the observed SSID implies a long stay duration at a place and should be more meaningful to the targeted user. Then, SSIDs are one-by-one verified by SSID informativeness assessment process (IAP) according to their frequency from high to low. If an SSID is pass through the SSID IAP, then the SSID enters into the SSID Information Enrichment process. In this study, we make use of the Google Web Search API to expand the information encoded behind the SSID. When an SSID is input into the Google Web Search API, the API will return a set of documents. We then process the returned web documents by tokenizing words and removing stop words. After the web document processing, the resultant tokens are accumulated and served as the user profile.

4 Evaluation

4.1 Experiment Setting

In our study, four participants are recruited by giving the voucher in exchange for contributing all the smartphone usage data in two years through installing our app in their smartphones. The app performs scanning available Wi-Fi signals at an interval of 30 seconds and sends the obtained Wi-Fi observations (Time, SSID, BSSID, and Level of Wi-Fi signal) to a data store server, where raw data from all participants are stored. The Wi-Fi data collection starts from 20 August 2016 to 20 April 2016. We use the collected raw Wi-Fi data trace as data sets, from which a user profile (a set of terms) is generated by using the scheme proposed in Section 3 for each participant. We employ a word cloud visualization form as illustrated in Figure 2 to render the terms discovered by the proposed scheme. In addition, we ask the participants to give a score range from 0 to 10 to judge what extent the reported terms are related to the users.



以下字詞有多少程度符合您的個人資訊

● 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10

Fig. 2. Privacy Leakage Survey Form

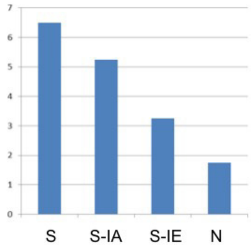


Fig. 3. Experiment Results

4.2 Experiment Results

In Figure 3, we show the experiment results, where the y-axis is the averaged score from the participants, and the x-axis are with the proposed scheme and its three variants. In the figure, we denote the proposed scheme by S, the scheme without information assessment component by $S-IA$, the scheme without information enrichment component by $S-IE$, and the scheme denoted by N simply show the collected highly frequently observed SSID strings without any process by information assessment and information enrichment components.

There are two primary observations. First, from the experiment results, we observe that the averaged score of the proposed scheme given by the participants is 6.4, which indeed raise the privacy leakage concerns for participants by applying a simple data analysis scheme over the collected Wi-Fi traces. Second, the experiment results also show that some data cleaning and information enrichment techniques are required. One can observe that the scores of the variants of our proposed scheme are all lower than the proposed scheme.

5 Conclusion

In this study, we have made an attempt to characterize the privacy leakage of Wi-Fi signals observed by a smartphone. We collected and analyzed Wi-Fi trace logs from four smartphone users. Some privacy sensitive information, e.g. occupation, are shown to be able to be deduced from Wi-Fi logs collected by a smartphone. The results are indeed alarming and clearly show that the privacy can be leaked by Wi-Fi logs without making many efforts. Our future work is to develop techniques for avoiding such privacy leakage.

Acknowledgment

This work was supported by Information and Communications Research Laboratories (ICL), Industrial Technology Research Institute (ITRI), Taiwan, under Grant 105-EC-17-A-24-0691 and Ministry of Science and Technology, Taiwan, under grant number 103-2221-E-005-051-MY2 and MOST 104-2218-E-001-002.

References

1. Liang Cai and Hao Chen. Touchlogger: Inferring keystrokes on touch screen from smartphone motion. *HotSec*, 11:9–9, 2011.
2. Ningning Cheng, Xinlei Oscar Wang, Wei Cheng, Prasant Mohapatra, and Aruna Seneviratne. Characterizing privacy leakage of public wifi networks for users on travel. In *Proceedings of the IEEE INFOCOM 2013, Turin, Italy, April 14-19, 2013*, pages 2769–2777. IEEE, 2013.
3. Nathan Eagle and Alex Sandy Pentland. Reality mining: sensing complex social systems. *Personal and ubiquitous computing*, 10(4):255–268, 2006.
4. Kassem Fawaz and Kang G Shin. Location privacy protection for smartphone users. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pages 239–250. ACM, 2014.
5. Saikat Guha, Mudit Jain, and Venkata N Padmanabhan. Koi: A location-privacy platform for smartphone apps. In *Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation*, pages 14–14. USENIX Association, 2012.
6. Hong Li, Limin Sun, Haojin Zhu, Xiang Lu, and Xiuzhen Cheng. Achieving privacy preservation in wifi fingerprint-based localization. In *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*, pages 2337–2345. IEEE, 2014.
7. Emmanuel Owusu, Jun Han, Sauvik Das, Adrian Perrig, and Joy Zhang. Accessory: password inference using accelerometers on smartphones. In *Proceedings of the Twelfth Workshop on Mobile Computing Systems & Applications*, page 9. ACM, 2012.
8. Roman Schlegel, Kehuan Zhang, Xiao-yong Zhou, Mehool Intwala, Apu Kapadia, and XiaoFeng Wang. Soundcomber: A stealthy and context-aware sound trojan for smartphones. In *NDSS*, volume 11, pages 17–33, 2011.
9. Yu Zheng, Xing Xie, and Wei-Ying Ma. Geolife: A collaborative social networking service among user, location and trajectory. *IEEE Data Eng. Bull.*, 33(2):32–39, 2010.

Enhancing Security of LTE using a Double Masking Technique

Jung-Chun Liu, Yi-Li Huang, Fang-Yie Leu
Department of Computer Science, Tunghai University, Taichung, Taiwan
{jcliu, yifung, leufy}@thu.edu.tw;

Abstract. LTE uses the Evolved Packet System Authentication and Key Agreement (EPS-AKA) procedure to establish and verify keys. However, the EPS-AKA is vulnerable to attacks such as disclosure of the user identity, man-in-the-middle attack and denial of services; therefore, a robust authentication mechanism is required. In this paper, we enhance security of LTE by using a double masking technique, in which both the identity key of the user equipment (UE), i.e., *IMSI*, and the random challenge key, i.e., *RAND*, are masked without being exposed in the authentication process. The proposed double masking technique is effective in performing mutual authentication of the user and the network. Security analysis shows that this technique is more secure than the original EPS-AKA since *IMSI* and *RAND* are well-protected and achieve practical security. Since all operators used in encrypting keys are simple and efficient, it works without degrading the performance of the existing LTE system.

1 Introduction

LTE utilizes the Evolved Packet System Authentication and Key Agreement (EPS-AKA) procedure to establish and verify keys [1-2]. The EPS-AKA consists of three stages: initiation, transfer of credentials, and challenge-response exchange. During the initiation stage, the user equipment (UE) provides the network with its identity, i.e., International Mobile Subscriber Identity (*IMSI*); based on this identity, the network initiates the authentication procedure. Three nodes involve in the authentication procedure: the UE, Mobility Management Entity (MME) and Home Subscriber Server (HSS). The HSS holds subscriber information; it can verify an authentication request from the UE as well as generate authentication data for the MME.

Mutual authentication is the process in which a network and a user authenticate each other. The EPS AKA authentication procedure consists of two stages. First, the HSS generates Evolved Packet System (EPS) authentication vectors, i.e., *AVs*= (*RAND*, *AUTN*, *XRES*, K_{ASME}), and delivers them to the MME. Second, the MME selects one of the *AVs* and uses it for mutual authentication with the UE, and thus UE and MME each other shares the same authentication key (K_{ASME}).

The designers of 4G wireless security should face following challenges: (i) the security concerns of accessing the Internet from a fixed location as well as the added

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_75

needs for flexibility and mobility; (ii) the impact on the performance and traffic handling capacity of the service provider's network due to additional cryptographic methods and security mechanisms applied to IP networks; (iii) protection of all new emerging 4G devices and applications from a growing variety of security threats [2,3].

In the traditional EPS-AKA protocol design, at the attach stage, the UE sends the UE ID, i.e., *IMSI*, to the eNB in plaintext. Hence, malicious attackers can decode the *IMSI* and copy it for illegal purposes. Besides, at the authentication request stage, the random challenge key *RAND* is exposed in the air when the MME sends *AVs* to the UE. To solve above two issues, in this paper, we propose an efficient augmented EPS-AKS protocol with a double masking technique, in which both *IMSI* and *RAND* are encrypted before being transmitted. By adding preprocess and customization procedures at the UE and HSS, the proposed protocol can be incorporated into existing LTE networks; moreover, the performance of the service provider's network suffers less impact from this cryptographic method since the operators used in encrypting keys are elementary and very efficient.

The rest of this paper is organized as follows. Section 2 introduces related works of this study. Section 3 presents the proposed authentication protocol. Security analysis is performed in Section 4. Section 5 concludes this paper and outlines our future studies.

2 Related Studies

4G networks offer an open environment where different wireless technologies and service providers share an IP-based core network to provide uninterrupted services to subscribers. Due to the fact that 4G is an open, heterogeneous, and IP-based environment, it suffers from new security threats as well as inherent ones. The key vulnerabilities include access control, communication security, data confidentiality, availability and privacy [4,5].

The EPS-AKA is vulnerable to attacks such as disclosure of the user identity, man-in-the-middle attack and denial of services (DoSs) [3]. To solve these problems, Abdrabou et al. [6] proposed a pre-authentication procedure based on Simple Password Exponential Key Exchange (SPEKE) and symmetric key cryptography to generate a dynamic key every time when users access to the network. However, this approach gives rise to higher communication overheads than the current EPS-AKA protocol does.

Rogue base-station attack can compromise user privacy by tracking their geographical movements, and intercept sensitive personal data, such as credit card information. Several existing studies are trying to remedy this situation. For example, Mazroa and Arozullah [7] proposed a rogue base-station identification protocol to protect UE privacy; a cloud server is used to maintain the locations of real base-stations, and the UE can ensure that it only sends sensitive information through legal base stations by contacting the cloud server first.

To avoid DoS attacks [3], Apostol and Răcuciu [8] proposed an EPS-AKA with a Security Vector approach, in which before computing the authentication vectors, the HSS first checks to see whether it has already received an authentication requests from the same hardware address of the UE. This approach can prevent a device from trying to sustain DoS attacks to overload the HSS/MME.

In this study, we used two efficient elementary operators to encrypt/decrypts keys, including exclusive-OR ($\oplus$) [9] and binary adder ($+_2$) [10]. We also used a dynamically accumulated shifting substitution (DASS for short) algorithm, which is a one-way function to encrypt a plaintext into an irreversible ciphertext introduced as follows.

Algorithm 1: DASS

```
Input: plaintext block $P$ of $n$ bits in length, and a 16 x 16 random table-box (RT-Box
        for short), where $n$ is a number which is multiple of 8.
Output: ciphertext block $C$.
{let $P = p_1 \parallel p_2 \parallel \cdots \parallel p_k$ and $C = c_1 \parallel c_2 \parallel \cdots \parallel c_k$, where $k = n/8$;
$ct = 0$;
for $i = 1$ to $k$   {$vp[i]$ = Int($pi$);   $ct = ct + vp[i]$;}
for $i = 1$ to $k$   {$ct = ct + vp[i]$;   $ch$ = str(($vp[i] + ct$) mod 256);
                      $c_i$ = the corresponding content in the RT-Box after $ch$ is substituted by
                          looking up RT-Box following the SubBytes step in Advanced
                          Encryption Standard (AES);
                          /*for example, if $ch = 65_{hex}$, then $c_i$ is found by looking up the
                          entry located at row 6 and column 5 in the RT-Box.*/}
```

3 Proposed Authentication Protocol

In this section, we propose a secure and efficient authentication and key agreement protocol for LTE networks.

On the UE side, there are parameters (shared with the HSS) stored in the Universal Subscriber Identity Module (USIM), including *IMSI* (usually a 15-digit number, of which the last 9 or 10 digits form the mobile subscription identification number, *MSIN*), subscriber authentication key K (128 bits), and authentication and key management field *AMF* (16 bits). Also, a global random box (GR-Box) and a unique user individual key (*UIK*) are used in our protocol.

On the HSS side, a GR-Box and a user individual keys table (UIKS-TABLE), which consists of *IMSI* and corresponding *UIK*, are used for customization of the HSS.

Similar to the original EPS-AKA protocol, our authentication and key agreement protocol comprises three stages: attach request, authentication request, and authentication response. In addition, it includes UE preprocess and HSS preprocess to establish keys described as follows.

A. UE Preprocess: it is called by UE at step 1 of the UE attach request stage to establish keys.

(1) The UE generates a random key $Rand_U$;

(2) $URAND$ = DASS (GR-Box, $Rand_U$);

(3) $URAND = UR_1 \parallel UR_2 \parallel UR_3$; /* key division */

(4) $CMSIN = (MSIN \oplus UR_2) +_2 UR_3$;

$CIMSI = MCC \parallel MNC \parallel CMISN$;

(Note that $IMSI = MCC \parallel MNC \parallel MISN$);

(5) $UIRK_1 = URAND \oplus UIK$;
$UIRK_2$= DASS (GR-Box, $UIRK_1$);
$UIRK_3 = UIRK_2 +_2 URAND$;
(6) Return;

B. HSS Preprocess: it is called by HSS at step 1 of the HSS attach request stage to establish keys.
(1) The HSS Retrieves $Rand_U$ and *CIMSI* from message 2 and retrieves *MCC*, *MNC* and *CMSIN* from *CIMSI*;
(2) $URAND$ = DASS (GR-Box, $Rand_U$);
(3) $URAND = UR_1 \parallel UR_2 \parallel UR_3$;
(4) $MSIN = (CMSIN -_2 UR_3) \oplus UR_2$;
$IMSI = MCC \parallel MNC \parallel MSIN$;
(5) According to *IMSI*, a corresponding *UIK* is retrieved from the UIKS-TABLE;
(6) $UIRK_1 = URAND \oplus UIK$;
$UIRK_2$ = DASS (GR-Box, $UIRK_1$);
$UIRK_3 = UIRK_2 +_2 URAND$;
(7) Return;

3.1 Attach Request Stage

In this stage, UE, MME and HSS are involved.

A. UE: Attach Request-1
When a mobile device is turned on, the UE does the following steps.
(1) The UE calls UE Preprocess to obtain $Rand_U$, *URAND*, and *CIMSI*, and then $UIRK_1$, $UIRK_2$, and $UIRK_3$;
(2) The UE sends message 1, which is an attach request, to eNB/MME, i.e.,

$$\text{UE} \xrightarrow{\text{message 1}} \text{eNB/MME}$$

where message 1 = $CIMSI/Rand_U/$UE Network Capability$/KSI_{ASME=8}$.

B. MME: Attach Request-2
When receiving message 1, MME sends message 2 to the HSS, i.e.,

$$\text{MME} \xrightarrow{\text{message 2}} \text{HSS}$$

where if $KSI_{ASME} = 8$, then message 2 = $CIMSI/Rand_U/SN\ ID/$Network type$/KSI_{ASME=8}$.
else message 2 = $IMSI/SN\ ID/$Network type

C. HSS: Attach Request-3
On receiving message 2, the HSS does the following steps.
(1) Retrieves KSI_{ASME} from message 2; if $KSI_{ASME} = 8$, then calls HSS Preprocess to obtain *IMSI*, and then $UIRK_1$, $UIRK_2$, and $UIRK_3$, otherwise retrieves *IMSI* from message 2.
(2) Retrieves LTE *K* based on *IMSI*;
(3) Generates a random number *RAND* and sequence number *SQN*;

(4) Generates $AUTN_{HSS}$, *XRES*, and K_{ASME} by employing the EPS-AKA algorithm where $AUTN_{HSS} = (SQN \oplus AK \parallel AMF \parallel MAC)$;
(5) Generates the authentication vector $AV = RAND \| XRES \| \ K_{ASME} \parallel AUTN_{HSS}$;
(6) If $KSI_{ASME} = 8$, then $CRAND=(RAND \oplus UIRK_2)+_2 UIRK_3$;
(7) Sends message 3 to the MME, i.e., HSS $\xrightarrow{\text{message 3}}$ MME
where if $KSI_{ASME} = 8$,
message 3 = $(AV/KSI_{ASME=8}/RandU/UIRK_1/UIRK_2/UIRK_3/CARND)$;
else message 3 = $(AV/KSI_{ASME} = 7)$;

3.2 Authentication Request Stage

In this stage, MME and UE are involved.

A. MME: Authentication Request-1
When receiving message 3, the MME does the following steps:
(1) Retrieves KSI_{ASME} from message 3;
(2) Transmits message 4 to the UE through eNB, i.e., MME/eNB $\xrightarrow{\text{message 4}}$ UE
where if KSI_{ASME} = 7, message 4 = $(RAND/AUTN_{HSS}/KSI_{ASME} = 1)$;
else message 4 = $(CRAND/AUTN_{HSS}/KSI_{ASME} = 9)$;

B. UE: Authentication Request-2
On receiving message 4, the UE does the following steps:
(1) Retrieves KSI_{ASME} from message 4;
(2) If KSI_{ASME} = 9, then $RAND = (CRAND -_2 UIRK_3) \oplus UIRK_2$;
(3) Computes *AK*, *SQN*, *XMAC*, *RES*, *CK*, *IK* and K_{ASME} by employing the EPS-AKA algorithm on the USIM side;
(4) Retrieves *MAC* from $AUTH_{HSS}$, and verifies whether *XMAC* = *MAC* or not. If not, discards the receiving message and waits for a correct one, otherwise continues with next step;
(5) Transmits the authentication response message *RES* to the MME/eNB;

3.3 Authentication Response Stage

MME: Authentication Response
On receiving the authentication response message *RES*, the MME verifies whether *XRES* = *RES* or not. If not, it discards the receiving message and waits for a correct one. Otherwise, the authentication process is completed.

4 Security Analysis

In this study, parameter *MSIN*, which is a part of UE ID or *IMSI*, has been encrypted as *CMSIN*, and parameter *RAND* has been encrypted as *CRAND* before wirelessly transmitted. Hence, at each communication session, these two most important security parameters, i.e., *IMSI* and *RAND*, which are exposed in the original EPS-AKA protocol, are masked and well-protected; therefore, the overall security of the LTE communication networks is greatly enhanced, and the reasons are chiefly as follows.

(1) In the LTE authentication phase, the major security mechanism is the EPS-AKA protocol, in which the most input parameters are LTE *K* and *RAND*. However, in the current LTE communication mechanism, *RAND* is exposed when wirelessly transmitted by the eNB to the UE; hence *RAND* may be intercepted by attackers, causing security threats on the EPS-AKA protocol. In this study, *RAND* has been encrypted as *CRAND* before wirelessly transmitted, so the attacker needs to decrypt the intercepted *CRAND* to recover *RAND*, and then attacks the EPS-AKA security mechanism. But, in this study, both LTE *K* and *RAND* parameters are unknown to the attacker, which greatly increases the difficulty of solving the EPS-AKA.

(2) The *IMSI*, as the identity of the *USIM* for a mobile device, is exposed when wirelessly transmitted by the UE to the eNB right after the mobile device is turned on. The attacker can intercept it and continue retrieving and recording the following communication messages. The attacker can eavesdrop and stealthily collect data of a specific *IMSI* for a long period of time. After collecting a considerable number of communication messages, he/she can try to decrypt *IMSI*. However, in this study, the *MSIN* of *IMSI* has been encrypted as *CMSIN* before being transmitted from the UE to the eNB. Since *CMSIN* itself is a random number, i.e., different *CMSIN* is generated each time when the mobile device is turned on, the attacker has no way to obtain the correct *MSIN* via the information accompanied by *CMSIN*; in other words, the attacker cannot know the identity with which to identify who is making phone calls, and has lesser opportunity to eavesdrop and stealthily collect data of a specific *IMSI*, implying that the chance to decrypt *IMSI* is also smaller.

Based on above reasons, the security of *CIMSI* and *CRAND* will determine the pros and cons of the proposed approach. In the following lemmas, we prove that *CIMSI* and *CRAND* are highly secured.

Lemma 1:

In the study, $Rand_U$/*URAND* is random/induced random key with 128 bits in length. Let *MSIN* be a decimal digit parameter with 40 bits long. Then the probability p with which to recover the value of *IMSI* from an illegally intercepted *CIMSI* on one trial is $p = \frac{1}{10^{10}}$.

Proof:

For *IMSI* = *MCC*//*MNC*//*MSIN* and *CIMSI* = *MCC*//*MNC*//*CMSIN*, the probability p with which to recover the value of *IMSI* from *CIMSI* is the same as the probability p of recovering the value of *MSIN* from *CMSIN*.

$\because CMSIN = (MSIN \oplus UR_2) +_2 UR_3$;

$URAND = UR_1 // UR_2 // UR_3$; and

$URAND =$ DASS (GR-Box, $Rand_U$);

Even if $Rand_U$ is known by the attacker, however, since the GR-Box is inside the user mobile phone, the attacker does not know it. The attacker cannot generate *URAND* without knowing the GR-Box, i.e., the attacker cannot know *URAND* from the known

$Rand_U$, implying that *UR_2* and *UR_3* are unknown to the attacker. Under this situation, the probability *p* with which to recover the value of *MSIN* from *CMSIN* is the same as that of a blind guess [10, 11]. Also, there are 10^{10} possible values for *MSIN*, ranging from 0 to 9,999,999,999.

Hence the probability *p* with which to recover the value of *IMSI* from *CIMSI* on one trial is also the probability *p* with which to recover the value of MSIN from CMSIN on one trial, and $p = \frac{1}{10^{10}}$. (Q.E.D.)

Lemma 2:

In the study, *RAND*, *CRAND* and the related keys, such as *URAND* and *UIK*, are 128 bits in length. The probability *p* with which to recover *RAND* from an illegally intercepted *CRAND* on one trial is $p = \frac{1}{2^{128}}$.

Proof:

For $CRAND = (RAND \oplus UIRK_2) +_2 UIRK_3$;

and $RAND = (CRAND -_2 UIRK_3) \oplus UIRK_2$;

then to obtain *RAND* from known *CRAND*, both *$UIRK_2$* and *$UIRK_3$* are indeed required. However,

$UIRK_3 = UIRK_2 +_2 URAND$;

$UIRK_2$ = DASS(GR-Box, $UIRK_3$); and

$UIRK_1 = URAND \oplus UIK$;

These equations show that only when *UIK* and GR-Box are known, then *$UIRK_2$* and *$UIRK_3$* can be derived, but *UIK* and GR-Box are inside the user mobile device, and attackers cannot know them, indicating that the attackers cannot know the induced keys *$UIRK_2$* and *$UIRK_3$*. Hence, under this situation, the probability *p* with which to recover the value of *RAND* from *CRAND* is the same as that of a blind guess [10,11]. Therefore, the probability *p* of recovering the value of *RAND* from *CRAND* on one trial is $p = \frac{1}{2^{128}}$. (Q.E.D.)

Lemma 1 and Lemma 2 show that *IMSI* and *RAND* are well-protected and achieve practical security, implying that our proposed approach is more secure than the original LTE security mechanism.

5 Conclusions and Future Studies

Since the LTE represents an open, heterogeneous, and IP-based environment [5, 12], it faces many security threats, such as disclosure of the user identity, man-in-the-middle, and DoS attacks. We propose a mechanism to enhance security of LTE by using a double masking technique, and prove that both important parameters, *IMSI* and *RAND*, which are exposed when wirelessly transmitted in the current EPS-AKA protocol, are masked from attackers. Besides, the proposed approach will not impact performance of the current LTE system, since it uses efficient elementary operators for encryption. In the future, we will study the feasibility of hardware and software deployment of our techniques over the existing LTE system. We will also like to derive the reliability and behavior models for the proposed system. These constitute our future studies.

References

1. M. Purkhiabani and A. Salahi, "Enhanced Authentication and Key Agreement Procedure of next Generation 3GPP Mobile Networks," *International Journal of Information and Electronics Engineering*, vol. 2, no. 1, January 2012, pp. 69-77.
2. N. Seddigh, B. Nandy, R. Makkar, and J.F. Beaumont, "Security Advances and Challenges in 4G Wireless Networks," *Annual International Conference on Privacy, Security and Trust*, 2010, pp. 62-71.
3. B. Rashidi and C. Fung, "A Survey of Android Security Threats and Defenses," *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, vol. 6, no. 3, September 2015, pp. 3-35.
4. M. Aiash, G.E. Mapp, A. Lasebae and R. Phan, "Providing Security in 4G Systems: Unveiling the Challenges," Advanced International Conference on Telecommunications, 2010, pp. 439-444.
5. A. Skovoroda and D. Gamayunov, "Securing mobile devices: malware mitigation methods," *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, vol. 6, no. 2, June 2015, pp. 78-97.
6. M.A. Abdrabou, A.D.E. Elbayoumy, E.A. El-Wanis, "LTE Authentication Protocol (EPS-AKA) Weaknesses Solution," *IEEE International Conference on Intelligent Computing and Information Systems*, 2015, pp. 434-441.
7. A.A. Mazroa, M. Arozullah, "Detection and Remediation of Attack by Fake Base Stations in LTE Networks," *International Journal of Soft Computing and Engineering* vol. 5, issue-2, May 2015, pp. 12-15.
8. C.G. Apostol and C. Răcuciu, "Improving LTE EPS-AKA using the Security Request Vector," *7th Edition Electronics, Computers and Artificial Intelligence*, June 2015, pp. 185-188.
9. T. kurokawa, R. Nojima, and S. Moriai, "On the security of CBC Mode in SSL3.0 and TLS1.0," *Journal of Internet Services and Information Security*, vol. 6, issue 1, February 2016, pp. 2-19.
10. Y.L. Huang, C.R. Dai, and F.Y. Leu, and I. You, "A Secure Data Encryption Method Employing a Sequential-Logic Style Mechanism for a Cloud System," *International Journal of Web and Grid Services*, vol. 11, no. 1, January, 2015, pp. 102-124.
11. Y.L. Huang, F.Y. Leu and K.C. Wei, "A Secure Communication over Wireless Environments by using a Data Connection Core," *Mathematical and Computer Modelling*, vol. 58, no. 5-6, September 2013, pp. 1459-1474.
12. A. Kitana, I. Traore, and I. Woungang, "Impact Study of a Mobile Botnet over LTE Networks," *Journal of Internet Services and Information Security*, vol. 6, issue 2, May 2016, pp. 1-22.

Part VI
The 5-th International Workshop on Robot Interaction, Control, Communication and Cooperation (RI3C-2016)

The effective flock control by two sheepdogs

Haruka Watanabe and Kaoru Fujioka

Abstract This paper concerns with the shepherding problem in which a small number of sheepdogs guide a large number of sheep to the target position. A research of a swarm control by external force with a small number of leader is less popular than modeling and simulation of swarm behavior. If we control a flock by comparatively simple systems, we expect to apply the technique such as constructing robots which prevent the spread of oil in sea or prevent the forest fires. In the previous research, it has been shown that sheepdogs use two rules to herd sheep. We come up with a new method with two sheepdogs which share the two rules respectively. Our method is compared with previously proposed methods using computer simulations.

1 Introduction

Many researches have investigated the flock of birds, fish, and insects to clarify the flock behavior [7]. Boids model proposed by Craig Raynolds in 1987 [4] is known as a well-known example of crowd study consisting of three simple rules: Separation, Alignment, and Cohesion. The those rules are referred to our model for simulating the flock of sheep.

On the other hand, researches to control flocks are still less popular, but Lien studied technical methods to control flocks [2], [3]. The shepherding problem, to guide a herd consisting of a large number of sheep by a small number of sheepdogs, is useful for the application of robotics and crowd control if we think of a sheepdog as a controller and a sheep as a target to be controlled. In practice, actual sheepdogs

Haruka Watanabe
International College of Arts and Sciences, Fukuoka Women's University, 1-1-1 Kasumigaoka, Higashi-ku, Fukuoka 813-8529, Japan, e-mail: 13ue076@mb2.fwu.ac.jp

Kaoru Fujioka
International College of Arts and Sciences, Fukuoka Women's University, 1-1-1 Kasumigaoka, Higashi-ku, Fukuoka 813-8529, Japan, e-mail: kaoru@fwu.ac.jp

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_76

have been kept for guiding a herd for a long time and must have some tips to control a flock.

In [6], it has been shown that a sheepdog switches its motion between *driving mode* and *collecting mode* by the data collected from actual sheep and sheepdogs. Inspired by the paper, we construct a sheepdog model using multi-agent systems in which two sheepdog agents control a flock of sheep agents. One sheepdog agent dedicates itself to the driving mode and the other to the collecting mode, independently. Actually, some breeds of dogs are skillful at driving and some breeds of dogs are skillful at collecting, which is consistent to our model. By comparing our model with two sheepdog agents and the previous model with one sheepdog agent, we consider the effects of two sheepdogs by computer simulations.

2 Description of the sheepdog model

First, we explain the basic configuration of our model. On a square field of size 250 × 250 meters, N_s sheep agents are randomly placed in the central square of size (50×50), called start area (Fig 1).

In this paper, the number of sheep agents N_s are set either 50 or 100. N_d sheepdog agents begin to guide the sheep agents from the position $(125, 0)$ toward the target position $(125, 250)$. We set the number of sheepdog agents N_d as either 1 or 2. In case of multiple sheepdog agents, there is a possibility that the two sheepdog agents locate at the same position, actually the sheepdog agents locate at the same start position.

Each trial is regarded as successful if the global center of mass (GCM), represented with the x-mark in Figure 1, is within 5 meters of the target position in 1000 steps. Otherwise, the trial is treated as unsuccessful.

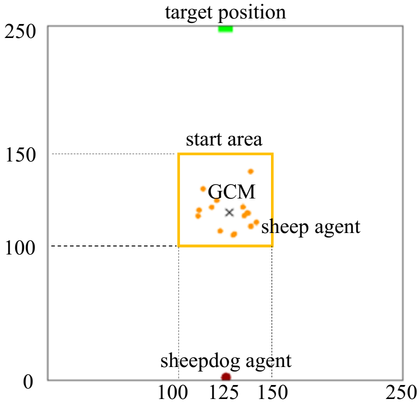


Fig. 1 Field of size 250 × 250 at the initial step.

Next, we describe the behavior of sheep agents which is defined based on the previous study [6]. Each sheep agent decides its next position by the following forces:

- Cohesive power.
  A sheep agent heads for the local center of mass (LCM) of the nearest n $(1 \leq n < N_s)$ sheep agents.
- Repulsion force.
  To prevent collisions between sheep agents, a repulsion force occurs if another sheep agent exists within the repulsion distance $d_r = 2$.

- Force to escape from a sheepdog agent.
  A sheep agent keeps away from the nearest sheepdog agent if the sheepdog agent is within the detection distance $d_{detect} = 65$.
- Inertial force.
  A sheep agent remains in the previous position by the inertial force.
- Noise.
  A sheep agent heads for in random directions due to noise.

As for sheepdog agent, in the previous researches [6] [5] [1], *GCM-targeting control* with one sheepdog agent was provided, in which the sheepdog agent selects *collecting mode* or *driving mode* alternatively according to the scatter of the flock. When the flock sticks together, the sheepdog agent drives the flock toward the target position, otherwise, the agent heads for the furthest sheep agent to collect the agent until all the sheep agents are cohesive.

In this paper, we consider the control with two sheepdog agents based on GCM-targeting control. Each shepherd has its own role for a guide and is called *driving sheepdog* or *collecting sheepdog* from its role, for simplicity. The roles of the two sheepdog agents are the followings:

- Driving sheepdog agent. (Fig 2)
  To drive the herd to the target position, the agent heads for the driving position $\mathrm{P}_{\mathrm{drive}}$ which is located $d_{\mathrm{drive}} = d_r\sqrt{N_s}$ meters behind the flock.
- Collecting sheepdog agent. (Fig 3)
  To collect the furthest sheep agent from the GCM, the collecting sheepdog agent heads for the collecting position $\mathrm{P}_{collect}$ which is located $d_{\mathrm{collect}} = d_r$ meters behind the target agent.

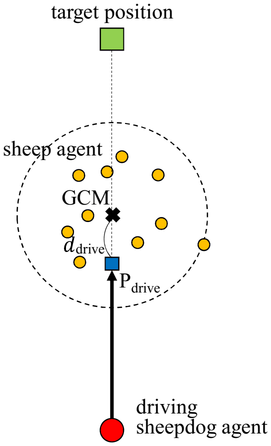


Fig. 2 Driving sheepdog agent and the driving position P_{drive}.

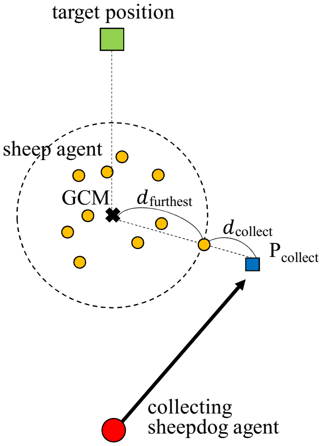


Fig. 3 Collecting sheepdog agent and the collecting position $\mathrm{P}_{collect}$.

3 Results

We consider the rate of success, the total number of steps required, and the dispersion when we change the nearest number of sheep agents n in order to compare the control with one sheepdog agent and the control with two sheepdog agents. Each plot in the following six line graphs is illustrated by 100 trials. The red line (resp. blue line) indicates the control with one sheepdog agent (resp. two sheepdog agents), respectively. In the graphs of the total number of steps for the successful experiments (Fig 6, Fig 7), the corresponding plot is not displayed if all the 100 trials are unsuccessful. The dispersion (Fig 8, Fig 9) implies the sum of the squares of the distance between the GCM of all the sheep agents and each sheep agent at the end of the simulation.

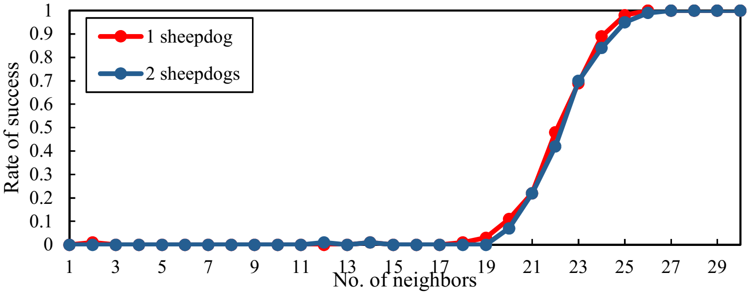


Fig. 4 Rate of success with $N_s = 50$ and $1 \le n \le 30$.

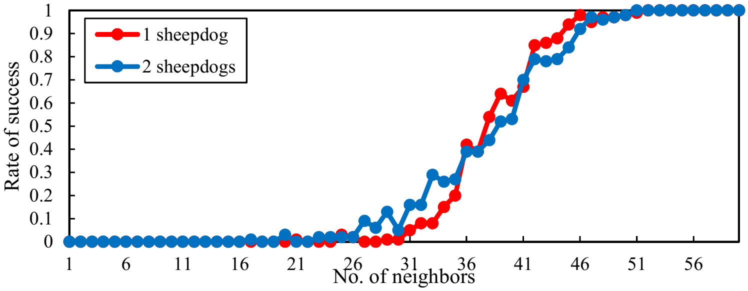


Fig. 5 Rate of success with $N_s = 100$ and $1 \le n \le 60$.

Figure 4 (resp. Figure 5) shows the rate of success for the 100 trials with the number of sheep agents $N_s = 50$ (resp. $N_s = 100$) and the nearest number of sheep agents $1 \le n \le 30$ (resp. $1 \le n \le 60$). The two lines are quite similar in Figure 4 but the rate of success by the control with two sheepdog agents is slightly higher than the one with one sheepdog agent for $25 \le n \le 35$ in Figure 5.

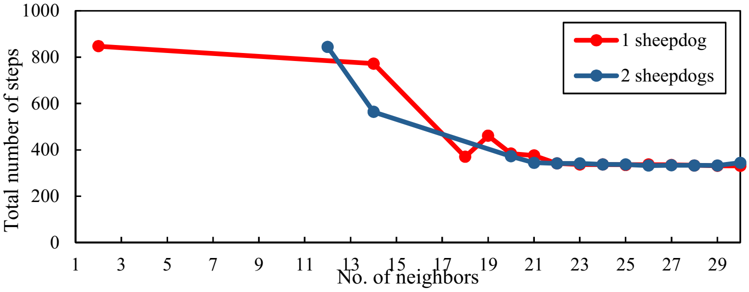


Fig. 6 Average number of time steps for the successful experiments with $N_s = 50$ and $1 \leq n \leq 30$.

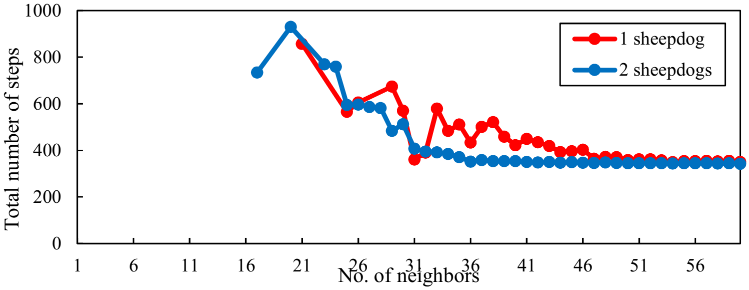


Fig. 7 Average number of time steps for the successful experiments with $N_s = 100$ and $1 \leq n \leq 60$.

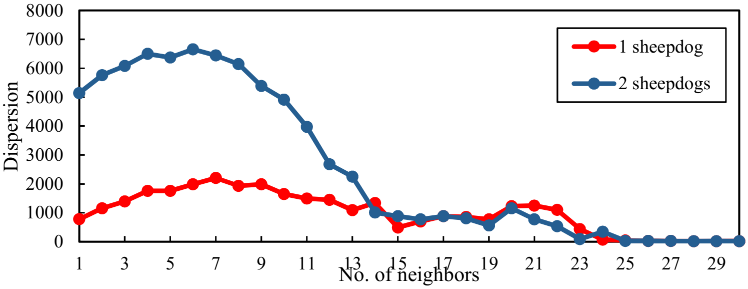


Fig. 8 Average dispersion at the end of the simulation with $N_s = 50$ and $1 \leq n \leq 30$.

For the average number of time steps in Figure 6 (resp. Figure 7), we focus on the plots of $n \geq 20$ (resp. $n \geq 30$) because from Figure 4 (resp. Figure 5) the rate of success is 10% or more. The control of 100 sheep agents with two sheepdog agents decreases the average number of time steps especially for $33 \leq n \leq 45$ in Figure 7.

As for the dispersion at the end of the trial, both Figure 8 and Figure 9 show that the control with one sheepdog agent results in lower dispersion than the control with two sheepdog agents. By the observations of the trials, the driving sheepdog spreads out the flock by forcefully heading for the driving position in Figure 1 in spite of the scatter of the sheep agents.

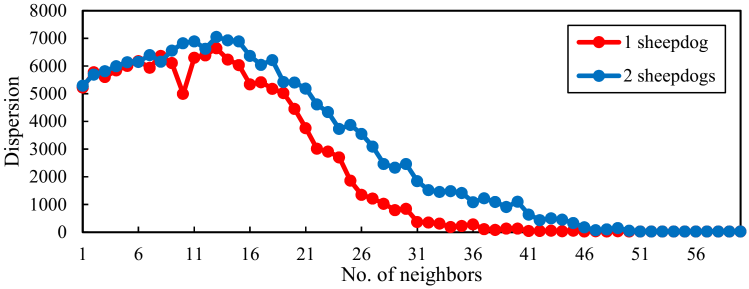


Fig. 9 Average dispersion at the end of the simulation with $N_s = 100$ and $1 \leq n \leq 60$.

4 Conclusion

In this paper, we introduced the simple crowd control method with two sheepdog agents called driving sheepdog agent and collecting sheepdog agent based on the previous researches with one sheepdog agent then compared our control with the previous one. Contrary to our expectation, the control with two sheepdog agents obtain few effects due to the driving sheepdog agent. In order to guide the sheep agents close to each other effectively by two sheepdog agents, a modification of the motion of sheepdog agents, especially for the driving sheepdog, is required for our future work.

Acknowledgements This work is supported in part by JSPS KAKENHI Grant Number 15K21308.

References

1. Sakiko Hayashi, Kaoru Fujioka, *A study on the efficient flock management by multi-agent systems* (in Japanese), In Proc. The 78th National Convention of IPSJ, pp. 2-379–2-380, 2016.
2. Jyh-Ming Lien, Burchan O. Bayazit, Ross T. Sowell, Samuel Rodriguez, Nancy M. Amato, *Shepherding Behaviors*, In Proc. IEEE ICRA'04, pp. 4159–4164, 2004.
3. Jyh-Ming Lien, Samuel Rodriguez, Jean-Phillipe Malric, Nancy M. Amato, *Shepherding behaviors with multiple shepherds*, In Proc. IEEE ICRA'05, pp. 3402–3407, 2005.
4. Reynolds, Craig W., *Flocks, herds and schools: A distributed behavioral model*, SIGGRAPH Comput. Graph., Vol. 21, No. 4, pp. 25–34, 1987.
5. Yuichiro Sueoka, Takuto Kita, Masato Ishikawa, Yasuhiro Sugimoto and Koichi Osuka, *Harnessing control and performance analysis for sheepdog system: approach from spatial discretization* (in Japanese), Transactions of the JSME, Vol. 80, No. 809, pp. 1–314, 2014.
6. Strömbom, Daniel, Richard P. Mann, Alan M. Wilson, Stephen Hailes, Jennifer Morton, David J. T. Sumpter, Andrew J. King, *Solving the shepherding problem: heuristics for herding autonomous, interacting agents*, *Journal of The Royal Society Interface*, Vol. 11, No. 100, 2014.
7. Richard T. Vaughan, Neil Sumpter, Jane V. Henderson, Andy Frost, Stephen Cameron, *Experiments in Automatic Flock Control*, *Robotics and Autonomous Systems*, Vol. 31, pp.109-117, 2000.

Design and Control of an Omnidirectional Wheelchair for Moving in Room Narrow Spaces

Keita Matsuo and Leonard Barolli

Abstract Because of aged tendency of population and rapid growth in the number of the disabled caused by diseases or injuries, the wheelchair with good performance for the aged and disabled is attracting more and more attention from the society. Also, the wheelchair can provide the user with many benefits, such as maintaining mobility, continuing or broadening community and social activities, conserving strength and energy, and enhancing quality of life. The wheelchair body must be compact enough and should be able to make different movements in order to have many applications. In this paper, we present the design and control of an omnidirectional wheelchair for moving in room narrow spaces. Finally, we discuss some implementation and application issues.

1 Introduction

Robots are being steadily introduced into modern everyday life and are expected to play a key role in the near future. Typically, the robots are deployed in situations where it is too dangerous, expensive, tedious, and complex for humans to operate.

Recently, the older age population is increased. According to WHO (World Health Organization) by 2025, the increase of population over aged 60 is predicted to reach 23% in North America, 17% in East Asia, 12% in Latin America and 10% in South Asia. There are over 600 million disabled persons in the world constituting nearly 10% of the global population.

Because of aged tendency of population and rapid growth in the number of the disabled caused by diseases or injuries, the wheelchair with good performance for

Keita Matsuo and Leonard Barolli
Department of Information and Communication Engineering
Fukuoka Institute of Technology (FIT)
3-30-1 Wajiro-Higashi, Higashi-Ku, 811-0295 Fukuoka, Japan
e-mail: kt-matsuo@fit.ac.jp, barolli@fit.ac.jp

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_77

the aged and disabled is attracting more and more attention from the society. There are many research works on wheelchairs including wheelchair for recovery, climbing stairs, and multifunction [1]. Therefore, it is necessary to design a wheelchair with the feature of easy-walking, convenient-use, and small-radius-swerving because the wheelchair is often used in a relatively narrow and small room.

The wheelchair body must be compact enough to go through narrow spaces. The wheelchair must be wide enough to prevent the patient from falling on the floor. A large footprint is therefore desirable for stability and safety, while wheelchairs must conform to dimensional constraints. Also the footprint must be compact since a large footprint does not allow the vehicle to move in a closely confined place. Stability and mobility are therefore conflicting requirements.

In recent years, more and more convenient facilities and equipments have been developed in order to satisfy the requirements of elderly people and disabled people. Among them, wheelchair is used widely. A wheelchair can provide the user with many benefits, such as maintaining mobility, continuing or broadening community and social activities, conserving strength and energy, and enhancing quality of life.

In this paper, we present the design and control of an omnidirectional wheelchair for moving in room narrow spaces. Then, we discuss some implementation and application issues.

The structure of this paper is as follows. In Section II, we introduce the related work. In Section III, we present the proposed omnidirectional wheelchair system. In Section IV, we discuss some implementation and application issues. Finally, conclusions and future work are given in Section V.

2 Related Work

Most of the work, for mobile robots has be done for improving the quality of life of disabled people. One of important research area is robotic wheelchairs. The persons having physical impairment often find it difficult to navigate the wheelchair themselves. The reduced physical function associated with the age or disability make independent living more difficult. Many research works have been undertaken to reduce the problem of navigation faced by the physically and mentally challenged people and also older age persons. One of the suggestive measures are the development of a Brain Control Interface (BCI), that assist an impaired person to control the wheelchair using his own brain signal. The research proposes a high-frequency SSVEP-based asynchronous BCI in order to control the navigation of a mobile object on the screen through a scenario and to reach its final destination [2]. This could help impaired people to navigate a robotic wheelchair. The BCIs are systems that allow to translate in real time the electrical activity of the brain in commands to control devices, provide communication and control for people with devastating neuromuscular disorders, such as the Amyotrophic Lateral Sclerosis (ALS), brainstem stroke, cerebral palsy, and spinal cord injury [3].

One of the key issue in designing wheelchairs is to reduce the caregiver load. Some of the research works deal with developing prototypes of robotic wheelchairs that helps the caregiver by lifting function or which can move with a caregiver side by side [4, 5]. The lifting function equipment facilitates easy and safe transfer from/to a bed and a toilet stool by virtue of the opposite allocation of wheels from that for a usual wheelchair. The use of lifting function and the folding of frames makes it more useful in indoor environments. Robotic wheelchair based on observations of people using integrated sensors can move with a caregiver side by side. This is achieved by a visual-laser tracking technique, where a laser range sensor and an omnidirectional camera are integrated to observe the caregiver.

Another important issue for the design of wheelchair is the collision detection mechanism. The omnidirectional wheelchairs with collaborative controls ensures better safety against collisions. Such wheelchairs possess high level of ability when moving over a step, through a gap or over a slope [6, 7]. To achieve omnidirectional motion, vehicles are generally equipped with an omniwheel consisting of a large number of free rollers or a spherical ball wheel. The development of such omniwheels attempts to replace the conventional wheel-type mechanism.

There are also other works which deal with vision design of robotic wheelchairs by equipping the wheelchair with camera for monitoring wheelchair movement and obstacle detection and pupil with gaze sensing [8, 9]. Prototype for robotic wheelchairs have been suggested in various research works, which are exclusively controlled by eye and are used by different users, while proving robust against vibration, illumination change, and user movement [10, 11].

To enable older person to communicate with other people the assisting devices have been developed. They can improve the quality of life for the elderly and disabled people by using robotic wheelchairs. The head gesture recognition is performed by means of real time face detection and tracking techniques. They developed a useful human-robot interface for RoboChair [12].

3 Proposed Omnidirectional Wheelchair System

In this section, we will describe the design and the implementation of the proposed wheelchair system.

In Fig. 1, we show a conventional wheelchair. In the case of kitchen space, the wheelchair can not move on the left or on the right sides. In order to move on right side as shown in Fig. 2, the wheelchair should make 5 movements. This is only one example of using the wheelchair, but when the wheelchair is used in indoor environment is difficult to make movements because of the small spaces.

In order to deal with these problems, we propose an omnidirectional wheelchair as shown in Fig. 3. The image of proposed omnidirectional wheelchair is shown in Fig. 4. The image of proposed omnidirectional wheelchair is shown in Fig. 5.

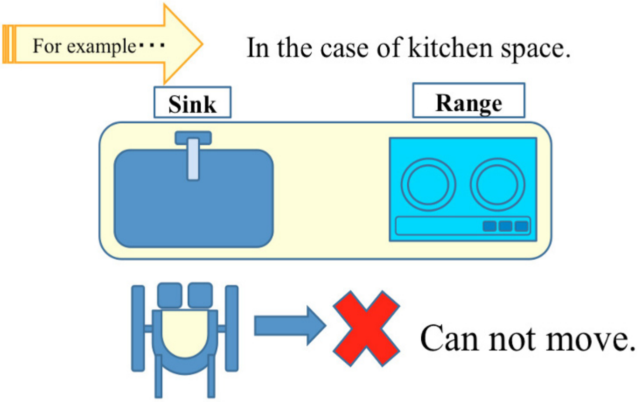


Fig. 1 Conventional wheelchair.

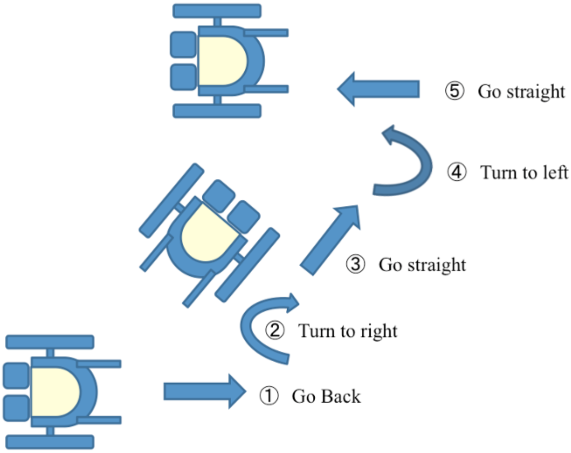


Fig. 2 Moving of conventional wheelchair.

3.1 Kinematics

For the control of the wheelchair are needed the omniwheel speed, wheelchair movement speed and direction.

Let us consider the movement of the wheelchair in 2 dimensional space. In Fig. 6, we show the onmiwheel model. In this figure, there are 3 onmiwheels which are placed 120 degree with each other. The omniwheels are moving in clockwise direction as shown in the figure. We consider the speed for each omniwheel M1, M2 and M3, respectively.

As shown in Fig. 6, the axis of the wheelchair are x and y and the speed is $v = (\dot{x}, \dot{y})$ and the rotating speed is $\dot{\theta}$. In this case, the moving speed of the wheelchair can be expressed by Eq. (1).

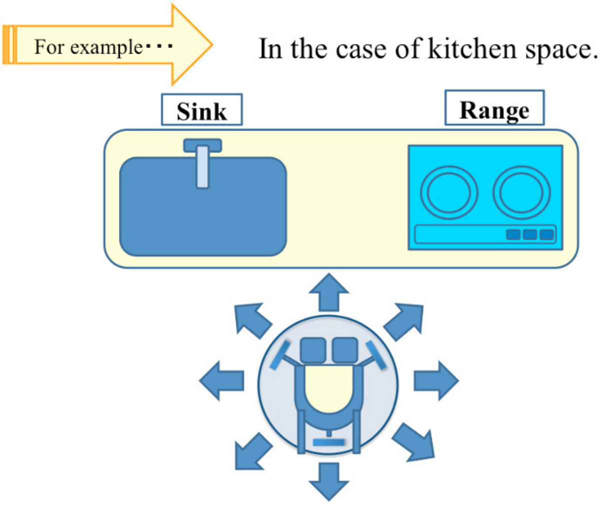


Fig. 3 Design of omnidirectional wheelchair.

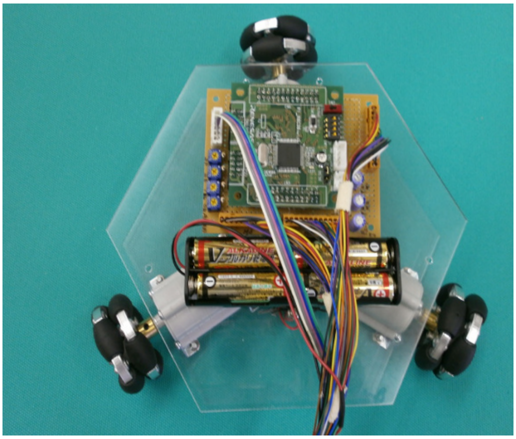


Fig. 4 Image of proposed test omnidirectional wheelchair.

$$V = (\dot{x}, \dot{y}, \dot{\theta}) \tag{1}$$

Based on Eq. (1), the speed of each omniwheel can be decided. By considering the control value of the motor speed ratio of each omniwheel as linear and synthesising the vector speed of 3 omniwheels, we can get Eq. (2) by using Reverse Kinematics, where (d) is the distance between the center and the omniwheels. Then, from the rotating speed of each omniwheel based on Forward Kinematics, we get the wheelchair moving speed. If we calculate the inverse matrix of Eq. (2), we get Eq. (3). Thus, when the wheelchair moves in all directions (omnidirectional movement), the speed for each motor (theoretically) is calculated as shown in Table 1.

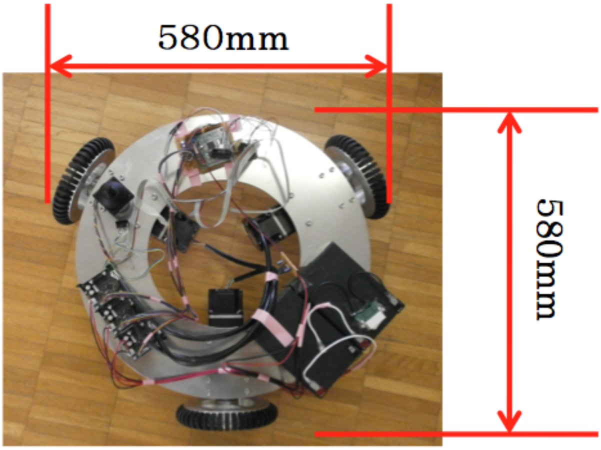


Fig. 5 Image of Proposed Omnidirectional Wheelchair.

$$\begin{vmatrix} M_1 \\ M_2 \\ M_3 \end{vmatrix} = \begin{vmatrix} 1 & 0 & d \\ -\frac{1}{2} & -\frac{\sqrt{3}}{2} & d \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} & d \end{vmatrix} \begin{vmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{vmatrix} \tag{2}$$

$$\begin{vmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{vmatrix} = \begin{vmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ 0 & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{3d} & \frac{1}{3d} & \frac{1}{3d} \end{vmatrix} \begin{vmatrix} M_1 \\ M_2 \\ M_3 \end{vmatrix} \tag{3}$$

3.2 Control System of the Proposed Omnidirectional Wheelchair

For the control of the proposed omnidirectional wheelchair, we considered R8C38 CPU board from Renesas Electronics Corporation. This CPU board has a small size and high speed processing time. The core of the CPU has a maximum frequency of 20 MHz. It is equipped with a flash memory, which is easy to rewrite. The R8C38 board has the following features:

- 8bit multi functions timer: 2,
- 16bit output competition timer: 5,

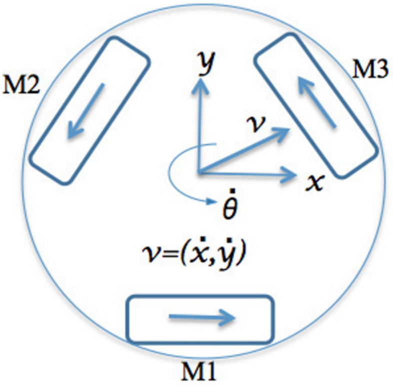


Fig. 6 Model of omniwheel.

Table 1 Motor speed ratio.

Direction	Motor Speed Ratio		
(Degrees)	Motor1	Motor2	Motor3
0	0.00	-0.87	0.87
30	0.50	-1.00	0.50
60	0.87	-0.87	0.00
90	1.00	-0.50	-0.50
120	0.87	0.00	-0.87
150	0.50	0.50	-1.00
180	0.00	0.87	-0.87
210	-0.50	-1.00	-0.50
240	-0.87	0.87	0.00
270	-1.00	0.50	0.50
300	-0.87	0.00	0.87
330	-0.50	-0.50	1.00
360	0.00	-0.87	0.87

- Real time clock timer: 1,
- UART/clock synchronization type serial interface: 3 channels,
- 10bit A/D converter: 20 channels,
- 8bit D/A converter: 2 circuits,
- Voltage detected circuit,
- Number of output and input port: 75,
- External interrupt input: 9.

In Fig. 7 is shown the control system for the proposed omnidirectional wheelchair. The direction movement of the wheelchair is decided by the Joystick. The Analog-Digital Converter changes the analog value to a digital value needed for R8C38 board. The R8C38 board based on the Eq. (2) calculates the motors control value. Based on this value, the Pulse Width Modulation (PWM) generator generates an

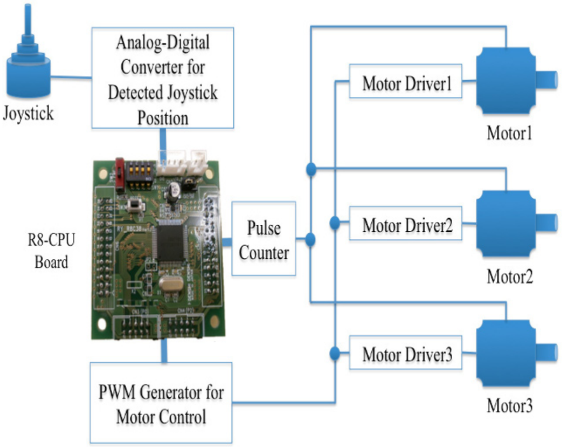


Fig. 7 Control system for omnidirectional wheelchair.

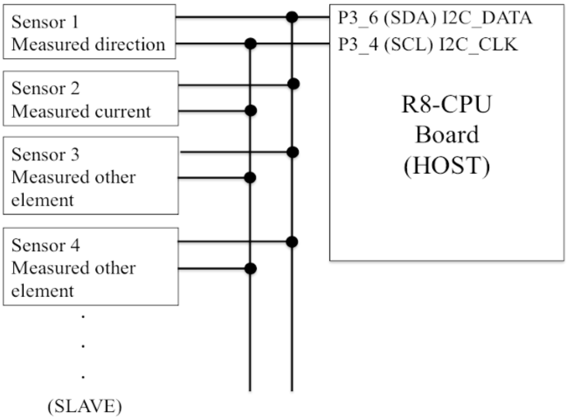


Fig. 8 Schematic diagram of communication between sensors and R8-CPU by I2C.

appropriate value for the control of each motor. The number of rotation of each motor is detected by Pulse Counter and is sent to the R8C38 board in order to make a correct feedback control.

We implemented a directional sensor on the wheelchair. The communication between the sensor and R8-CPU is done by I2C (Inter Integrated Circuit). The directional sensor can keep the directions of the wheelchair when playing sports. The I2C is used for different kind of sensors. The maximum number of sensors that can communicate on the same bus of R8-CPU is 112. The schematic diagram of communication between sensors and R8-CPU by I2C is shown in Fig. 8. While, the implemented circuit is shown in Fig. 9.

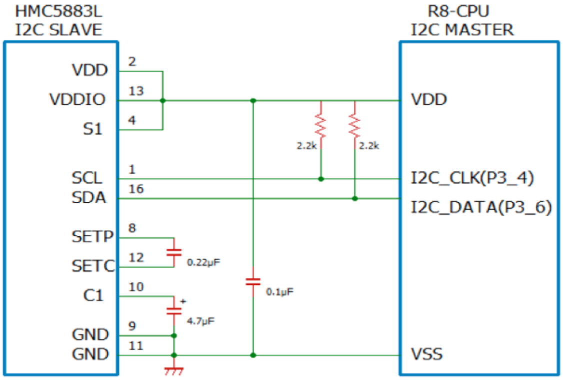


Fig. 9 Implemented circuit (HMC5883L and R8-CPU connection).

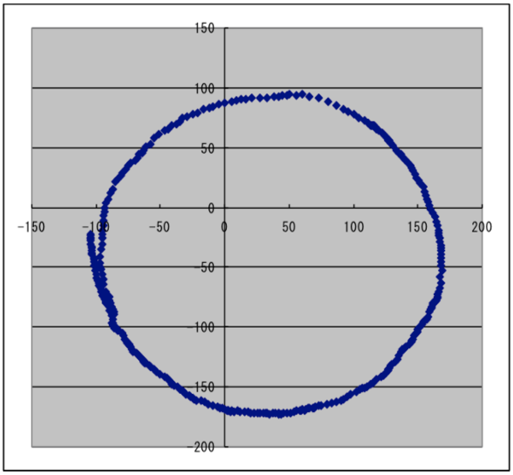


Fig. 10 Measured directions by HMC5883L.

4 Implementation and Application Issues

In this research, we used the implemented wheelchair for moving in room narrow spaces.

In Fig. 10 are shown measured data by HMC5883L. From the data, we can see that the directional sensor should be calibrated in order that the circle be in the center. Also, when playing sports, the players always should be face to face. However, the direction of the wheelchair may change during the movement. For this reason, we decided to keep the body of the player in the same direction by using directional sensor.

The proposed omnidirectional wheelchair can be controlled remotely using WiFi communication system embedded on the wheelchair as shown In Fig. 11. Using this

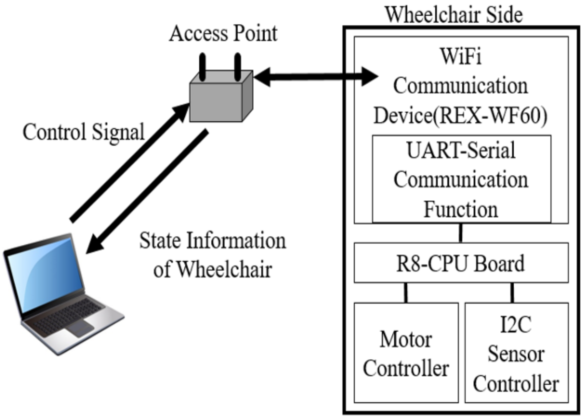

Access Point
Wheelchair Side
WiFi Communication Device(REX-WF60)
UART-Serial Communication Function
Control Signal
State Information of Wheelchair
R8-CPU Board
Motor Controller
I2C Sensor Controller

Fig. 11 Communication system for omnidirectional wheelchair.

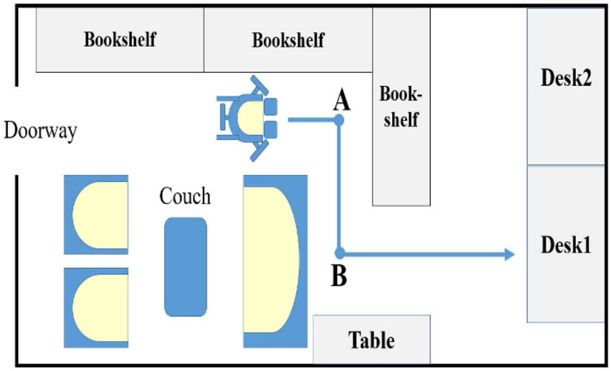


Fig. 12 Image of narrow spaces in the room.

communication system, we can get the number of rotation for each motor, direction of the wheelchair and the value of the Joystick controller. In particular, real speed of each motor is very important because when there is difference between the value of control signal and real speed, it causes incorrect moving of the wheelchair. So, we have to modify the control signal based on this value.

Fig. 12 shows the model of a room in our building. For instance, when wheelchair user want to go to the Desk1, if is used our omnidirectional wheelchair he doesn't need to turn the wheelchair. But if is used ordinary wheelchair, it is needed to turn two times (point of A and B). Moreover, move of the omnidirectional wheelchair is more convenient when the desks are close to each other. In Fig. 13 are shown some snapshots of the proposed omnidirectional wheelchair when moving in a room.

Now, the movement of wheelchair is done by joystick. We want to make the control of the wheelchair more convenient for disabled persons. For this reason, we want the implement an automatic control.

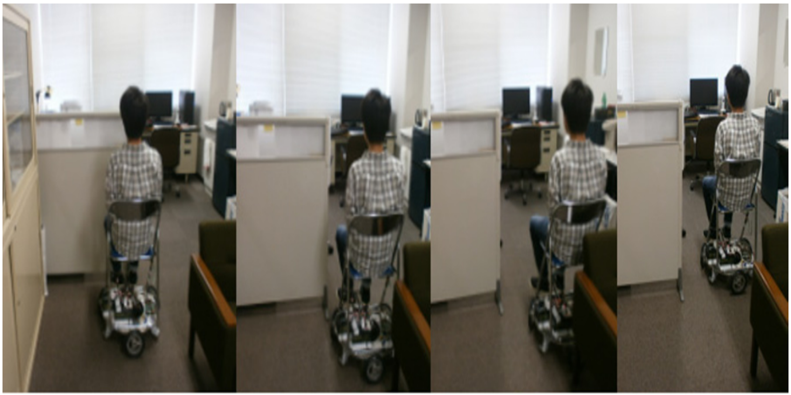


Fig. 13 Snapshots of using omnidirectional wheelchair in narrow spaces.

In this work, we used the implemented wheelchair for moving in narrow spaces of a room, but it can be used for sports such as tennis, basketball, and playing badminton. Also, the application of the proposed wheelchair for transport in plants and factories will be considered.

5 Conclusions and Future Work

In this paper, we presented the design and implementation of an omnidirectional wheelchair for moving in narrow spaces of a room. The implemented wheelchair can be used also for sports such as tennis, basketball and playing badminton.

We introduced some of the previous works and discussed the related problems and issues. Then, we presented in details the kinematics and the control system for the proposed omnidirectional wheelchair. Finally, we discussed some implementation and application issues.

In the future work, we want to implement a system to detect the environment by using a map recognition method in order that the wheelchair avoid the collision with other objects. Also, we want the implement an automatic control.

References

1. T. Lu, K Yuan, H. Zhu, ”Research Status and Development Trend of Intelligent Wheelchair”, Application Technology of Robot, No. 2, pp.1-5, 2008.
2. P. F. Diez, V. A. Mut, E. M. A. Perona, E. L. Leber, ”Asynchronous BCI Control Using High-frequency SSVEP”, Journal of NeuroEngineering and Rehabilitation, Vol. 8, No. 39, 8 pages, doi:10.1186/1743-0003-8-39, July 2011.

3. S. M. Grigorescu, T. Luth, C. Fragkopoulos, M. Cyriacks, A. Graser, ”A BCI-controlled Robotic Assistant for Quadriplegic People in Domestic and Professional Life”, Robotica, Cambridge University Press, Vol. 30, No. 3, pp. 419-431, 2012.
4. Y. Mori, N. Sakai, K. Katsumura, ”Development of a Wheelchair with a Lifting Function”, Advances in Mechanical Engineering, Volume 2012, Article ID: 803014, 9 pages, doi:10.1155/2012/803014, 2012.
5. Y. Kobayashi, Y. Kinpara, T. Shibusawa, Y. Kuno, ”Robotic Wheelchair Based on Observations of People Using Integrated Sensors”, Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 11-15, October 2009.
6. S. Ishida, H. Miyamoto, ”Collision Detecting Device for Omni directional Electric Wheelchair”, Robotics, Hindawi Publishing Corporation, Volume 2013, Article ID: 672826, 2013.
7. T. Carlson, Y. Demiris, ”Robotic Wheelchair with Collaborative Control”, Proc. of IEEE International Conference on Robotics and Automation, pp. 5582-5587, 2010.
8. P. Jia, H. H. Hu, T. Lu, K. Yuan, ”Head Gesture Recognition for Hands-free Control of an Intelligent Wheelchair”, Industrial Robot: An International Journal, Vol. 34, No. 1, pp.60-68, doi: 10.1108/01439910710718469, 2007.
9. K. Arai, R. Mardiyanto, ”Electric Wheelchair Controlled by Eye-Only for Paralyzed User”, Journal of Robotics and Mechatronics, Vol. 23, No. 1, pp. 66-74, 2011.
10. A. Escobedo, A. Spalanzani, C Laugier, ”Multimodal Control of a Robotic Wheelchair: Using Contextual Information for Usability Improvement”, Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2013), doi: 10.1109/IROS.2013.6696967, pp. 4262-4267, 2013.
11. J. Gonzalez, A. J. Munoz, C. Galindo, J. A. Fernandez-Madrigal, J. L. Blanco, ”A Description of the SENA Robotic Wheelchair”, Proc. of IEEE Mediterranean Conference (MELECON-2006), pp. 437-440, 2006.
12. H. Wang, G. G. Grindle, J. Candiotti, C. Chung, M. Shino, E. Houston, R. A. Cooper, ”The Personal Mobility and Manipulation Appliance (PerMMA): A Robotic Wheelchair with Advanced Mobility and Manipulation”, Proc. of IEEE Eng Med Biol Soc., pp. 3324-3327. doi: 10.1109/EMBC.2012.6346676, 2012.

Development of Training System for Pedaling Skill by Visualizing Muscle Activity Pattern

Takuhiro Sato[1], Shoma Kushizaki[1], Shimpei Matsumoto[2],
Tomoki Kitawaki[3], Tatsushi Tokuyasu[1]

[1] Fukuoka Institute of Technology, 3-30-1 Wajiro-higashi, Higashi-ku, Fukuoka Japan
mdm15002@bene.fit.ac.jp
[2] Hiroshima Institute of Technology, 2-1-1 Miyake, Saeki-ku, Hiroshima, Japan.
s.matsumoto.gk@cc.it-hiroshima.ac.jp
[2] Kansai Medical University, 2-5-1 Shin-machi, Hirakata City, Osaka, Japan.
kitawaki@hirakata.kmu.ac.jp

Abstract. This paper presents a training system of pedaling skill for a cyclist. In order to maximize the competition performance of a cyclist, improving the pedaling skill, that converts his/her physical strength to the impulsive force of a bicycle, is effective. In the field of cycling competition, the pedaling skill has been vaguely discussed for a long time. Then, this study have proposed a method to visualize the pedaling skill by using the kinetic information of a cyclist who pedals on a bicycle. In this paper, we make a training system for the improvement of pedaling skill and discuss the experimental results.

1 Introduction

The feature of a bicycle which had been utilized in a cycling race is to use the binding pedal, that attaches the sole of shoe on the body of a pedal. According to the categories of cycling races the basic specification of a bicycle varies, however it is necessary for a racer to use the binding pedal in order to convert his/her physical power to the impulsive force of the bicycle. In the field of cycling competition, the pedaling skill that realizes effective pedaling exercise, has been well known and practically utilized [1][2]. By the way, there is three degrees of freedoms at the parts of both a saddle and a handle of a competitive bicycle as shown in Fig. 1. The settings of both a saddle and a handle determine the fundamental riding posture of a rider since the use of binding pedals must be minimum requirement among riders. Whereas the importance of setting both a saddle and a handle has been recognized, these components are adjusted according to the feelings of real riding. The setting of both a saddle and a handle influences the activity levels of the muscles recruited in pedaling motion and the range of motion of lower limbs [3][4]. Then, the authors have thought the competitive ability of a rider should be increased by practicing pedaling skill under appropriate setting mechanical components. For inexperienced rider or beginner cyclist, their inner senses enough to assess the feeling of riding have not been developed with comparing to skilled cyclists. Therefore, it would become an effective method to enhance the training effectiveness to assess the pedaling skill of a

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_78

rider based on his/her objective information measured during pedaling exercise. Then, this paper firstly discusses about the relationship between pedaling skill and the stability of muscle activity pattern of lower limbs, and aims to construct a training system for pedaling skill which is subjected to intermediate grade riders. Especially, we aim to develop the function which enables a trainee to monitor own inner information such as muscle activity pattern to obtain the ideal riding form.

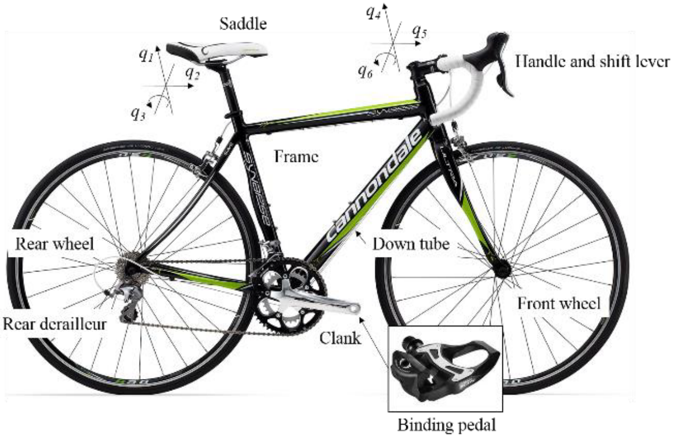


Fig. 1 Basic constitution of a cycle road racer

2 Experimental device

In general, the kinematic principle of cycling exercise differs according to the slope of road. For example, we are completely unaffected by the force of gravity while riding on a flat road, only the air resistance increases corresponding to the velocity. By the way, on a sloping road, we have to pedal against both of the gravitation and the resistance of air. As mentioned above, the condition of pedaling exercise varies according to the situation of course. In order to make our concern as simple as possible, we consider the pedaling skill utilized while riding on a flat course and this study aims to visualize the pedaling skill. This study employed a commercially available cycle trainer to fix a competitive bicycle for road race. Fig. 2 shows the experimental device we have developed in this study, where a road bicycle (RS8, Bridge stone anchor) is set to the cycle trainer (Elite Crono Fluid Elastogel, Elite). At the part of down tube, a rotary encoder (E6C2-CWZ1X, Omron) was mounted in order to measure the rotation angle of a clank. The output signal of rotary encoder is inputted to a computer via a counter board (PCI-6205C, Interface Inc.). And the hub of rear wheel plays a role of power meter and the measured pedaling load is transmitted to the cycle computer (Edge 500, Garmin Inc.) which is mounted on a handle. A trainee can monitor the riding information by viewing the cycle computer. On the monitor, the velocity[km/h], the pedal's rotation speed [rpm], and the heart rate[bpm] are expressed, too. This study has confirmed the accuracy of the pedal's rotation speed measured the rotary encoder. The muscle activity of lower limb during pedaling exercise can be measured by using a multi-channel high sensitive amplifier (MEG-6108,　NIHON KODEN. Inc.)

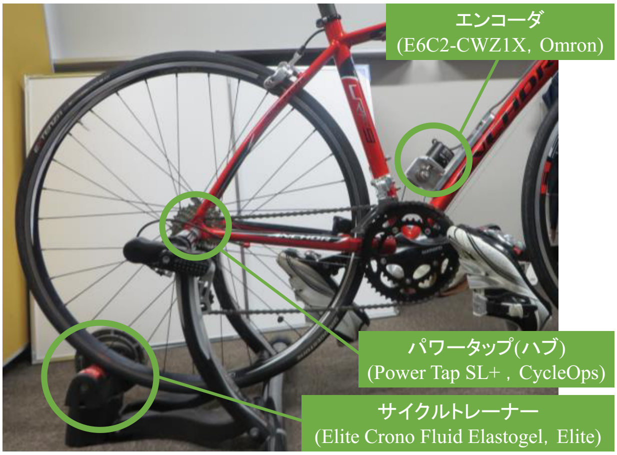


Fig. 2 Construction of experimental device

3 Pedaling skill visualization

3.1 Signal processing of electromyogram

This section describes how to process the electromyogram measured in pedaling exercise. In our experiment, a test subject pedals at the constant speed of 90 rpm with monitoring the cycle computer. The pedaling exercise at 90 rpm has been known as the moderate exercise load for muscle and lung and most of cyclists have referred this exercise load in their training on a flat road [5]. This study focused on the muscles of lower limb that are recruited in pedaling motion as shown in Fig. 3 and measures the electromyograms of them. In general, electromyogram is easily influenced of muscle fatigue then this study measures only thirty seconds after the pedaling speed became stable [6][7]. The electromyograms of each muscle measured in 45 times of clank rotation was applied to root mean square, and the rotation angle data was averaged at 15 degrees intervals. Additionally, we standardized them with all 24 period to derive the muscle activity, and each muscle activity is expressed as e15n (n=1～24). The standardization of electromyogram is necessary to eliminate the effects of sensitivity of sensors and/or the density of muscle. The equation (1) factorized the muscle activity for one rotation of pedaling exercise, where e means the average value and the index number is for angular period. Since we measure four muscles to assess the activity pattern of lower limb, then the equation (2) is derived by applying the equation (1) to these four muscles.

$$q_m = \left[e_{15}, e_{30}, \cdots e_{345}, e_{360}\right] \qquad (1)$$

$$mp = \left[q_{RF}, q_{BF}, q_{AT}, q_{GM}\right]^T \qquad (2)$$

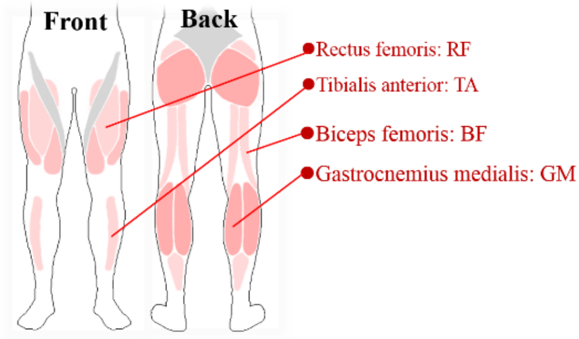


Fig. 3 Schematic diagram of the leg muscles used to evaluate muscle activity pattern in pedaling exercise

3.2 Strategy of visualizing pedaling skill

This study noticed the racing career more than the physical strength and performance to employ a skilled test subject. And the beginner for binding pedal was defined as a beginner subject and the other was defined as intermediate subjects. This paper employed one skilled cyclist who has over fourteen cycling career and one intermediate cyclist who has three cycling career as our test subjects. According to [2], the muscle activity pattern varied slightly among the highly trained cyclists, then there is not problem to employ one skilled cyclist as a standard subject on progressing our basic research. Firstly, this study implements preparatory experiment with the skilled subject to obtain the standard of muscle activity pattern during pedaling exercise.

In the beginning of experiment, a subject shifts the gear ratio to make the exercise load 150 watt, and continues to pedal at 90 rpm of the pedaling speed. About the electromyogram measured in pedaling exercise, two of muscle activity pattern are generated from the randomly selected pedal rotation, and described as mp_{s1}^T and mp_{s2}^T, respectively. These two patterns of muscle activity will be utilized as the standard to assess the pedaling skill of other test subjects. Our proposal to visualize the pedaling skill is realized by clarifying the difference between the skilled cyclist and a test subject, where the principal component analysis applies to the muscle activity patterns, so that the muscle activity pattern of a test subject is described as mp_B^T and then the equation (3) can be obtained.

$$X = \begin{bmatrix} mp_B^T & mp_{S1}^T & mp_{S2}^T \end{bmatrix} \tag{3}$$

The contribution ratio and the singular vector of each component derived by the process of principal component analysis conducted on the equation (3) are related to the roles of each muscle in pedaling exercise.

4 Experimental result

In order to understand the means of the scores of each principal component derived from the equation (3), an another muscle activity pattern from a voluntary rotation of pedaling motion is made and plugged it into the first element. Consequently, the cumulative contribution ratio of the first principal component became 91% and the signs of all eigenvector were positive. Then, the authors considered that the first principal score includes important factors as evaluation index of pedaling skill. It can be considered that pedaling skill is composed of the combinations of contraction and relaxation among the muscles recruited in pedaling exercise in order to output the power toward the direction of tangent to the clank rotational circle, and this skill might be improved though training. Then, in this study, the sign of the first principal component score became positive, then the pedaling exercise performed during the measurement period was realized with both contraction and relaxation of the muscles. By the way, the authors understood that the negative sign of the first principal component score mean that over muscle contraction appeared in the pedaling exercise performed during the measurement period. From this perspective in respect to the sign of the first principal component score, this study visualized the pedaling skill as described in Fig. 4, where the colors (Blue, Red, Green, and Purple)of circles corresponds to the muscles (RF, BF, AT, and GM).

The activity level of either contraction or relaxation of muscle is reflected to the radius of circle. Hence, Fig. 4 indicates that both the degree and the timing of muscular contraction and relaxation comparing to the standard data established from pedaling exercise of a skilled cyclist. Next, the principal component analysis was implemented, where the first element was replaced with the muscle activity pattern mp_B^T derived from the pedaling exercise of the intermediate subject, where the pedaling load was set to 150 watt. The contribution ratios of the first and the second principal components were 68.3% and 29.5%, respectively. Then, the cumulative contribution ratio became 97.9 %. Therefore, we need to understand the means of both the first and the second principal component. According to the sign of eigenvectors, the pedaling skill can be evaluated from the principal component scores plotted in four quadrants as shown in Fig. 5. The type of subject is judged by the sign of the first eigenvector and the state of muscle such as contraction or relaxation is judged by the sign of the second eigenvector. From the equation (3), it can be thought that the first eigenvector and the second eigenvector are deeply related to the muscle activity pattern of the intermediate subject and the skilled subject. And, the scores plotted in the second quadrant and third quadrant were derived from the muscle activity pattern of the intermediate cyclist because the sign of first element of the first eigenvector was positive.

Additionally, the scores plotted in the first quadrant and the fourth quadrant indicate the level of muscular contraction or relaxation of the intermediate subject, because the sign of all second eigenvector were positive. On the other hand, the scores plotted in the second quadrant and the third quadrant indicate the level of muscular contraction or relaxation of the skilled subject. This study could realized the online usage of the mentioned visualization method for pedaling skill, then a trainee is able to pedal with comparing to the pedaling skill of a skilled cyclist.

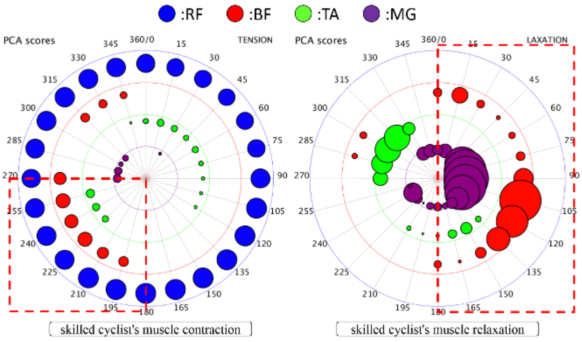


Fig. 4 Results of visualizing pedaling skill in respect to muscle contraction and relaxation [150watt]

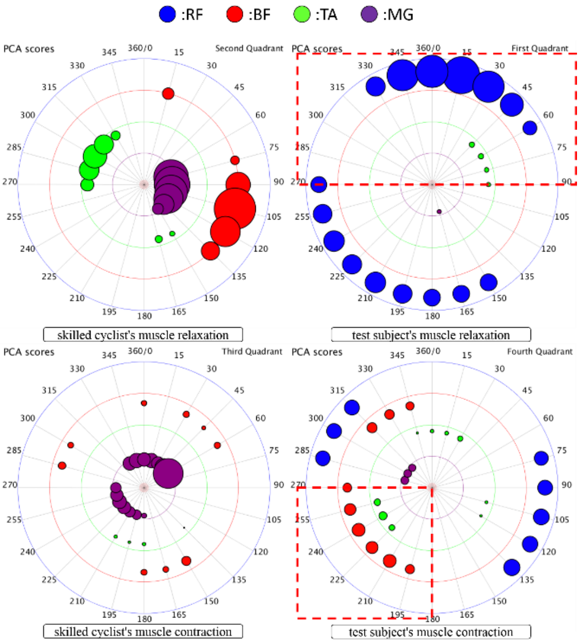


Fig. 5 Results of visualizing pedaling skill of an intermediate cyclist comparing to a skilled cyclist [150watt]

5 Discussion

This study made it possible to visualize the pedaling skill by make the process of principal component analysis the online usage during pedaling exercise. Though this means the accomplishment of basic function of our system, it was not cleared whether a trainee would improve the pedaling skill. Because there are individual difference of the physical constitution and the amount of muscle among cyclists.

For future, we will construct the numerical simulation model of lower limb motion and attempt to estimate the optimal muscle activity pattern according to the physical constitution of a test subject. After them, we like to substitute the optimized muscle activity pattern to the first and second element of the equation (3).

Acknowledgments. This work was supported by JSPS KAKENHI Grant Number, 26750293

References

1. Schmidt A, "Handbook of competitive cycling: Training, keep fit, tactics", Oxford, Meyer & Meyer, pp. 13-16. 1998.
2. Andrew R. Chapman, Bill Vicenzino, Peter Blanch, Paul W. Hodges. :Patterns of leg muscle recruitment vary between novice and highly trained cyclists, Journal of Electromyography and Kinesiology, Vol.18, No.3, pp. 359-371. 2008.
3. Tatsushi Tokuyasu, Shoma Kushizaki, Shimpei Matsumoto, Tomoki Kitawaki. :Saddle height setting of cycling road bike focusing on variance of muscle activity, ICIC Express Letters, Express Letters, Vol.9, No.1, pp.53-58. 2014.

4 Andrew R. Chapman, Bill Vicenzino,, Peter Blanch, Joanna J. Knox, Steve Dowlan, Paul W. Hodges. :The influence of body position on leg kinematics and muscle recruitment during cycling, Journal of Science and Medicine in Sport. Vol.11, No.6, pp. 519-526. 2008.

5. J. M. Hagberg, J. P. Mullin, M. D. Giese, E. Spitznagel, Effect of pedaling rate on submaximal exercise responses of competitive cyclists, Journal of Applied Physiology, Vol. 51, No.2, pp. 447-451, Aug 1981.
6. Raymond C.H. So, Joseph K.-F. Ng, Gabriel Y.F. Ng. :Muscle recruitment pattern in cycling: a review, Physical Therapy in Sport, Vol.6, No.2, pp. 89-96. 2005.
7. Hug F and Dorel S. :Electromyographic analysis of pedaling: A review", Journal of Electromyography and Kinesiology, Vol.19, No.2, pp. 182-198. 2009.

Proposal for a strategy to discover the students with need for a learning support by using text analysis

Chicako Miyamoto[1], Naoko Furukawa[1], Tatsushi Tokuyasu[1]
[1] Fukuoka Institute of Technology, 3-30-1 Wajiro-higashi, Higashi-ku, Fukuoka, Japan
c-miyamoto@fit.ac.jp

Abstract. This paper proposes a strategy to discover the students who might need some learning support by using text analysis for a brief interview with a student. In order to find the psychological trends such as depression, autism, and interpersonal fear from the result of text analysis, firstly the authors define the significant keywords in respect to these psychological trends based on the experimental rules of clinical psychotherapists. Next, the scenario of interview, the manual of interview, and the check sheet for the language and behavior of a student are built. Finally, the operation procedure of text mining to constantly discover the student with need for some kind of learning support is discussed.

1 Introduction

Recently, the number of the students who were certified as disability person has been increasing and the higher education institutions are strongly requested to have a positive approach to encourage their motivation for learning in Japan. In fact, Japanese government implemented "Disability discrimination method" in April 2015, and the securement and the support of learning opportunity in higher education for disability persons were statutorily obligated. By the way, there are lots of students who have not been officially certified and then they are learning with other ordinary students. However, at present, most of them had some troubles and had been found in the state of high anxiety and/or maladaptation in a academic year. Some of them are had been forced to choose a leave of absence and/or to withdraw from school. In order to understand the causes of them, Japan University Health Association had developed a screening test and has been widely utilized in Japan. This test has been called as University Personality Inventory (UPI)[1][2]. The researches of the development of questionnaire to investigate the reality of support needs has been implemented. And the relationship between the result of UPI and the occurrence of psychological problem or mental disorder has been investigated.

In Fukuoka Institute of Technology (FIT), which is the affiliation of authors, a brief interview about 10 minutes has been conducted on all admitted students in April every year. Because the number of student is large, the general officers who are not a psychotherapist have to play the role of interviewer. Additionally, it is difficult to tentatively employ the enough number of psychological therapists.

The purpose of interview is first to dispel anxiety about college life and check the psychological trends of a student. But, about 100 students had passed this

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_79

interview and they had had psychological problem during their academic year. We have conducted UPI test on the students who had psychological problem, however it is not realistic to implement UPI test to all admitted students because this test requires additional work to obtain the test result.

Then, the authors aim to develop an intelligent system which can estimate the psychological trends of a student based on the interview result with text analysis. This paper describes the strategy to complete our system and describes the technical factors necessary for the system.

2 Methods

2.1 Objective

Fig. 1 shows a flow of constructing the system that estimates the psychological trends of a student by using the interview results. This study is planning to develop the system through six steps from (a) to (f) as described in Fig. 1. The basic strategy is to embed the experimental rules of psychological therapists into an intelligent algorithm. Each technical factor will be shown in the following sections,

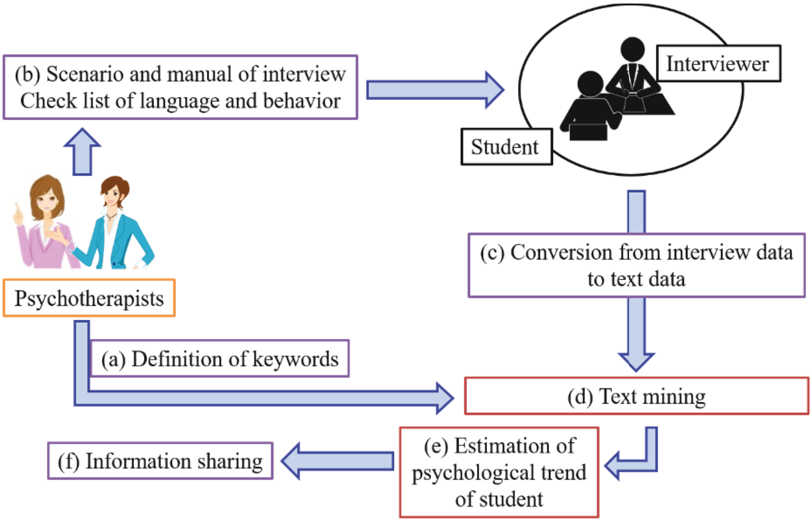

(b) Scenario and manual of interview
Check list of language and behavior
Interviewer
Student
Psychotherapists
(c) Conversion from interview data to text data
(a) Definition of keywords
(d) Text mining
(f) Information sharing
(e) Estimation of psychological trend of student

Fig. 1 A flow of the strategy and the six steps of constructing the system

2.2 Definition of keywords

This study aims to extract the information related to the psychological trends of a student by applying Text Mining to the interview results that is converted from the recorded voice data to text data. A computer can output any information with obeying the rule of Text Mining, however it is impossible for the computer to judge which data includes the information related to psychological trends of a student suffering from troubles. Therefore, it needs to determine the keywords related to psychological trends in the beginning of this study. Then, this study choose the keywords based on the experimental rules of psychotherapist, especially the keywords have to be the words which are talked by the students. Actually, both the first author and the second author of this paper have work experience over 10 years as a psychotherapist in the education institution in Japan. This study will refer their work experiences to define the keywords and weigh them according to the level of significant of psychological trends.

2.3 Materials for interview

This study assumed that a general officer working in a college plays role of an interviewer, because there is not enough number of psychotherapists in Japan against the number of psychological patients. Because the purpose of the interview is not only to investigate the psychological trends of them but also to eliminate the anxiety of them to their first campus life, additionally most of interviewers do not have the experience as a psychological therapist, we need to prepare the flow of interview. The authors assumed that the psychological trends of a student naturally emerge even in a dairy conversation. And it is necessary to make an interview manual for the interviewers. The interviewers have to elicit a talk related to psychological trend from a student. Additionally, we build a check list that an interviewer can easily record the features of language and behavior of a student during in the interview.

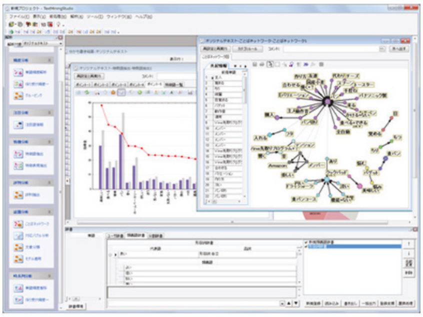
2.4 Data acquisition

The fundamental strategy is to educe the psychological trends hidden in language and behavior of a student during in the interview, in which the conversation between an interviewer and the student is recorded and converted to text data. This study applies Text Mining to the text data. Actually, the voice data contains the qualitative information such as voice tone, volume, and pitch. These qualitative information is eliminated by the conversion from voice data to text data. This study compensates them by using the check list. There is the terms of voice information, a sign of emotion, line of sight, and eye motion and so on in the check lists. In order to convert voice data into text data, a commercially available software (Dragon speech 11, NUANCE) has been adopted in this study. In order to utilize this software, this study has to improve the conversion function manually and we are trying to rise the accuracy of conversion by using some test data.

2.5 Text analysis

This study will conduct Text Mining on the interview data converted to text format in order to extract the psychological trends of a student. Text Mining is one of data mining methods for big data analysis, that enables us to objectively observe the conversation during the interview from the view point of the number of keywords appearance and/or the relationship between the keywords. So, we can assess the thought of a student and the temporal transition of his/her consciousness.

In practical, we adopt a commercial text mining software, Text Mining Studio produced by NTT data, in which a variety of types of analysis can be done by simple mouse operation. Fig. 2 shows a graphical user interface and the functions of Text Mining Studio we had chosen in this study. This study will categorize the words related to the keywords and make it possible to numerically evaluate the psychological trends of a student.



Basic function

- Morphology
- Dictionary
- Synonym extraction
- Attribution processing
- Filtering
- Grouping
- Original text reference
- Text edition
- Output & Print
- Reporting
- Categorization
- Window layout

Types of analysis

- Basic information
- Frequency analysis
- Mark analysis
- Feature analysis
- Evaluation analysis
- Topic analysis
- Temporal analysis

Fig. 2 Graphical user interface and the functions of Text Mining Studio

2.6 Psychological trends

According to our previous investigation toward this study, most of students who had troubles during academic years had the psychological aspects of dysphoria, autism, and interpersonal fear. As mentioned above, they were not checked as the students having any psychological troubles.

For the beginning of our study, we tentatively candidate these aspects as the psychological trends that our system assesses quantitatively by using the result of Text Mining.

The keywords we described in the section 2.2 would be related to the psychological trends defined in this section. The authors are considering the necessity of continuously revising the keywords, the flow of interview, and the psychological trends though the progression of this study.

2.7 Information Sharing

After applying Text Mining to the text data, the student information might be obtained according to our proposed method. Since the purpose of assessing the psychological trends of a student is to share among the workers of university, such as as teachers, office workers, and school therapists. However, most of the workers are lack of the psychological knowledge and/or the experiences of taking care of the students who have psychological troubles. Therefore, the authors are considering that it would be helpful to attempt the ways of dealing with the features of psychological trends. Then, this study proposes to visualize the psychological trends of a student by using a radar chart as shown in Fig. 3, where the average scores of all students and the score of the student assessed with our proposed method are concurrently drawn. To define the items of noting the psychological trends is our future work and we need to consider how to estimate the level of psychological trends. The details of our strategy would be described in the next section.

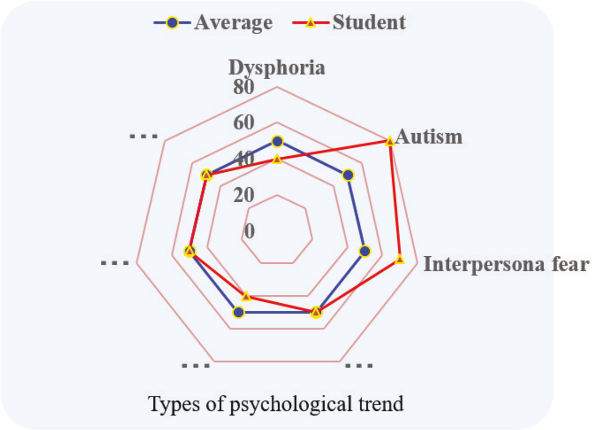


Fig. 3 Radar chart of psychological trends of a student

3. Discussions and future work

The outcomes of this study will construct the learning environment of students who needs some learning support and contribute to maintain and rise their motivation for learning. In the fields of education and psychology, it has been discussed about the effective learning support from the view point of communication skill and/or school adaptation. However, these are researched into many university students. Therefore, no research has considered how to estimate the level of psychological trends of a student and to share their information among the workers.

In our proposal, the keywords detected from the interview with Text Mining would be given the weights according to the significant level of psychological trends. For future, this study has to construct to an intelligent algorithm to automatically estimate the level of psychological trends of a student. This study is on the way of choosing the method to construct the algorithm. At present, we tentatively utilize Neural Network to connect the result of Text Mining to the psychological trends of a students, because in the phase of learning Neural Network we can embed the experimental rules of psychological therapists. ¥

References

1. http://health-uv.umin.ac.jp/
2. Masatoshi Nakagawa, Research on relation between UPI and occurrence of psychological problem, and relation between UPI and study accomplishment, Bulletin of DEN-EN CHOFU UNIVERSITY, vol. 1, pp. 51-67, 2006 (In Japanese).

Part VII
The 3-rd International Workshop on Secure Cloud Computing (SCC-2016)

Construction of Boolean Functions With Optimal Algebraic Immunity

Hang Liu and Dong Zheng and Qinglan Zhao

Abstract Boolean functions with good cryptographic properties act as important nonlinear components in symmetric cryptography which is often used to encrypt stored data for cloud computing. In this paper, we develop a new class of Boolean functions with optimal algebraic immunity by utilizing the Reed-Muller code. In addition, our new functions are balanced and have good nonlinearity.

1 Introduction

In the design of stream ciphers and block ciphers in cryptography, Boolean functions act as the main nonlinear component [6]. Boolean functions should have balancedness, high algebraic degree, high nonlinearity and high algebraic immunity(AI) to resist kinds of cryptanalytic attacks. However, there is a successful algebraic attack on stream ciphers. In 2003, Courtois and Meier successfully proposed an algebraic attack on several stream ciphers[2]. Algebraic attacks allow to cryptanalyse a large class of stream ciphers, satisfying all the previously known design criteria.To resist algebraic attacks, Boolean functions should achieve optimal AI. It was shown in [2, 3] that optimal AI of an n-variable Boolean function is $\lceil \frac{n}{2} \rceil$. Many researchers have been presented some constructions of Boolean functions with optimal AI [7, 8, 9].

Hang Liu
NELWS lab, Xi’an University of Posts & Telecommunications, Xi’an 710121, China

Dong Zheng
NELWS lab, Xi’an University of Posts & Telecommunications, Xi’an 710121, China
Westone Cryptologic Research Center, Beijing 100070, China

Qinglan Zhao
Shanghai Jiao Tong University, Shanghai 200240, China
NELWS lab, Xi’an University of Posts & Telecommunications, Xi’an 710121, China
e-mail: zhaoqinglan@foxmail.com

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_80

A famous method for constructing Boolean functions with optimal AI based on the generator matrix $G(k,n)$ of the kth-order Reed-Muller code $\mathrm{RM}(k,n)$ was given by Carlet in 2006[5]. In 2013, Su et al gave a further study of the linear relations of the column vectors in $G(k,n)$ studied in paper [11], and proposed some constructions of odd-variable Boolean functions . In reference [12], Su proposed a construction of even-variable Boolean functions. There are few constructions of Boolean functions based on the generator matrix of Reed Muller code with optimal algebraic immunity. In this paper, we construct new Boolean functions with optimal AI using the generator matrix $G(k,n)$ of Reed-Muller code $\mathrm{RM}(k,n)$, and also prove that they have high nonlinearity.

The rest of this paper is organized as follows. Some notions and preliminaries are described in Section 2. A new class of Boolean functions with optimal AI is presented in Section 3. The concluding remarks are given in Section 4.

2 Preliminaries

Let $\mathbb{F}_2^n$ be the n-dimensional vector space over the finite field $\mathbb{F}_2 = \{0,1\}$. Given a vector $\alpha = (a_1,a_2,\dots,a_n) \in \mathbb{F}_2^n$, we define its support $supp(\alpha)$ as the set $\{i|a_i = 1, 1 \le i \le n\}$, and its Hamming weight $wt(\alpha)$ as the cardinality of its support, i.e., $wt(\alpha) = |supp(\alpha)|$. From now on, we always assume $k = \lceil \frac{n}{2} \rceil - 1$ in this paper.

Constructing an n-variable Boolean function f with $wt(f) = s$ and optimal AI is equivalent to find out a nonsingular $s \times s$ sub-matrix of the generator matrix G of the kth-order Reed-Muller code $\mathrm{RM}(k,n)$, denoted by $G(k,n) = (a_{i,j})_{s\times 2^n} = (c_{\alpha_1}, c_{\alpha_2}, \dots, c_{\alpha_{2^n}})$, for $1 \le i \le s$, $1 \le j \le 2^n$, $s = \sum_{t=0}^{k} \binom{n}{t}$.

Lemma 1. *[11] For any vector $u \in \mathbb{F}_2^n$, such that $wt(u) = k + j, 1 \le j \le n-k$, we have*

$$c_u = \bigoplus_{i=0}^{k} a_i^{(j)} \left(\bigoplus_{\substack{\alpha \preceq u,\\ wt(\alpha)=k-i}} c_\alpha \right) \tag{1}$$

where $a_i^{(j)} \in \mathbb{F}_2, 0 \le i \le k$, which satisfies $a_0^{(j)} = 1$ and $a_i^{(j)} = 1 \oplus \bigoplus_{l=0}^{i-1} a_l^{(j)} \binom{i+j}{i-l}, 1 \le i \le k$.

Lemma 2. *[11, 12] For $1 \le j \le n-k$, let u be a vector in $\mathbb{F}_2^n$ with $wt(u) = k + j$. In the linear expression of c_u in Equation (1), the coefficients $a_i^{(j)}$ of c_α with $\alpha \preceq u$ and $wt(\alpha) = k-i$ satisfy $a_i^{(j)} = \binom{i+j-1}{i} (\bmod 2)$, for $0 \le i \le k$ and $1 \le j \le n-k$. Furthermore, the first s column vectors in $G(k,n)$ form a basis of the vector space $\mathbb{F}_2^s$. And any column vector in $G(k,n)$ can also be linearly expressed by the last s column vectors in $G(k,n)$.*

For an integer $1 \le l \le s$, choose two vector subsets $U = \{u_1,\dots,u_l\} \subseteq W^{\ge k+1}$ and $T = \{\beta_1,\dots,\beta_l\} \subseteq W^{\le k}$. Set $W^{\ge k+1} \backslash U = \{\gamma_1,\dots,\gamma_{s-l}\}$. Then, based on the basis $\{c_{u_1},\ldots,c_{u_l},c_{\gamma_1},\ldots,c_{\gamma_{s-l}}\}$, the submatrix $[c_{\beta_1},\ldots,c_{\beta_l},c_{\gamma_1},\ldots,c_{\gamma_{s-l}}]$ can be expressed as

$$[c_{u_1},\ldots,c_{u_l},c_{\gamma_1},\ldots,c_{\gamma_{s-l}}] = [c_{\beta_1},\ldots,c_{\beta_l},c_{\gamma_1},\ldots,c_{\gamma_{s-l}}]\begin{pmatrix} B & 0 \\ C & I \end{pmatrix} \tag{2}$$

where $B = (b_{i,j})$ is an $l \times l$ matrix, 0 is a zero matrix, and I is an identity matrix of order $s-l$.

Concerning a function $f \in \mathbb{B}_n$ with optimal AI, we have the following sufficient and necessary condition.

Proposition 1. *[4] A function $f \in \mathbb{B}_n$ has optimal AI if and only if both of the vector sets $R_f^{(1)}(k,n) = \{c_\alpha \in G(k,n)|\alpha \in supp(f)\}$, $R_f^{(0)}(k,n) = \{c_\alpha \in G(k,n)|\alpha \in zeros(f)\}$ have rank $\sum_{i=0}^{k}\binom{n}{i}$, where $k = \lceil n/2 \rceil - 1$.*

The important task is to properly choose two vector subsets $U = \{u_1,\ldots,u_l\} \subseteq W^{\geq k+1}$ and $T = \{\beta_1,\ldots,\beta_l\} \subseteq W^{\leq k}$, satisfying the following two conditions.

C1. The coefficient of c_{u_i} in the linear expression of c_{β_i} is 1, i.e., $b_{i,i} = 1$ for $1 \leq i \leq l$;

C2. The coefficient of c_{u_j} in the linear expression of c_{β_i} is 0, i.e., $b_{i,j} = 0$ for all $1 \leq j < i \leq l$, (or for all $1 \leq i < j \leq l$);

The simplest Boolean function with optimal AI is the so-called majority function

$$F(x) = \begin{cases} 1, & wt(x) \geq \lceil \frac{n}{2} \rceil \\ 0, & otherwise \end{cases} \tag{3}$$

which is proved that the function achieves optimal AI[1, 10].

3 Construction of Boolean functions

This section will construct balanced Boolean functions with optimal AI. We will propose a new construction of odd-variable function $f \in \mathbb{B}_n$ with $supp(f) = (W^{\geq k+1} \backslash U)\bigcup T$, where T and U are two properly chosen subsets of $W^{\leq k}$ and $W^{\geq k+1}$, respectively. And we will describe the sort of vectors in set T and U in detail.

3.1 Construction of Boolean functions on odd variables

From now on, we always assume that n is odd. Denote $m = \lceil \frac{n}{3} \rceil$.Then $n = 3m$ or $n = 3m+2$, if m is odd, or $n = 3m+1$, if m is even. Further, we always denote

$t = \left\lfloor \frac{m}{2} \right\rfloor$ and $p = \lceil log_2 t \rceil$. Set $\mathbb{F}_2^p = \left\{ e_1^{(p)}, e_2^{(p)}, \ldots, e_{2^p}^{(p)} \right\}$, where the vectors are listed according to the Hamming weight firstly and the lexicographic order secondly.

Note that, if $n = 3m$ with n odd, then m is odd. We define m subsets T_i of $\mathbb{F}_2^n$ in Table 1, $1 \leq i \leq m$.

Table 1 Vectors in T_i, for $n = 3m, 1 \leq i \leq m$

i	vectors in $T_i \subseteq W^k \bigcup W^{k-2}$
$[1,t]$	$(y_1, 0, y_2, 0, y_3, e_i^{(p)}, 0) \in \mathbb{F}_2^{3i-3} \times \mathbb{F}_2^3 \times \mathbb{F}_2^{3t-3i} \times \mathbb{F}_2 \times \mathbb{F}_2^{n-3t-p-2} \times \mathbb{F}_2^p \times \mathbb{F}_2$
$[t+1, \min\{2t, m\}]$	$(0, e_{i-t}^{(p)}, y_1, 0, y_2, 1) \in \mathbb{F}_2 \times \mathbb{F}_2^p \times \mathbb{F}_2^{3i-4-p} \times \mathbb{F}_2^3 \times \mathbb{F}_2^{n-3i-1} \times \mathbb{F}_2$
$[\min\{2t, m\}+1, m]$	$(1, e_{i-t}^{(p)}, y_1, 1, y_2, 0) \in \mathbb{F}_2 \times \mathbb{F}_2^p \times \mathbb{F}_2^{3t-1-p} \times \mathbb{F}_2 \times \mathbb{F}_2^{n-3t-3} \times \mathbb{F}_2^3$

Similarly, when $n = 3m + 2$ with m odd, we define m subsets T_i of $\mathbb{F}_2^n$ in Table 2, $1 \leq i \leq m$, where $\lambda = m + 3 - min\{2t, m\}$, with the exception of $T_{m+1} = \{\beta \in W^k | \beta = (1, e_\lambda^{(p)}, y_1, 1, y_2, 0) \in \mathbb{F}_2 \times \mathbb{F}_2^p \times \mathbb{F}_2^{3t-1-p} \times \mathbb{F}_2 \times \mathbb{F}_2^{3m-3t-1} \times \mathbb{F}_2^{n-3m}\}$.

Table 2 Vectors in T_i, for $n = 3m+2, 1 \leq i \leq m$

i	vectors in $T_i \subseteq W^k \bigcup W^{k-2}$
$[1,t]$	$(y_1, 0, y_2, 0, y_3, e_i^{(p)}, 0) \in \mathbb{F}_2^{3i-3} \times \mathbb{F}_2^3 \times \mathbb{F}_2^{3t-3i} \times \mathbb{F}_2 \times \mathbb{F}_2^{n-3t-p-2} \times \mathbb{F}_2^p \times \mathbb{F}_2$
$[t+1, \min\{2t, m\}]$	$(0, e_{i-t}^{(p)}, y_1, 0, y_2, 1) \in \mathbb{F}_2 \times \mathbb{F}_2^p \times \mathbb{F}_2^{3i-4-p} \times \mathbb{F}_2^3 \times \mathbb{F}_2^{n-3i-1} \times \mathbb{F}_2$
$[\min\{2t, m\}+1, m]$	$(1, e_{i-t}^{(p)}, y_1, 1, y_2, 0, y_3) \in \mathbb{F}_2 \times \mathbb{F}_2^p \times \mathbb{F}_2^{3t-1-p} \times \mathbb{F}_2 \times \mathbb{F}_2^{3i-4-3t} \times \mathbb{F}_2^3 \times \mathbb{F}_2^{n-3i}$

Similarly, when $n = 3m + 1$ with m even, we define m subsets T_i of $\mathbb{F}_2^n$ in Table 3, $1 \leq i \leq m + 1$, where λ is the same with the above definition, with the exception of $T_{m+1} = \{\beta \in W^k | \beta = (1, e_{2^p}^{(p)}, y_1, 1, y_2, 0) \in \mathbb{F}_2 \times \mathbb{F}_2^p \times \mathbb{F}_2^{3t-1-p} \times \mathbb{F}_2 \times \mathbb{F}_2^{3m-3t-1} \times \mathbb{F}_2^{n-3m}\}$.

Table 3 Vectors in T_i, for $n = 3m+1, 1 \leq i \leq m$

i	vectors in $T_i \subseteq W^k \bigcup W^{k-2}$
$[1,t]$	$(y_1, 0, y_2, 0, y_3, e_i^{(p)}, 0) \in \mathbb{F}_2^{3i-3} \times \mathbb{F}_2^3 \times \mathbb{F}_2^{3t-3i} \times \mathbb{F}_2 \times \mathbb{F}_2^{n-3t-p-2} \times \mathbb{F}_2^p \times \mathbb{F}_2$
$[t+1, m]$	$(0, e_{i-t}^{(p)}, y_1, 0, y_2, 1) \in \mathbb{F}_2 \times \mathbb{F}_2^p \times \mathbb{F}_2^{3i-4-p} \times \mathbb{F}_2^3 \times \mathbb{F}_2^{n-3i-1} \times \mathbb{F}_2$

And then, we also need to define $|T|$ subsets $U_i \subseteq \mathbb{F}_2^n$ as $U_i = \{\beta + \beta_0 | \beta \in T_i, \beta_0 = (0, 1, 0) \in \mathbb{F}_2^{3i-3} \times \mathbb{F}_2^3 \times \mathbb{F}_2^{n-3i}\}$, for $1 \leq i \leq |T|$.

Now, we describe vectors in T and U for three odd numbers, as follows.

Example 1. For $n = 19$, 21 and 23, some specific elements in T_i and U_i are illustrated in Tables 4, 5, 6.

Table 4 This is a table about specific elements in T_i and U_i, for $n = 19$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
T_1	0	0	0							0							0	0	0
U_1	1	1	1							0							0	0	0
T_2				0	0	0				0							1	0	0
U_2				1	1	1				0							1	0	0
T_3							0	0	0	0							0	1	0
U_3							1	1	1	0							0	1	0
T_4	0	0	0							0	0	0							1
U_4	0	0	0							1	1	1							1
T_5	0	1	0										0	0	0				1
U_5	0	1	0										1	1	1				1
T_6	0	0	1													0	0	0	1
U_6	0	0	1													1	1	1	1
T_7	1	1	1							1									0
U_7	1	1	1							1									1

Table 5 This is a table about specific elements in T_i and U_i, for $n = 21$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
T_1	0	0	0							0									0	0	0
U_1	1	1	1							0									0	0	0
T_2				0	0	0				0									1	0	0
U_2				1	1	1				0									1	0	0
T_3							0	0	0	0									0	1	0
U_3							1	1	1	0									0	1	0
T_4	0	0	0							0	0	0									1
U_4	0	0	0							1	1	1									1
T_5	0	1	0										0	0	0						1
U_5	0	1	0										1	1	1						1
T_6	0	0	1													0	0	0			1
U_6	0	0	1													1	1	1			1
T_7	1	1	1							1									0	0	0
U_7	1	1	1							1									1	1	1

Table 6 This is a table about specific elements in T_i and U_i, for $n = 23$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
T_1	0	0	0							0											0	0	0
U_1	1	1	1							0											0	0	0
T_2				0	0	0				0											1	0	0
U_2				1	1	1				0											1	0	0
T_3							0	0	0	0											0	1	0
U_3							1	1	1	0											0	1	0
T_4	0	0	0								0	0	0										1
U_4	0	0	0								1	1	1										1
T_5	0	1	0											0	0	0							1
U_5	0	1	0											1	1	1							1
T_6	0	0	1														0	0	0				1
U_6	0	0	1														1	1	1				1
T_7	1	1	1							1									0	0	0		
U_7	1	1	1							1									1	1	1		
T_8	1	1	1							1												0	0
U_8	1	1	1							1												1	1

Based on the subsets T_i and U_i, set

$$T = \bigcup_{i=1}^{|T|} T_i \text{ and } U = \bigcup_{i=1}^{|T|} U_i. \tag{4}$$

Now, we give a new construction of Boolean functions with optimal AI. With T and U being subsets of $\mathbb{F}_2^n$ given by Equation (4), define $f \in \mathbb{B}_n$ as

$$f(x) = \begin{cases} F(x)+1, & x \in T \bigcup U \\ F(x), & otherwise \end{cases}. \tag{5}$$

where $F(x)$ is the majority function on n variables.

3.2 Analysis of Boolean function f

For convenience, we respectively arrange all vectors in T_i and U_i, $1 \le i \le |T|$, according to the Hamming weight firstly and the lexicographic order secondly. Suppose $T_i = \{\beta_1^{(i)}, \beta_2^{(i)}, \ldots, \beta_{|T_i|}^{(i)}\}$, $U_i = \{u_1^{(i)}, u_2^{(i)}, \ldots, u_{|T_i|}^{(i)}\}$, for $1 \le i \le |T|$. By the definition of T_i and U_i, we can know that any the jth vector in set T_i is said to be covered by the jth vector in set U_i. All entries of the jth vector in set T_i are less than or equal to all the entries in the jth vector in set U_i according to the order.

First, we compute and prove coefficients of Boolean function f satisfying Condition C1. It follows that any the jth vector in set T_i is said to be covered by the jth vector in set U_i, for $1 \le i \le |T|$, $1 \le j \le 2^n$. If $wt(\beta_j^{(i)}) = k - j'$, with $j' = 0, 2,$

then $wt(u_j^{(i)}) = k+3-j'$, for $1 \le i \le |T_i|$, $1 \le i \le m$. From Lemma 2, we know that the corresponding coefficient $b_{j,j}$ in Equation (2) is $b_{j,j} = a_{j'}^{(3-j')} = \binom{3-j'+j'-1}{j'} = 1(\bmod 2)$, $j' = 0, 2$. So, the vectors in T_i and U_i, $1 \le i \le m$, satisfy Condition C1. And there are 2 cases about $i = m+1$ as following:

Case 1: When $n = 3m+2$, m is odd, $wt(\beta_j^{(m+1)}) = k$, $wt(u_j^{(m+1)}) = k+1$, $1 \le j \le |T_{m+1}|$. We know that $b_{j,j} = a_0^{(1)} = \binom{0+1-1}{0} = 1(\bmod 2)$.

Case 2: When $n = 3m+1$, m is even, $wt(\beta_j^{(m+1)}) = k$, $wt(u_j^{(m+1)}) = k+1$, $1 \le j \le |T_{m+1}|$. We know that $b_{j,j} = a_0^{(1)} = \binom{0+1-1}{0} = 1(\bmod 2)$.

Hence, the vectors in T_i and U_i, $1 \le i \le |T|$, satisfy Condition C1. On the other hand, we know $\beta_{j_2}^{(i)} \not\preceq \beta_{j_1}^{(i)}$, which implies $\beta_{j_2}^{(i)} \not\preceq u_{j_1}^{(i)}$, for $1 \le i \le |T|$ and $1 \le j_1 < j_2 \le |T_i|$, since all $\beta_j^{(i)}$ are arranged according to the Hamming weight firstly and the lexicographic order secondly, for $1 \le i \le |T|$, $1 \le j \le 2^n$. Furthermore, for $1 \le i_1 < i_2 \le m$, $1 \le j_1 \le |T_{i_1}|$ and $1 \le j_2 \le |T_{i_2}|$, we know $\beta_{j_2}^{(i_2)} \not\preceq \beta_{j_1}^{(i_1)}$ by the $e_i^{(p)}$'s and the first, the $(3t+1)$th and the last entries of the vectors in T_{i_1} and T_{i_2}, which implies $\beta_{j_2}^{(i_2)} \not\preceq u_{j_1}^{(i_1)}$. Hence, we have $b_{i,j} = 0$, $i > j$. In short, the vectors in T_i and U_i, $1 \le i \le m$, satisfy Condition C1 and Condition C2, which implies that the vector set $\{c_\alpha | \alpha \in (W^{\ge k+1} \backslash U) \bigcup T\}$ has rank s.

We compute the nonlinearity of Boolean functions f given in Equation (5). The nonlinearity of Boolean function f can be expressed according to Walsh spectrum as $nl_f = 2^{n-1} - \frac{1}{2} \max_{\omega \in \mathbb{F}_2^n} |W_f(\omega)|$. The Walsh spectrum of f is less than that of majority function F. Therefore, we can obtain the nonlinearity of f as follows.

Theorem 1. *For $n \ge 11$ being odd, the nonlinearity of $f \in \mathbb{B}_n$ constructed in Equation (5) is $nl_f = 2^{2k} - \binom{2k}{k} + 2\left[\binom{2k-4-p}{k} + \binom{2k-4-p}{k-2}\right]$, where $p = \lceil log_2 \lceil \frac{m}{2} \rceil \rceil$, $m = \lfloor \frac{n}{3} \rfloor$ and $k = \frac{n-1}{2}$.*

For $11 \le n \le 21$, with n odd, the comparison of our function with nonlinearity of the majority function, and nonlinearity of the function [11] as table 7. It should be noted that the nonlinearity of our function is as good as that of function[11] for $n = 13, 15, 17$ and 19.

Table 7 Comparison of the nonlinearity for $11 \le n \le 21$ with n odd

n	11	13	15	17	19	21
nonlinearity of the function in reference[11]	824	3256	13276	53920	218386	882696
nonlinearity of the function in this paper	794	3256	13276	53920	218386	871828

4 Conclusion

In this paper, we proposed a new construction of Boolean functions with optimal algebraic degree and high nonlinearity. In fact, the nonlinearity of our function is not higher than some constructions which are not based on the generator matrix of Reed-Muller code. Nevertheless, the most useful properties of Boolean function based on the generator matrix of Reed-Muller code are the efficient computation and easy implementation. In addition, there are some aspects for further study, such as how to define the algebraic degree and how to give a rigorous proof of our function on the behavior against fast algebraic attack.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant Nos. 61472472, 61272037 and 61402366) and the Natural Science Basic Research Plan in Shaanxi Province of China (Grant Nos. 2016JM6033).

References

1. Ding C, Xiao G, Shan W. *The Stability Theory of Stream Ciphers*[M]. Springer Berlin Heidelberg, 1991.
2. Nicolas T. Courtois, Willi Meier. *Algebraic Attacks on Stream Ciphers with Linear Feedback.*[C]. Advances in Cryptology - EUROCRYPT 2003, International Conference on the Theory and Applications of Cryptographic Techniques, Warsaw, Poland, May 4-8, 2003, Proceedings. 2003:345-359.
3. Nicolas T. Courtois, Willi Meier. *Fast Algebraic Attacks on Stream Ciphers with Linear Feedback.*[C]. Advances in Cryptology - EUROCRYPT 2003, Lecture Notes in Computer Science. 2003:176-194.
4. Carlet C, Gaborit P. *On the construction of balanced boolean functions with a good algebraic immunity.*[C]// Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on. 2005:1101-1105.
5. Carlet C. *A Method of Construction of Balanced Functions with Optimum Algebraic Immunity*[J]. Iacr Cryptology Eprint Archive, 2006, 2006(2):131-6.
6. Carlet C. *Vectorial Boolean functions for cryptography*[J]. Boolean Models & Methods in Mathematics, 2006.
7. Carlet C. Constructing balanced functions with optimum algebraic immunity[C]// IEEE International Symposium on Information Theory. 2007:451-455.
8. Carlet C, Feng K. An Infinite Class of Balanced Functions with Optimal Algebraic Immunity, Good Immunity to Fast Algebraic Attacks and Good Nonlinearity.[C]// Advances in Cryptology - ASIACRYPT 2008, International Conference on the Theory and Application of Cryptology and Information Security, Melbourne, Australia, December 7-11, 2008. Proceedings. 2008:425-440.
9. Carlet C, Zeng X, Li C, et al. Further properties of several classes of Boolean functions with optimum algebraic immunity[J]. Designs Codes & Cryptography, 2009, 52(3):303-338.
10. Dalai D K, Maitra S, Sarkar S. Basic Theory in Construction of Boolean Functions with Maximum Possible Annihilator Immunity[J]. Designs Codes & Cryptography, 2006, 40(1):41-58.
11. Su S, Tang X, Zeng X. *A systematic method of constructing Boolean functions with optimal algebraic immunity based on the generator matrix of the Reed-Muller code*[J]. Designs Codes & Cryptography, 2014, 72(3):653-673.
12. Su S. Construction of balanced even-variable Boolean functions with optimal algebraic immunity[J]. International Journal of Computer Mathematics, 2014, 92(11):1-14.

Location-Sensitive Data Sharing in Mobile Cloud Computing

Zhiwei Zhang, Yunling Wang, Jianfeng Wang, Xiaofeng Chen, and Jianfeng Ma

Abstract Mobile could computing (MCC) enhances computation and storage capabilities of mobile devices by leveraging services in mobile clouds, and more and more mobile users tend to outsource their data to clouds. According to laws or user demands, some data can only be accessed by users locating at specified regions, which requires location-based access control mechanisms. There are mainly two techniques involved: secure location and location verification. Distance Bounding Protocols (DBP) is a foundation for secure location and location verification, however, existing studies require strict time synchronization which is a difficult problem in itself. In this paper, we propose a novel location verification protocol that can verify a mobile user's position more accurately and efficiently without time synchronization. Furthermore, we design a secure location-sensitive data sharing scheme based on our location protocol in mobile cloud computing.

Zhiwei Zhang
State Key Laboratory of Integrated Service Networks, Xidian University, China e-mail: zwzhang@xidian.edu.cn

Yunling Wang
State Key Laboratory of Integrated Service Networks, Xidian University, China e-mail: ylwang0304@163.com

Jianfeng Wang
State Key Laboratory of Integrated Service Networks, Xidian University, China e-mail: wjf01@163.com

Xiaofeng Chen
State Key Laboratory of Integrated Service Networks, Xidian University, China e-mail: xfchen@xidian.edu.cn

Jianfeng Ma
State Key Laboratory of Integrated Service Networks, Xidian University, China e-mail: jfma@mail.xidian.edu.cn

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_81

1 Introduction

Wireless communication technologies make the concept of ABC (Always Best Connected) a reality, and mobile cloud computing integrates cloud computing into mobile environment. Nowadays, more and more mobile applications enable data owners to outsource their data to the cloud and enjoy fascinating advantages brought by cloud computing.

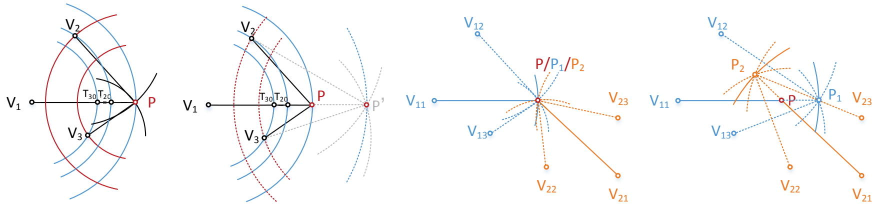
However, data outsourcing raises many new security problems. A trivial method to protect data is to encrypt data and controlling access. Traditional authentication and access control usually base on the following attributes: what you know (eg. a password), what you have (eg. a smart card), and who you are (eg. a fingerprint). But there is a lack of consideration of where you are [2], which is valuable in a number of applications. For examples: (1) in the field of e-health or m-health, electronic medical records should require to be accessed by the authorized doctors who are staying in the hospitals. (2) There are three types of physical threats to mobile equipments: lending, loss, and theft, all of which raise the possibility of enabling unauthorized persons to access to the outsourced data with a legal device. If users' positions are verified, these threats to cloud storage can be removed.

Peterson et al. defined data sovereignty in [7], and emphasized the neessarity for developing algorithms for establishing the integrity, authenticity, and geographical location of data stored in the cloud [3]. In addition, how to verify and guarantee data usage position has also been widely investigated. There are mainly two techniques involved: secure localization and location verification [2], [4], and the basic block for these works is DBP protocol [1]. Specifically, in [2], Chandran et al. proved a strong impossibility result on achieving secure location in the Vanilla model, and gave the bounded-storage framework to study the foundations of position-based cryptography.

In this paper, we focus on accurately verifying mobile user's location in 2-dimensional space (it can be extended to 3-dimensional space), and securely sharing outsourced data to other mobile cloud users whose positions are specified. There are two contributions: (1) our location verification protocol can verify user's position with no more than two verifiers executing DBP protocols, which overcomes the difficulty of time synchronization in the existing location verification methods. (2) we design an SLDS scheme based on our location protocol. SLDS achieves location-sensitive data sharing in mobile cloud computing, and access-related key management as well as symmetrical encryption re-keying with all-or-nothing transformation (AONT) [8].

2 Deterministic Location Using Single Verifier

We present a novel location verification method that can work with all existing DBP protocols. Specifically, to locate a wireless node, our method needs at least one master verifier to execute a DBP protocol. Moreover, our method can also locate



(a) Correct 1rd (b) Lengthen 1rd (c) Correct 2rd (d) Lengthen 2rd

Fig. 1 One round location and two round location

the real position when a malicious prover try to lengthen the distance between the master verifier and it.

2.1 One Round Location

We use V and P to denote a verifier and a prover in DBP protocols. Our one round location module gives out P's position with one master verifier V_1 and two slave verifiers V_2, V_3 , as shown in Fig.1. The master verifier runs DBP protocol with the prover to get the distance between them, and calculates distances between slave verifiers and prover. Slave verifiers listen communications between the master verifier and the prover, record time durations between the master verifier's challenge and the prover's response. Note that slave verifiers are randomly chosen by master verifier.

Take Fig.1(a) for example, the master verifier V_1 knows positions of itself, V_2 and V_3, so it knows the length of $\overline{V_1V_2}$, $\overline{V_1V_3}$, the prover P should locate at the conjoint point of three circle whose centers are V_1, V_2, V_3 and radiuses are $\overline{V_1P}$, $\overline{V_2P}$, $\overline{V_3P}$.So, the problem of locating P is transformed into calculating $\overline{V_1P}$, $\overline{V_2P}$, $\overline{V_3P}$.

At the time of t_{20} and t_{30}, V_2 and V_3 detect V_1's challenge of DBP quick exchanges. Note that, at these moment, signal can be also detected at point T_{20} and T_{30} which should be on the line V_1P. In other words, $\overline{V_1T_{20}} = \overline{V_1V_2}$ and $\overline{V_1T_{30}} = \overline{V_1V_3}$. And at time t_{21} and t_{31}, P's response signal is detected, then $\Delta t_2 = t_{21} - t_{20}$ and $\Delta t_3 = t_{31} - t_{30}$ are the time durations that electromagnetic waves travel from T_{20} and T_{30} to P and from P to V_2 and V_3 (P's process time is omitted in DBPs). Because speed of electromagnetic wave is a constant v, $\Delta t_2 v$ and $\Delta t_3 v$ can be calculated:

$$\begin{cases} \Delta t_2 v = \overline{T_{20}P} + \overline{PV_2} = (\overline{V_1P} - \overline{V_1T_{20}}) + \overline{V_2P} \\ \Delta t_3 v = \overline{T_{30}P} + \overline{PV_3} = (\overline{V_1P} - \overline{V_1T_{30}}) + \overline{V_3P} \end{cases} \Rightarrow \begin{matrix} \overline{V_2P} = \Delta t_2 v - (\overline{V_1P} - \overline{V_1V_2}) \\ \overline{V_3P} = \Delta t_3 v - (\overline{V_1P} - \overline{V_1V_3}) \end{matrix},$$

where $\overline{V_1P}$ is the result of DBP protocol, $\overline{V_2P}$ and $\overline{V_3P}$ can be figured out, so far the problem of locating P is solved.

2.2 Two Rounds Location

Our one round location protocol can locate a prover's position by only one master verifier, while the existing location protocols need at least three master verifiers and request strict time synchronization. However, our one round location will be fail when provers are not honest who try to lengthen or shorten distances on purpose.

It's easy to lengthen the distance (an adversary prover can add an extra delay), see Fig.1(b), but it's not easy to detect this type of distance attack for the exiting DBPs. In other hand, to shorten the distance, there must be adversaries colluding with the prover, and adversaries replace prover to send responses to verifier to make the verifier believe that they were the real prover. However, this attack has been studied by many works, so we focus on the first mentioned attack.

One round location gives the prover's suggested position which must be on the line V_1P (P is the prover's real position). This ability is beyond of any existing protocols. Because one prover can not present at two places at the same time and an adversary can not have all of the prover's information, we can execute one round location protocol twice by two master verifiers. Specifically, if the prover presents a distance attack to the master verifiers V_{11} and V_{21}, two wrong positions P_1 and P_2 of the prover will be given out. However, the prover's real position P should be on both lines $V_{11}P_1$ and $V_{21}P_2$, namely that the intersection of lines $V_{11}P_1$ and $V_{21}P_2$ is prover's real position P. The relationships of P, P_1 and P_2 is shown in Fig.1(c) and Fig.1(d). In this way, our two rounds location protocol can always give the position of the prover or a judgement of location attack.

3 SLDS: Secure Location-Sensitive Data Sharing Scheme

In this section, we present the SLDS (Secure Location-Sensitive Data Sharing) scheme. SLDS is designed to ensure that only users whose identities and positions are both authenticated can access to and decrypt the outsourced data, and protect the outsourced data with a modified AONT encryption which enables outsourced data to be re-encrypted by new keys for later users. SLDS consists of **User Center**, **Data Owner**, **Data User**, **Storage Cloud**, and **Location Cloud** (including **Location Verifiers** belonging to it).

The main goal of our SLDS scheme is to guarantee that only permitted data users can access to and decrypt data outsourced by data owner if their positions are located in specified regions. There are three security assumptions: (1) The storage cloud is honest-but-curious. storage cloud holds uploaded data safely and reliably, and it honestly follows the proposed protocol. However, it wants to access to outsourced data without permission. (2) The user center and location cloud are trusted. The user center and location cloud are trusted by all data owner and data users. (3) The adversaries can obtain part of key parameters. If adversaries have all of the users' information, adversaries and users will be indistinguishable [2].

3.1 Scheme Framework

3.1.1 Initialization Stage

For every set of data (denoted as *DS*), data owner and user center prepare key parameters and access policies for system participants.

Encryption Keys. Data owner chooses one symmetric key *DK* to encrypt *DS* (result is noted as *EDS*) and one symmetric key SK_0 to encrypt a randomly selected subsegment of all-or-nothing transformed *EDS*.

User List and Position Policy. For each data set to be shared with other users, data owner tells user center who can access to it by a list: $UserList_{DS} = \{UID_1, UID_2, \cdots\}$. Data owner generates a data access position policy which clearly and unambiguously specify for *DS* where it can or can not be accessed to: $PosPolicy_{DS} = \{Action@(Modifier)LocationSet\}$, in which *Action* can be *Reject* or *Accept*, *Modifier* can be *None* or *NOT*, and *LocationSet* is a positions set.

User Management Polynomial Coefficient Matrix. User center generates a $(n+1) \times (n+1)$ matrix $PCM = [\alpha_{ij}]$, where $\alpha_{ij} \in_R \{0,1\}^\lambda$ $(0 \le i, j \le n)$ and n decides upper of user number (n is suggested to be the length of AES's key or 128 fixed). *PCM* will be used to compute other key parameters, *PCM* can be changed or unchanged for different data sets or owners depending on operators' policy.

3.1.2 Configuration Stage

In this stage, more parameters will be computed, and the data users, the storage cloud as well as the location cloud will be configured with these parameters.

Data encryption and AON transformation. There are two operations on owner outsourced data *DS*. The first is data encryption, our SLDS uses AES to encrypt *DS* with *DK*, $EDS = AES(DS, DK)$ is the ciphertext of *DS*. The second is AON transformation, we use a AON scheme which is similar with the AON scheme in [6] to transform *EDS* into $TDS = AONT(EDS) = C||S$. The AON transformation ensures that users must have both *C* and *S* to recover the *EDS*, otherwise, they can obtain nothing about *EDS*, and it makes our symmetric re-encryption possible.

Data center and users configuration. For each *UID* in the $UserList_{DS}$, user center uses a hash function $hash(\cdot)$ to compute $H = hash(hash(C)||UID||x||RK_{DS})$, where $hash(C)$ is provided by the data owner, unique x is selected randomly, and RK_{DS} is a unchanged for all *DS*'s users selected randomly by data center. Then, the data center computes $y = |f(x)|$ for each user as follows,

$$f(x) = \begin{bmatrix} h_0 & 0 & \cdots & 0 \\ 0 & h_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_n \end{bmatrix} \cdot \begin{bmatrix} \alpha_{00} & \alpha_{01} & \cdots & \alpha_{0n} \\ \alpha_{10} & \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{n0} & \alpha_{n1} & \cdots & \alpha_{nn} \end{bmatrix} \cdot \begin{bmatrix} x^0 \\ x^1 \\ \vdots \\ x^n \end{bmatrix} \mod p, \quad (1)$$

where h_i is the $i-$th bit (0 or 1) of H, the middle matrix is the PCM, and p is a large prime number (the length of p should be at least equal to or bigger than the length of AES's key). Data users who are in the $UserList_{DS}$ can request their own (x,y) pairs of DS from the data center. Note that x is used as user's pseudo name and can be public, while y is used to compute keys and should be kept secret.

Storage cloud configuration. The data user uploads C of TDS to storage cloud directly and sends $hash(C)$ to user center. Before uploaded to storage cloud, S of TDS should be encrypted to S' with AES and a randomly key SK_0 by data owner. In order to improve the efficiency of re-encryption, we can just encrypted a small part of S instead of entirety of it [5]. Besides, a sub-matrix PCM_{sc} consisting of all odd columns of PCM as well as modulus p are sent to storage cloud by the user center (n is generally considered to be a even number).

Location Cloud Configuration. There are two pieces of information are needed, they are position policy $PosPolicy_{DS}$ from the data owner and sub-matrix PCM_{lc} consisting of all non-zero even columns of PCM and modulus p from the data center.

3.1.3 Data Access Stage

In this stage, the data user will get C, S' and keys to decrypt, otherwise data access request will be denied. A correct process of this stage is given as following:

Storage cloud authentication. The data user sends pseudo name x to storage cloud, storage cloud computes $Y_{sc} = PCM_{sc} \cdot X_{odd} = \left[y_1, y_3, \cdots, y_{(n-1)}\right]^T \mod p$, and sends (x, Y_{sc}) to user center. User center checks x, and computes $y_{sc} = H_{odd} \cdot Y_{sc} = \left[h_1, h_3, \cdots, h_{(n-1)}\right] \cdot \left[y_1, y_3, \cdots, y_{(n-1)}\right]^T \mod p$, where h_i is the i-th bit of H, then user center re-encrypts S from S' to $NewS'$ with a SK_{i+1}. User center keeps these SKs in a list whose newest node is indexed by $SKIndicator$. Finally, $(x, y_{sc} \oplus y)||SKIndex||CSIndicator||NewS'$ is sent to storage cloud, where $SKIndex = SKIndicator$ denotes the index of SK_i, $CSIndicator$ tells the storage cloud which C and S' should be used and $NewS'$ tells storage cloud to replace S' with $NewS'$. Besides, the user center makes $SKIndicator$ point to SK_{i+1}, and sets $SKIndex$ timer going. The storage cloud then responses to data user with $(y_{sc} \oplus y)||C||S'||SKIndex$.

Location cloud authentication. After the storage cloud authentication, the data user starts interacting with the location cloud. The data user sends $x||SKIndex$ to the location cloud, and the location cloud forwards it to the user center. The user center checks x and $SKIndex$, especially freshness of $SKIndex$, and then sends $PosPolicyIndicator$ to the location cloud to start location protocol. The location cloud executes our two rounds location protocol to verify the data user's position under $PosPolicy$ indicated by $PosPolicyIndicator$. If location result is ok, location cloud computes $Y_{lc} = PCM_{lc} \cdot X_{even} = \left[y_2, y_4, \cdots, y_n\right]^T \mod p$, and sends $(x, Y_{lc})||SKIndex$ to user center. User center checks x and $SKIndex$ again, and computes $y_{lc} = H_{even} \cdot Y_{lc} = \left[h_2, h_4, \cdots, h_n\right] \cdot \left[y_2, y_4, \cdots, y_n\right]^T \mod p$, where h_i is the i-th bit of H, and sends $(x, y_{lc} \oplus y)||(DK \oplus y)||(SK_{index} \oplus \alpha_0)$ to location cloud, where

$\alpha_0 = [h_0, h_1, \cdots, h_n] \cdot [\alpha_{00}, \alpha_{10}, \cdots, \alpha_{n0}]^T \mod p = y - y_{sc} - y_{lc} \mod p$. The location cloud finally sends $(y_{lc} \oplus y)||(DK \oplus y)||(SK_{index} \oplus \alpha_0)$ to the data user.

3.1.4 Decryption Stage

After the identity and position have been both authenticated, the data user recovers y_{sc}, y_{lc} and DK from $y_{sc} \oplus y$, $y_{lc} \oplus y$ and $DK \oplus y$ with local y, then computes $\alpha_0 = y - y_{sc} - y_{lc}$ and recovers SK_{index} from $SK_i \oplus \alpha_0$. Then, the data user gets S by decrypting S' with SK_{index}, and transforms $C||S$ to EDS by reverse AONT [6]. Finally, the data user can get plaintext DS by decrypting ciphertext EDS with DK.

4 Conclusion

Location is being accepted as one of security factors, in this paper, we propose a novel location verification protocol which avoids the time synchronization problem in exciting methods, and we design the SLDS scheme which enables the data owner to share data with valid data users whose positions are securely verified.

Acknowledgements This work is supported by the Doctoral Fund of Ministry of Education of China (No. 20130203110004) and the National Natural Science Foundation of China (No. 61272455).

References

1. Brands S., Chaum D.: Distance-Bounding Protocols. In: EUROCRYPT ’93 Workshop on the theory and application of cryptographic techniques on Advances in cryptology, 344-359 (1994)
2. Chandran N., Goyal V., Moriarty R., Ostrovsky R.: Position Based Cryptography. In: International Cryptology Conference on Advances in Cryptology, 391-407 (2009)
3. Gondree M., Peterson Z.N.J.: Geolocation of data in the cloud. In: ACM Conference on Data and Application Security and Privacy, 25-36 (2013)
4. Gungor O., Chen F., Koksal C.E.: Secret Key Generation Via Localization and Mobility. IEEE Transactions on Vehicular Technology 64(6), 2214-2230 (2014)
5. Peterson Z.N.J., Burns R., Herring J., Stubblefield A., Rubin A.D.: Rekeying for Encrypted Deduplication Storage. In: IEEE/IFIP International Conference on Dependable Systems and Networks, (2016)
6. Peterson Z.N.J., Burns R., Herring J., Stubblefield A., Rubin A.D.: Secure Deletion for a Versioning File System. In: FAST ’05 Conference on File and Storage Technologies, 143-154 (2005)
7. Peterson Z.N.J., Gondree M., Beverly R.: A Position Paper on Data Sovereignty: The Importance of Geolocating Data in the Cloud. In: Usenix Workshop on Hot Topics in Cloud Computing, (2011)
8. Rivest R.L.: All-or-nothing encryption and the package transform. In: FSE ’97 Proceedings of the 4th International Workshop on Fast Software Encryption, 210-218 (1998)

Efficient and Expressive Anonymous Attribute-Based Encryption for Mobile Cloud Computing

Yinghui Zhang and Dong Zheng

Abstract As a kind of attribute-based encryption, ciphertext-policy attribute-based encryption (CP-ABE) is a potential technique for realizing fine-grained access control on shared data. However, traditional CP-ABE is not suitable for mobile cloud computing, where mobile users are resource-limited and privacy is fragile. In this paper, we propose an anonymous CP-ABE scheme supporting offline key generation and offline encryption. In the proposed scheme, sensitive attribute values specified in an access structure are not explicitly sent along with a ciphertext. The online/offline encryption mechanism alleviates the computational burden of mobile users by performing most of encryption tasks without draining the battery. In addition, the online/offline key generation mechanism allows the attribute authority to finish most of operations in the key generation process in advance, which enables efficient mobile user registration. Finally, the proposed scheme is proven fully secure in the standard model and the performance analysis shows its effectiveness in mobile cloud computing.

Yinghui Zhang
National Engineering Laboratory for Wireless Security, Xi'an University of Posts and Telecommunications, Xi'an 710121, P.R. China;
State Key Laboratory of Cryptology, P.O. Box 5159, Beijing 100878, P.R. China;
Westone Cryptologic Research Center, Beijing 100070, P.R. China.
e-mail: yhzhaang@163.com

Dong Zheng
National Engineering Laboratory for Wireless Security, Xi'an University of Posts and Telecommunications, Xi'an 710121, P.R. China;
Westone Cryptologic Research Center, Beijing 100070, P.R. China.
e-mail: zhengdong@xupt.edu.cn

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_82

1 Introduction

Towards cloud computing security, a promising public key primitive, attribute-based encryption (ABE), can be adopted. The concept of ABE was proposed by Sahai and Waters [21], in which scalable and fine-grained access rights can be assigned to individual users. ABE comes into two categories [10]: Key-Policy ABE (KP-ABE) and Ciphertext-Policy ABE (CP-ABE). CP-ABE is more suitable for realizing outsourced data security in cloud computing in that it puts access decisions in the hands of data owners. However, traditional CP-ABE schemes cannot be directly used in mobile cloud computing environment where security and efficiency requirements are more higher. In fact, traditional CP-ABE schemes cannot preserver users' attribute privacy because the sensitive access structure is sent along with ciphertexts explicitly. Besides, the key generation phase, the encryption phase and the decryption phase involve a large number of computation tasks. To the best of the authors' knowledge, most of existing CP-ABE schemes either suffer privacy disclosure or bad efficiency.

1.1 Our Contributions

We propose an anonymous CP-ABE scheme supporting offline key generation and offline encryption. In the proposed scheme, the computation tasks required in the key generation process and the encryption phase are split into an offline phase and an online phase. In the offline phase, the attribute authority can finish the majority of the work to issue attribute secret keys before knowing users' attributes. The mobile data owner does most of the computation tasks in encryption without needing the message and the access structure. Furthermore, the online phase can easily assemble the final secret key and ciphertexts once related specifications become known. In particular, the proposed scheme preserve users' attribute privacy by hiding the attribute values specified in the access structure in ciphertexts. Our scheme is proven fully secure in the standard model.

1.2 Related Work

Since the introduction of ABE [21], a plenty of researches have been done on various ABE schemes. In [10], Goyal *et al.* [10] presented a KP-ABE scheme by generating the private key according to the monotonic access structures. The first CP-ABE scheme was proposed by Bethencourt *et al.* [2], which is proven secure in the generic group model. To improve the security proof, Cheung and Newport [7] proposed another CP-ABE construction and proved its security in the standard model.

Although ABE can be directly adopted to enable secure data sharing, there is an increasing need to preserve users' attribute privacy in mobile cloud computing environment. In order to tackle this issue, anonymous ABE was introduced in [14]. The CP-ABE scheme [14] can realize hidden AND gate policies with positive and negative attributes, but it is not collusion-resistant. Based on the technique of hidden vector encryption, Boneh and Waters [4] proposed a predicate encryption scheme, which can realize anonymous CP-ABE by using the opposite semantics of subset predicates. An inner product predicate encryption scheme was presented by Katz et al. [15]. Based on this predicate scheme, we can achieve hidden CP-ABE schemes. A more efficient anonymous CP-ABE scheme was constructed in [18]. The security was based on the decisional bilinear Diffie-Hellman assumption and the decision linear assumption. Li et al. [17] proposed an accountable anonymous CP-ABE scheme. To achieve full security and expressiveness, Lai et al. [16] proposed an anonymous CP-ABE scheme under new assumptions. There are many other researches on anonymous ABE [23, 20, 13, 25, 19, 24].

To improve the efficiency of ABE, online/offline ABE schemes have recently been presented in [12]. The idea of online/offline was initiated by Even *et al.* [9] for digital signatures. An online/offline signature scheme consists of two phases and it can efficiently enables handover authentication in wireless networks [22]. To solve the key exposure problem, a special double-trapdoor hash family was proposed by Chen *et al.* [5], and they applied the hash-sign-switch paradigm to propose a much more efficient generic online/offline signature scheme [6]. The technique of online/offline encryption was introduced by Guo *et al.* [11]. The first fully secure online/offline predicate encryption and attribute-based encryption schemes have recently been presented by Datta *et al.* [8]. Constant-size ABE [26] has also been used in mobile cloud computing.

2 Preliminaries

2.1 Cryptographic Background

Definition 1 (Composite Order Bilinear Groups). Composite order bilinear groups are widely used in IBE and ABE systems [3]. We choose $(p_1, p_2, p_3, p_4, \mathbb{G}, \mathbb{G}_T, \hat{e})$, where p_1, p_2, p_3, p_4 are distinct primes, $\mathbb{G}$ and $\mathbb{G}_T$ are two cyclic groups of order $N = p_1 p_2 p_3 p_4$, and $\hat{e} : \mathbb{G} \times \mathbb{G} \to \mathbb{G}_T$ is a bilinear map satisfying: Bilinear: $\hat{e}(g^a, h^b) = \hat{e}(g,h)^{ab}$ for all $a, b \in \mathbb{Z}_N$ and $g, h \in \mathbb{G}$. Non-degenerate: there exists $g \in \mathbb{G}$ such that $\hat{e}(g,g)$ has order N in $\mathbb{G}_T$. Assume that group operations in $\mathbb{G}$ and $\mathbb{G}_T$ as well as the bilinear map $\hat{e}$ are computable in polynomial time with respect to λ. Let $\mathbb{G}_{p_i}$ be the subgroup of order p_i in $\mathbb{G}$ for $1 \leq i \leq 4$. Note that for any $X_i \in \mathbb{G}_{p_i}$ and $X_j \in \mathbb{G}_{p_j}$, $\hat{e}(X_i, X_j) = 1$ holds for $i \neq j$.

2.2 Access Structures and Linear Secret Sharing Schemes

The adopted access structure can be represented by a linear secret sharing scheme.

Definition 2 (Linear Secret Sharing Schemes (LSSS) [1]). Let $\mathscr{U}$ be the attribute universe, where each attribute includes two parts: attribute name and its values. Each attribute has multiple values. An LSSS can be used to represent an access structure $(\mathbf{A},\rho)$ on $\mathscr{U}$, where $\mathbf{A}$ is an $\ell \times n$ matrix which is called the share-generating matrix and ρ maps a row of $\mathbf{A}$ into an attribute name index. An LSSS consists of two algorithms: Share and Reconstruction [1].

We say that $I \subseteq \{1,2,\ldots,\ell\}$ satisfies $(\mathbf{A},\rho)$ if there exists constants $\{\omega_i\}_{i\in I}$ such that $\sum_{i\in I}\omega_i A_i = (1,0,...,0)$. Suppose a user has a secret key associated with a set of attribute name indexes I_S and the corresponding attribute value set is $S=(s_1,s_2,\ldots,s_n)$. We use $\mathbb{A}=(\mathbf{A},\rho,T)$ to represent the adopted access structure, where $T=(t_{\rho(1)},t_{\rho(2)},\ldots,t_{\rho(n)})$ is the attribute value set specified by $(\mathbf{A},\rho)$. We also say that S matches $\mathbb{A}$ if there exist an $I \subseteq \{1,2,\ldots,\ell\}$ satisfying $(\mathbf{A},\rho)$, $I \subseteq I_S$ and $s_{\rho(i)} = t_{\rho(i)}$ for each $i \in I$.

3 Definition and Security Model

3.1 Definition of Anonymous CP-ABE with Offline Computation

An anonymous CP-ABE scheme with offline mechanisms is defined as:

- **Setup**$(1^{\lambda}) \rightarrow (PK,MK)$: The setup algorithm takes as inputs the security parameter λ, and it outputs the system public key PK and the master key MK.
- **Offline.KeyGen**$(PK,MK) \rightarrow SK_{\text{off}}$: The offline key generation algorithm takes as inputs PK and MK. It outputs SK_{off} as an offline key.
- **Online.KeyGen**$(PK,SK_{\text{off}},S) \rightarrow SK_S$: Upon receiving an attribute set S, the online key generation algorithm takes as inputs PK and an offline key SK_{off}. It generates SK_S as the secret key associated with S.
- **Offline.Enc**$(PK) \rightarrow CT_{\text{off}}$: The offline encryption algorithm takes as input PK, and it generates an offline ciphertext CT_{off}.
- **Online.AnonEnc**$(PK,CT_{\text{off}},M,\mathbb{A}) \rightarrow CT_{\mathbb{A}}$: To encrypt a message M with the access structure $\mathbb{A}$, the online anonymous encryption algorithm generates the final ciphertext $CT_{\mathbb{A}}$ based on PK and an offline ciphertext CT_{off}. It's noted that the values of attributes in $\mathbb{A}$ cannot be explicitly included in $CT_{\mathbb{A}}$.
- **AnonDec**$(PK,SK_S,CT_{\mathbb{A}}) \rightarrow M$ or $\perp$: The anonymous decryption algorithm takes as inputs PK, a secret key SK_S with respect to S and a ciphertext $CT_{\mathbb{A}}$ associated with $\mathbb{A}$ which is hidden in $CT_{\mathbb{A}}$. If S matches $\mathbb{A}$, it outputs the potential message M, and it outputs $\perp$ otherwise.

3.2 Security Model

We define the indistinguishability against chosen access structure and chosen plaintext attacks in anonymous CP-ABE supporting offline key generation and offline encryption. It is defined by a game between an adversary $\mathscr{A}$ and a challenger $\mathscr{B}$.

Setup: The challenger $\mathscr{B}$ runs $(PK, MK) \leftarrow \textbf{Setup}(1^\lambda)$. It gives the system public key PK to $\mathscr{A}$ and keeps MK secret.
Phase 1: $\mathscr{A}$ issues a polynomially bounded number of queries $\mathscr{O}_{KeyGen}$: $\mathscr{A}$ submits an attribute set S. The challenger $\mathscr{B}$ runs $SK_{\text{off}} \leftarrow \textbf{Offline.KeyGen}(PK, MK)$ and $SK_S \leftarrow \textbf{Online.KeyGen}(PK, SK_{\text{off}}, S)$, then gives $\mathscr{A}$ the secret key SK_S for S.
Challenge: Once $\mathscr{A}$ decides that **Phase 1** is over, it submits to $\mathscr{B}$ two messages M_0, M_1 of equal length and two access structures $\mathbb{A}_1^* = (\mathbf{A}^*, \rho^*, T_0)$, $\mathbb{A}_2^* = (\mathbf{A}^*, \rho^*, T_1)$ with the restriction that $\mathbb{A}_1^*$ and $\mathbb{A}_2^*$ cannot be satisfied by any of the queried attribute sets in **Phase 1**. $\mathscr{B}$ flips a random coin $b \in \{0,1\}$, and encrypts M_b under $\mathbb{A}$ by running $CT_{\text{off}} \leftarrow \textbf{Offline.Enc}(PK)$ and $CT_{\mathbb{A}_b^*} \leftarrow \textbf{Online.AnonEnc}(PK, CT_{\text{off}}, M_b, \mathbb{A}_b^*)$. Then it sends $CT_{\mathbb{A}_b^*}$ to $\mathscr{A}$.
Phase 2: The same as **Phase 1** with the restriction that $\mathbb{A}_1^*$ and $\mathbb{A}_2^*$ cannot be satisfied by any of the queried attribute sets.
Guess: The adversary $\mathscr{A}$ outputs a guess bit $b' \in \{0,1\}$ and wins the game if $b' = b$. The advantage of an adversary $\mathscr{A}$ in the above game is defined as $\left|\Pr[b' = b] - \frac{1}{2}\right|$.

4 Anonymous CP-ABE Scheme Supporting Offline Key Generation and Offline Encryption

4.1 Our Construction

Setup(1^λ): The setup algorithm first generates $(p_1, p_2, p_3, p_4, \mathbb{G}, \mathbb{G}_T, \hat{e})$ with $\mathbb{G} = \mathbb{G}_{p_1} \times \mathbb{G}_{p_2} \times \mathbb{G}_{p_3} \times \mathbb{G}_{p_4}$, where p_1, p_2, p_3, p_4 are distinct primes, $\mathbb{G}$ and $\mathbb{G}_T$ are cyclic groups of order $N = p_1 p_2 p_3 p_4$, and $\hat{e} : \mathbb{G} \times \mathbb{G} \to \mathbb{G}_T$ is a bilinear map. The attribute universe is $\mathscr{U} = \{1, 2, \ldots, U\} \subseteq \mathbb{Z}_N$. Then it uniformly chooses $\alpha, a, a_1, a_2, \ldots, a_n \in_R \mathbb{Z}_N$, $g, h \in_R \mathbb{G}_{p_1}$, $X_3 \in_R \mathbb{G}_{p_3}$, $Z, X_4 \in_R \mathbb{G}_{p_4}$ and computes $Y = \hat{e}(g,g)^\alpha, H = hZ$. The system public parameters are published as $PK = (N, g, g^a, \{a_i\}_{1 \le i \le U}, Y, H, X_4)$, and the master key is $MK = (\alpha, h, X_3)$.

Offline.KeyGen(PK, MK): The offline key generation algorithm uniformly chooses $t, \hat{s}_1, \hat{s}_2, \ldots, \hat{s}_U \in_R \mathbb{Z}_N$ and $R, R', R_1, R_2, \ldots, R_U \in_R \mathbb{G}_{p_3}$. It computes $u_i = g^{a_i}$, for $1 \le i \le U$, and outputs the offline secret key $SK_{\text{off}} = (K, K', \{\hat{s}_i, K_i\}_{1 \le i \le U})$, where $K = g^\alpha g^{at} R, K' = g^t R', K_i = (u_i^t)^{\hat{s}_i} h^t R_i$.

Online.KeyGen(PK, SK_{off}, S): Upon receiving an attribute set $S = (s_1, s_2, \ldots, s_n)$, based on $SK_{\text{off}} = (S, K, K', \{\hat{s}_i, K_i\}_{1 \le i \le U})$, the online key generation algorithm outputs $SK_S = (S, K, K', \{L_i, K_i\}_{i \in I_S})$ as the final secret key associated with S, where $I_S \subseteq \{1, 2, \ldots, U\}$ is the attribute name index set corresponding to the attribute value set S, $|I_S| = n$ and $L_i = s_i - \hat{s}_i$. Without loss of generality, it is supposed that the i-th attribute name in S has attribute value s_i for simplicity of description.

Offline.Enc(PK): The offline encryption algorithm chooses $s, s' \in_R \mathbb{Z}_N$ and $\hat{t}_k \in_R \mathbb{Z}_N$ for $1 \le k \le U$. It also uniformly chooses $\hat{\lambda}'_x, \hat{\lambda}_x, r'_x, r_x \in_R \mathbb{Z}_N$ and $Z_{0,x}, Z'_{0,x}, Z_{1,x}, Z'_{1,x} \in_R \mathbb{G}_{p_4}$, for $1 \le x \le U$. Then it calculates $u_k = g^{a_k}$ for $k \in \{1,2,\ldots,U\}$ and sets the offline ciphertext as

$$CT_{\text{off}} = (\{\hat{t}_k\}_{1\le k\le U}, s', \tilde{C}_0, \bar{C}_0, \{\hat{\lambda}'_x, C_{0,x,k}, D_{0,x}\}_{1\le x\le U, 1\le k\le U}, s, \hat{C}_1, \bar{C}_1, \{\hat{\lambda}_x, C_{1,x,k}, D_{1,x}\}_{1\le x\le U, 1\le k\le U}),$$

where $\tilde{C}_0 = Y^{s'}, \bar{C}_0 = g^{s'}, C_{0,x,k} = g^{a\hat{\lambda}'_x}(u_k^{\hat{t}_k}H)^{-r'_x}Z_{0,x}, D_{0,x} = g^{r'_x}Z'_{0,x}$, $\hat{C}_1 = Y^s, \bar{C}_1 = g^s, C_{1,x,k} = g^{a\hat{\lambda}_x}(u_k^{\hat{t}_k}H)^{-r_x}Z_{1,x}, D_{1,x} = g^{r_x}Z'_{1,x}$.

Online.AnonEnc$(PK, CT_{\text{off}}, M, \mathbb{A})$: To encrypt a message $M \in \mathbb{G}_T$ under an access structure $\mathbb{A} = (\mathbf{A}, \rho, T)$, where $\mathbf{A}$ is an $\ell \times m$ matrix, ρ is a map from each row A_x of $\mathbf{A}$ to an attribute name index in $\{1,2,\ldots,U\}$, and $T = (t_{\rho(1)}, t_{\rho(2)}, \ldots, t_{\rho(\ell)}) \in \mathbb{Z}_N^{\ell}$. The online anonymous encryption algorithm chooses $v'_i, v_i \in_R \mathbb{Z}_N$, for $i \in \{2,3,\ldots,m\}$, and sets $v' = (s', v'_2, \ldots, v'_m)$ and $v = (s, v_2, \ldots, v_m)$. Then, it computes $\lambda'_x = \mathbf{A}_x \cdot v'$, $\lambda_x = \mathbf{A}_x \cdot v$, $F_{0,x} = \lambda'_x - \hat{\lambda}'_x$, $F_{1,x} = \lambda_x - \hat{\lambda}_x$, $E_{\rho(x)} = t_{\rho(x)} - \hat{t}_{\rho(x)}$, for $1 \le x \le \ell$. Finally, it sets the final ciphertext as

$$CT_{\mathbb{A}} = ((\mathbf{A}, \rho), \{E_{\rho(x)}\}_{1\le x\le \ell}, \tilde{C}_0, \bar{C}_0, \{F_{0,x}, C_{0,x}, D_{0,x}\}_{1\le x\le \ell}, \tilde{C}_1, \bar{C}_1, \{F_{1,x}, C_{1,x}, D_{1,x}\}_{1\le x\le \ell}),$$

where $\tilde{C}_1 = \hat{C}_1 \cdot M$, $C_{0,x} = C_{0,x,\rho(x)}$ and $C_{1,x} = C_{1,x,\rho(x)}$ for $1 \le x \le \ell$.

AnonDec$(PK, SK_S, CT_{\mathbb{A}})$: Let $CT_{\mathbb{A}} = ((\mathbf{A}, \rho), \{E_{\rho(x)}\}_{1\le x\le \ell}, \tilde{C}_0, \bar{C}_0, \{F_{0,x}, C_{0,x}, D_{0,x}\}_{1\le x\le \ell}, \tilde{C}_1, \bar{C}_1, \{F_{1,x}, C_{1,x}, D_{1,x}\}_{1\le x\le \ell})$, $SK_S = (S, K, K', \{L_i, K_i\}_{i\in I_S})$ and $S = (s_1, s_2, \ldots, s_n)$. The anonymous decryption algorithm first calculates $\mathbf{I}_{\mathbf{A},\rho}$ from $(\mathbf{A}, \rho)$, where $\mathbf{I}_{\mathbf{A},\rho}$ denotes the set of minimum subsets of $\{1,2,\ldots,\ell\}$ that satisfies $(\mathbf{A}, \rho)$. Then it checks if there exists an $I \in \mathbf{I}_{\mathbf{A},\rho}$ that satisfies

$$\tilde{C}_0 = \frac{\hat{e}(\bar{C}_0, K)}{\prod_{i\in I}\left(\hat{e}(C_{0,i} \cdot (g^a)^{F_{0,i}} \cdot D_{0,i}^{-a_{\rho(i)}E_{\rho(i)}}, K')\hat{e}(D_{0,i}, K_{\rho(i)} \cdot (K')^{a_{\rho(i)}L_{\rho(i)}})\right)^{\omega_i}},$$

where $I \subseteq \{i|\rho(i) \in I_S\}$ and $\sum_{i\in I}\omega_i A_i = (1,0,\ldots,0)$ for some constants $\{\omega_i\}_{i\in I}$. If no such I exists, it outputs $\perp$ to indicate that S does not satisfy the hidden access structure $\mathbb{A}$. Otherwise, it returns $M = \frac{\tilde{C}_1}{E}$, where

$$E = \frac{\hat{e}(\bar{C}_1, K)}{\prod_{i\in I}\left(\hat{e}(C_{1,i} \cdot (g^a)^{F_{1,i}} \cdot D_{1,i}^{-a_{\rho(i)}E_{\rho(i)}}, K')\hat{e}(D_{1,i}, K_{\rho(i)} \cdot (K')^{a_{\rho(i)}L_{\rho(i)}})\right)^{\omega_i}}.$$

4.2 Security Analysis

Theorem 1. *The proposed anonymous CP-ABE scheme supporting offline key generation and offline encryption is fully secure in the standard model.*

Proof. The proposed anonymous CP-ABE scheme Π is an improved version of the scheme Π_o [16]. Suppose there exists a PPT attacker $\mathscr{A}$ with a non-negligible advantage ε in the proposed security game against Π. We show how to design a PPT simulator $\mathscr{B}$, which can break the security of Π_o with an advantage ε. We will give more details in the full version due to the space limitation. □

4.3 Performance Analysis

In our scheme, it easily follows that only n subtraction operations in arithmetic are needed for the attribute authority to generate a secret key in the online phase, where n means the number of attributes in the attribute set. In the online encryption phase, a data owner only needs to perform k multiplication operations in arithmetic, where k represents the complexity of the access structure. The final ciphertext generated in the online phase does not explicitly include the attribute values specified in the access structure. Accordingly, the proposed scheme can preserve users' attribute privacy. Similar to the anonymous CP-ABE scheme [16], our scheme supports any monotonic access structures.

5 Conclusion

We propose an anonymous CP-ABE scheme supporting online/offline key generation and online/offline encryption. The proposed scheme is proven fully secure in the standard model. Because the attribute values of access structures are hidden in ciphertexts, our scheme can protect users' attribute privacy.

Acknowledgements This work is supported by National Natural Science Foundation of China (No. 61402366,61272037,61472472), Natural Science Basic Research Plan in Shaanxi Province (No. 2015JQ6236), Scientific Research Program Funded by Shaanxi Provincial Education Department (No. 15JK1686). Yinghui Zhang is supported by New Star Team of Xi'an University of Posts and Telecommunications.

References

1. Beimel, A.: Secure schemes for secret sharing and key distribution. Ph.D. thesis, Technion-Israel Institute of Technology (1996)

2. Bethencourt, J., Sahai, A., Waters, B.: Ciphertext-policy attribute-based encryption. SP'07, pp. 321–334. IEEE, Oakland (2007)
3. Boneh, D., Goh, E.J., Nissim, K.: Evaluating 2-dnf formulas on ciphertexts. TCC'05, *LNCS*, vol. 3378, pp. 325–341. Springer, Berlin (2005)
4. Boneh, D., Waters, B.: Conjunctive, subset, and range queries on encrypted data. TCC'07, *LNCS*, vol. 4392, pp. 535–554. Springer, Berlin (2007)
5. Chen, X., Zhang, F., Susilo, W., Mu, Y.: Efficient generic on-line/off-line signatures without key exposure. ACNS'07, *LNCS*, vol. 4521, pp. 18–30. Springer, Berlin (2007)
6. Chen, X., Zhang, F., Tian, H., Wei, B., Susilo, W., Mu, Y., Lee, H., Kim, K.: Efficient generic on-line/off-line (threshold) signatures without key exposure. Information Sciences **178**(21), 4192–4203 (2008)
7. Cheung, L., Newport, C.: Provably secure ciphertext policy abe. CCS'07, pp. 456–465. ACM, New York (2007)
8. Datta, P., Dutta, R., Mukhopadhyay, S.: Fully secure online/offline predicate and attribute-based encryption. ISPEC'15, *LNCS*, vol. 9065, pp. 331–345. Springer, Berlin (2015)
9. Even, S., Goldreich, O., Micali, S.: On-line/off-line digital signatures. Journal of Cryptology **9**(1), 35–67 (1996)
10. Goyal, V., Pandey, O., Sahai, A., Waters, B.: Attribute-based encryption for fine-grained access control of encrypted data. CCS'06, pp. 89–98. ACM, New York (2006)
11. Guo, F., Mu, Y., Chen, Z.: Identity-based online/offline encryption. FC'08, *Lecture Notes in Computer Science*, vol. 12, pp. 247–261. Springer, Berlin (2008)
12. Hohenberger, S., Waters, B.: Online/offline attribute-based encryption. PKC'14, *LNCS*, vol. 8383, pp. 293–310. Springer, Berlin (2014)
13. Jung, T., Li, X.Y., Wan, Z., Wan, M.: Control cloud data access privilege and anonymity with fully anonymous attribute-based encryption. IEEE Transactions on Information Forensics and Security **10**(1), 190–199 (2015)
14. Kapadia, A., Tsang, P.P., Smith, S.W.: Attribute-based publishing with hidden credentials and hidden policies. NDSS'07, pp. 179–192. The Internet Society (2007)
15. Katz, J., Sahai, A., Waters, B.: Predicate encryption supporting disjunctions, polynomial equations, and inner products. EUROCRYPT'08, *Lecture Notes in Computer Science*, vol. 4965, pp. 146–162. Springer, Berlin (2008)
16. Lai, J., Deng, R.H., Li, Y.: Expressive cp-abe with partially hidden access structures. ASIACCS'12, pp. 18–19. ACM, New York (2012)
17. Li, J., Ren, K., Zhu, B., Wan, Z.: Privacy-aware attribute-based encryption with user accountability. ISC'09, *LNCS*, vol. 5735, pp. 347–362. Springer, Berlin (2009)
18. Nishide, T., Yoneyama, K., Ohta, K.: Abe with partially hidden encryptor-specified access structure. ACNS'08, *Lecture Notes in Computer Science*, vol. 5037, pp. 111–129. Springer, Berlin (2008)
19. Phuong, T.V.X., Yang, G., Susilo, W.: Hidden ciphertext policy attribute-based encryption under standard assumptions. IEEE Transactions on Information Forensics and Security **11**(1), 35–45 (2016)
20. Rao, Y.S., Dutta, R.: Fully secure bandwidth-efficient anonymous ciphertext-policy attribute-based encryption. Security and Communication Networks **8**(18), 4157–4176 (2015)
21. Sahai, A., Waters, B.: Fuzzy identity-based encryption. EUROCRYPT'05, *LNCS*, vol. 3494, pp. 557–557. Springer, Berlin (2005)
22. Zhang, Y., Chen, X., Li, J., Li, H.: Generic construction for secure and efficient handoff authentication schemes in eap-based wireless networks. Computer Networks **75**, 192–211 (2014)
23. Zhang, Y., Chen, X., Li, J., Wong, D.S., Li, H.: Anonymous attribute-based encryption supporting efficient decryption test. ASIACCS'13, pp. 511–516. ACM, New York (2013)
24. Zhang, Y., Chen, X., Li, J., Wong, D.S., Li, H., You, I.: Ensuring attribute privacy protection and fast decryption for outsourced data security in mobile cloud computing. Information Sciences (2016). DOI 10.1016/j.ins.2016.04.015
25. Zhang, Y., Li, J., Chen, X., Li, H.: Anonymous attribute-based proxy re-encryption for access control in cloud computing. Security and Communication Networks **9**(14), 2397–2411 (2016).
26. Zhang, Y., Zheng, D., Chen, X., Li, J., Li, H.: Efficient attribute-based data sharing in mobile clouds. Pervasive and Mobile Computing **28**, 135–149 (2016)

Flexible Attribute-Based Keyword Search Via Two Access Policies

Peilin Zhou, Zhenhua Liu, and Shuhong Duan

Abstract Attribute-based keyword search (ABKS) allows users, whose credentials satisfy the owner's access control policy, to search over the encrypted data in cloud environment. However, most current schemes can not simultaneously achieve that owners dominate the data while users retrieve the interested files more accurately, either in ciphertext policy or key policy setting. Furthermore, majority of ABKS schemes ignore the decryption of retrieved files, or focus little about decryption efficiency. Therefore, aiming at these limitations, we propose a flexible attribute-based keyword search scheme via two access policies, which allows: (1) the data owner to control the access policy in order to gain control over their data and specify who can access the files; (2) the user to define the search policy so that he can search the interested encrypted files more accurately. In addition, we employ an online/offline technique to improve the efficiency. And the ciphertexts can be decrypted with two pairings while it grows linear with the number of attributes in most existing schemes. Our system is selectively secure against chosen-plaintext attack and chosen-keyword attack, and it also achieves keyword secrecy.

1 Introduction

Cloud computing enables data users to outsource their data to remote cloud servers. Among these servers, cloud storage server is of great use for data availability, efficient data management and low-cost pay-per-use. Despite numerous advantages, data outsourcing leads to confidential problem due to the fact that the

Peilin Zhou, School of Mathematics and Statistics, Xidian University, Xi'an 710071, P.R. China
e-mail: plzhou1224@163.com ·
Zhenhua Liu, School of Mathematics and Statistics, Xidian University, Xi'an 710071, P.R. China
e-mail: zhualiu@hotmail.com ·
Shuhong Duan, School of Mathematics and Statistics, Xidian University, Xi'an 710071, P.R. China
e-mail: 0801duanshuhong@163.com

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_83

cloud server are not fully trusted. Encryption-before-outsourcing has been regarded as a fundamental means of protecting data's privacy against the cloud server.

However, some useful data stored in the cloud should be shared with users who want to fine-grained access control them. Therefore, attribute-based encryption (ABE) which is introduced by Sahai and Waters [9], can be a resultful cryptographic tool. In ABE, data owners can share data with users whose attributes satisfy a specific access policy. ABE can be classified into two types: key-policy ABE (KP-ABE) [4] and ciphertext-policy ABE (CP-ABE) [7]. Later on, ABE has been enriched with various features, but those schemes have little application in some areas for lacking of functionality such as searchability.

Attribute-based keyword search (ABKS) allows users with proper cryptographic credentials to search over the outsourced encrypted data. Zheng et al. [10] proposed this new primitive ABKS and construct two schemes. Aiming at Zheng's scheme, Dong et al. [3] proposed a new ABKS construction by using online/offline technique [6] to improve the efficiency. However, these two schemes do not support decryption service. In fact, most previous studies such as [1, 8] also pay little attention to decryption of the files while conducting the search operation. Besides, several other ABKS schemes [2, 5] with decryption service may incur high computational cost, and the decryption cost leaves much to be desired.

To address the above problems, we propose a flexible attribute-based keyword search via two access policies (FABKS) scheme. The main contribution is as follows:

- We build a construction that combines a keyword search system and a ABE system into one system, which also supports large universe. In such a system, data owner can control the access policy. Meanwhile, users can control the search policy and decrypt the encrypted files.
- The proposed scheme can achieve multi-keywords search through applying different policies. It is owing to the fact that users can define the search policy according to the attributes corresponding the specific keyword during the search operation. We can regard these "keyword attributes" as "keywords" in some sense. Furthermore, our scheme is built on the multi-owner/multi-user(M/M) setting.
- To improve the efficiency, we adopt the online/offline technique in Index Generation and Token Generation algorithm. And through adopting a fast decryption transformation, the decryption cost of our scheme is greatly cut down.

In the following section, for limited space, we will omit the following section: (1) Preliminaries, which include bilinear map [4], access structure [4] and linear secret sharing schemes (LSSS) [7]; (2) Complexity assumptions, namely Decisional q-Bilinear Diffie-Hellman Exponent (Dq-BDHE) Assumption[7] and Decisional Linear (DL) Assumption [10]; (3) Syntax and security model for FABKS construction. Lastly, we will present the system entities, Fig.1 shows that our scheme consists of four entities:

- Trusted Authority (TA): It is fully trusted and it generates the decryption keys and search keys for data users.

- Multiple Data owners (DO): They outsource their encrypted files and keywords to the cloud for storage and sharing.
- Multiple Data users (DU): They create search tokens according to some interested keywords and launch the keyword search. Besides, they can decrypt the files returned from cloud server.
- Cloud Server (CS): It is honest-but-curious, and it conducts the search operations and returns the search results to data users.

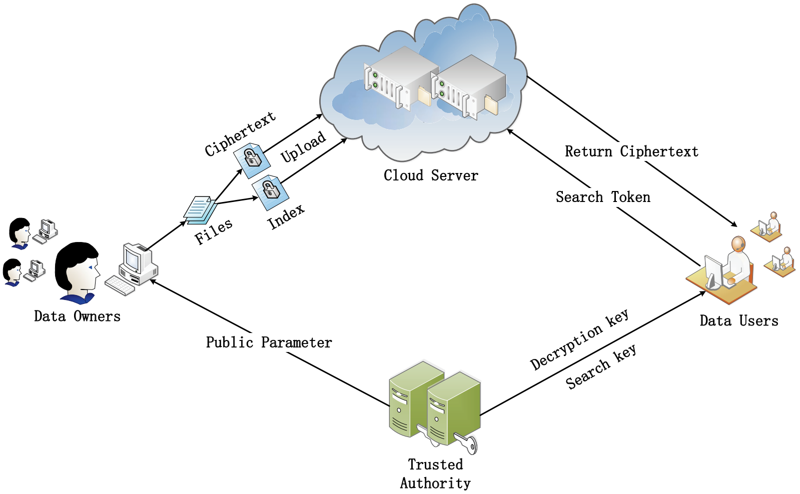


Fig. 1 The Framework of the FABKS Scheme

2 Our Construction

In this section, we will present the concrete FABKS construction. Let $a \in_R S$ denote selecting an element a from a set S uniformly at random. Let S be an attribute set and (M,ρ) an access structure. We define the function f: if S satisfy (M,ρ), $f(S,(M,\rho)) = 1$; Otherwise $f(S,(M,\rho)) = 0$. Now we describe the scheme:

Setup(1^κ,$Attr_{max}, l_{max}$): Given a security parameter 1^κ, the maximum number of attributes $Attr_{max}$ a user's secret key may have and the maximum number of columns l_{max} in a ciphertext access matrix. Let $\mathbb{G}$ and $\mathbb{G}_T$ be cyclic groups of prime order p, g be the generator of $\mathbb{G}$, and $e : \mathbb{G} \times \mathbb{G} \rightarrow \mathbb{G}_T$ be a bilinear map. Choose $a,b,c,\alpha \in_R \mathbb{Z}_p$. We define a hash function $H_1 : \mathbb{Z}_p \rightarrow \mathbb{G}$. It dose this by implicitly choosing a polynomial $p(x) \in \mathbb{Z}_p$ of degree $m = Attr_{max} + l_{max} - 1$ according to [7], and then computes $h_0 = g^{p(0)}, \ldots, h_m = g^{p(m)}$. Anyone will be able to compute $g^{p(x)}$ for any $x \in \mathbb{Z}_p$ when published these $m+1$ values. Let $H_2 : \{0,1\}^* \rightarrow \mathbb{Z}_p$ be a one-way hash function. The public parameters are

$$PK = \langle e, g, g^a, g^b, g^c, e(g,g)^{\alpha}, H_1, H_2, h_0, \ldots, h_m \rangle \ . \tag{1}$$

The master secret key is set as

$$MSK = \langle g^{\alpha}, a, b, c \rangle \ . \tag{2}$$

ABE-KeyGen(MSK, S): The key generation algorithm takes as input the master secret key MSK and an attribute set $S \subseteq \mathbb{Z}_p$. It chooses $t \in_R \mathbb{Z}_p$ and creates the private key SK as

$$SK = \langle K = g^{\alpha} g^{at}, K_1 = g^t, \{K_x = H_1(x)^t\}_{x \in S} \rangle \ . \tag{3}$$

KSF-KeyGen$(MSK, (M', \rho'))$: If DU wants to search messages containing a keyword kw with attributes set $\hat{S} \subseteq \mathbb{Z}_p$, to generate a query private key SK' associated with the search policy for a DU, the following protocol will be executed.

- DU defines a search policy $(M', \rho')(M'$ is a $l' \times n'$ matrix, and the function ρ' maps rows of M' to attributes) according to $\hat{S}$ and sends it to TA.
- Upon receiving (M', ρ'), TA selects a random vector $v = (ac, y'_2, \ldots, y'_n)$, where $y'_2, \ldots, y'_n \in_R \mathbb{Z}_p$. For $i = 1$ to l', sets $\lambda'_i = \upsilon' \cdot M'_i$, where M'_i is the vector corresponding to the ith row of M'.
- For $i = 1$ to l', TA sets $SK' = \langle D_i = g^{\lambda'_i} H_1(\rho'(i))^t, K_1 = g^t \rangle$, and returns SK' to DU.

Encrypt$(PK, (M, \rho), m)$: The algorithm takes as input public parameters PK, an $LSSS$ access structure $(M, \rho)(M$ is a $l \times n$ matrix, and the function ρ to be an injective function which maps rows of M to attributes) and a message m.

It then chooses a random vector $v = (s, y_2, \ldots, y_n) \in \mathbb{Z}_p^n$ which used to share the encryption exponent s. For $i = 1$ to l, set $\lambda_i = \upsilon \cdot M_i$, where M_i is the vector corresponding to the ith row of M. The ciphertext is published as

$$CT = \langle (M, \rho), C = me(g,g)^{\alpha s}, C_0 = g^s, \{C_i = g_i^{a\lambda} H_1(\rho(i))^{-s}\}_{i \in [l]} \rangle \ . \tag{4}$$

Offline.Index(PK): The purpose of this algorithm which takes in the public parameter only is to do a preparation task for generating the secure index. It selects $r_1, r_2 \in_R \mathbb{Z}_p$ and compute$W_0 = g^{cr_1}, W_1 = g^{a(r_1+r_2)}, W_2 = g^{r_2}$. For each $A'_j \in \mathbb{Z}_p$, compute$W_j = H_1(A'_j)^{r_2}$. The intermediate keyword ciphertext is

$$IX_{kw} = \langle r_1, r_2, W_0, W_1, W_2, \{W_j\}_{A'_j \in \mathbb{Z}_p} \rangle \ . \tag{5}$$

Online.Index(IX_{kw}, S', KW): Suppose a data owner wants to share a message m containing a keyword set KW with an attributes set $S' \subseteq \mathbb{Z}_p$. The algorithm takes as input IX_{kw}, an attributes set S' and a keyword set KW. It computes $W = W_1 \cdot g^{br_1 H_2(kw_i)}$ for each $kw_i \in KW$, and sets the keyword ciphertext as

$$IX = \langle S', W, W_0, W_2, \{W_j\}_{A'_j \in S'} \rangle \ . \tag{6}$$

Offline.TokenGen(SK', PK): The algorithm is a preparation for generating a token. It takes as input PK and DU's query secret key SK'. It selects $\beta \in_R \mathbb{Z}_p$, and computes $D'_i = D_i^{\beta}, K'_i = K_1^{\beta}$. Then computes $tk_1 = g^{a\beta}$ and $tk_2 = g^{c\beta}$. Finally sets the intermediate token as

$$IT = \langle \beta, tk_1, tk_2, \{(D'_i, K'_i\}_{i\in[l']} \rangle \ . \tag{7}$$

Online.TokenGen(IT, kw'): It takes as input an intermediate token IT and a keyword kw'. Compute $tk'_1 = tk_1 \cdot g^{b\beta H_2(kw')}$ and set the search token as

$$TK = \langle tk'_1, tk_2, \{(D'_i, K'_i\}_{i\in[l']} \rangle \ . \tag{8}$$

Test(TK, IX): To perform keyword test, the algorithm can be done as follows:

- DU initiates a keyword search request by sending the search token TK along with the attribute set S related to DU's decryption key to CS.
- First, CS verifies $f(S,(M,\rho)) \overset{?}{=} 1$. If satisfied, CS then searches for the correspondence ciphertext CT with the desired keyword kw'. Given attribute set S' specified in IX, select an attribute set $AS' \subseteq (S' \bigcap \hat{S})$ (Note that if $S' \bigcap \hat{S} = \emptyset$ or AS' doesn't exist, then output $\perp$) and verify $f(AS',(M',\rho')) \overset{?}{=} 1$.
- If satisfied, Let $I' = \{i : \rho'(i) \in AS'\} \subset \{1,\ldots l'\}$, then there must exists coefficients $\{\omega'_i | i \in I'\}$ such that $\sum_{i\in I} \omega'_i M'_i = (1,0,\ldots,0)$, so $\sum_{i\in I} \omega'_i \lambda'_i = ac$. CS computes
$$E = \prod_{i\in I'} \left(\frac{e(D'_i, W_2)}{e(K'_i, W_j)}\right)^{w'_i} \ . \tag{9}$$
- CS checks whether the kw' in the token TK matches the kw in the secure index IX by verifying the following equation
$$e(W_0, tk'_1)E = e(W, tk_2) \ . \tag{10}$$

If it holds, output 1 and CS sends the search result that include ciphertext CT to DU. Otherwise, output $\perp$.

Decrypt(CT, SK): Given a private key SK for a set S and a ciphertext CT for a linear access structure (M,ρ). If $f(S,(M,\rho)) = 0$, abort. Otherwise, let $I = \{i : \rho(i) \in S\} \subset \{1,\ldots l\}$, then there must exists coefficients $\{\omega_i | i \in I\}$ such that $\sum_{i\in I} \omega_i \cdot M_i = (1,0,\ldots,0)$, so $\sum_{i\in I} \omega_i \cdot \lambda_i = s$. The decryption algorithm first compute
$$Z = e(\prod_{i\in I} C_i^{-\omega_i}, K_1) e(C_0, K \prod_{i\in I} K_{\rho(i)}^{-\omega_i}) = e(g,g)^{\alpha s} \ . \tag{11}$$
and then obtain m by computing $m = C/Z$.

3 Security Analysis

Similar to Waters' system [7] in appendix B and Zheng's scheme [10] in the key policy setting, our construction can achieve the following properties. For limited space, the proof of the following theorems are not shown here and can be provided on request.

Theorem 1. *The above FABKS scheme is sCPA-secure assuming that the scheme of Waters (appendix B) is a sCPA-secure CP-ABE system.*

Theorem 2. *Given the DL assumption and one-way hash function H_2, the above FABKS scheme is sCKA-secure in the random oracle model under the assumption that Zheng's scheme is sCKA-secure in the random oracle model.*

Theorem 3. *Given the one-way hash function H_2, the above FABKS scheme achieves keyword secrecy in the random oracle model if Zheng's scheme achieves keyword secrecy in the random oracle model.*

4 Discussion

4.1 Functionality and Features Analysis

we first analyse the functionality and features of our scheme and previous schemes[2, 3, 5, 7]. Table 1 shows that the proposed FABKS scheme has a few advantages over these schemes. Note that MK represents multi-keyword search, LU represents large universe and K-secrecy represents Keyword secrecy.

Table 1 Functionality and features comparison

Schemes	Access policy	Search policy	MK	Provable security			LU
				CPA	SCKA	K-secrecy	
[2]	CP(AND)	CP(AND)	√	random	√	×	×
[3]	–	KP/CP(Access Tree)	×	–	√	√	×
[5]	CP(LSSS)	KP(OR and AND)	√	random	×	×	×
[7]	CP(LSSS)	–	–	standard	–	–	√
Ours	CP(LSSS)	KP(LSSS)	√	standard	√	√	√

- First, the schemes proposed in Dong's scheme [3] and Waters' scheme [7] achieve data access control and keyword search separately, while ours supports both. Furthermore, the scheme in Dong [3] only supports single-keyword search. Whereas, we can achieve multi-keywords search through separating the data access and data retrieving into two different policies.

- Second, the schemes proposed in [2, 5] and ours both achieve multi-keyword search and decryption service, but ours supports large universe construction and any monotonous access structure by using LSSS. In addition, our scheme exploits offline/online technique in Dong's scheme [3] so there is efficiency superiority.
- Finally, we utilize different policies to control data access and retrieve in our design, so we prove the security of our proposed scheme separately. For data access, we can prove that our basic ABE construction is sCPA-secure according to Waters' scheme [7]. And for data retrieve, we prove the sCKA security and keyword secrecy in the random oracle based on Zheng's scheme [3].

4.2 Efficiency Analysis

In this subsection, we analyze the efficiency of the schemes proposed in [2, 5] and our scheme, because the above two constructions also achieve keyword search and decryption service simultaneously. For convenience, we list all the parameters used in Table 2. E, P represent the exponentiation and pairing operation, respectively. Denote H_1 to be the hash operation, $|S|$ and $|S'|$ to be the number of a DU's attributes and the number of attributes corresponding to a secure index for keywords, l and $l'$to be the number of attributes that are involved in an access policy (M, ρ) and (M', ρ'), to be the number of attributes that are involved in an access policy, I and I' to be a subset of $\{1, 2, \ldots, l\}$ and $\{1, 2, \ldots, l'\}$, respectively. E_{SE} and D_{SE} represents the operation of symmetric encryption and decryption separately [5].

Table 2 Efficiency comparison

Computation	Our scheme	[2]	[5]'s improved scheme
Encrypt	$(l+2)E$	$(2\lvert l\rvert+3)E$	E_{SE}
Offline-Index	$(3+\lvert S'\rvert)E+\lvert S'\rvert H_1$	–	–
Online-Index	E	$> (l+1)E$	$2P+\lvert S'\rvert E+\lvert S'\rvert H_1$
Offline-TokenGen	$(l'+3)E$	–	–
Online-TokenGen	E	$> (\lvert S\rvert+1)E$	$< (2(l')^2+5l'+4)E+H_1$
Test	$(2\lvert I'\rvert+2)P$	$> (\lvert S\rvert+1)P+E$	$3P+(2\lvert I'\rvert+2)E$
Decrypt	$2P+(2+\lvert I\rvert)E$	$(\lvert S\rvert+2)P$	$(2\lvert I\rvert+1)P+E+D_{SE}$

As shown in Table 2, it is obvious that the efficiency of the proposed scheme is higher than [2, 5] during Index and TokenGen phase, because we move the majority task into an offline phase. In addition, the ciphertexts can be decrypted with a constant number of pairings in our scheme, while in [2, 5] it grows linear with the number of attributes. Finally, the efficiency of the proposed scheme is lower than schemes proposed in [2, 5] in terms of Test phase, but the proposed scheme is more features than the two schemes analyzed above. Generally, the proposed scheme is practical in that it can simultaneously support multi-keywords search, large universe and decryption service under the LSSS access structure.

5 Conclusions

In this paper, a flexible attribute-based keyword search scheme via two access policies scheme has been proposed. The proposed scheme guarantees that all the users whose credentials satisfy the data owner's access control policy can conduct keyword search operation and decrypt the retrieved encrypted files. Furthermore, it permits the data owner to control the access policy and delegate the search policy to data user who wants to search the interested files, hence it is more expressive. Besides, the construction employs an online/offline technique which spilt the computation of Index and TokenGen algorithm into two phases to improve the efficiency. And the decryption cost only needs two parings operation. Finally, the proposed scheme is proven selectively secure against chosen-plaintext attack and chosen-keyword attack, and it also achieves keyword secrecy.

Acknowledgements We would like to thank the anonymous reviewers for their valuable comments and suggestions. This work is supported by the National Natural Science Foundation of China under Grants No.61472470, the Science and Technology Bureau Project of Weiyang District of Xi'an City under Grants No.201403 and the National Natural Science Foundation of Shaanxi Province under Grants No.2014JM2-6091 and 2015JQ1007.

References

1. Alderman, J., Janson, C., Martin, K. M., & Renwick, S. L. (2015). Extended functionality in verifiable searchable encryption. In International Conference on Cryptography and Information Security in the Balkans (pp. 187-205).
2. Chaudhari, P., & Das, M. (2015). Privacy-preserving attribute based searchable encryption. IACR Cryptol ePrint Arch, 899.
3. Dong, Q., Guan, Z., & Chen, Z. (2015). Attribute-based keyword search efficiency enhancement via an online/offline approach. In Parallel and Distributed Systems (pp. 298-305).
4. Goyal, V., Pandey, O., Sahai, A., & Waters, B. (2006). Attribute-based encryption for fine-grained access control of encrypted data. In Proceedings of the 13th ACM conference on Computer and communications security (pp. 89-98).
5. Li, J., & Zhang, L. (2014). Attribute-based keyword search and data access control in cloud. In Computational Intelligence and Security (CIS)(pp. 382-386).
6. Jiang, P., Mu, Y., Guo, F., Wang, X., & Wen, Q. (2015). Online/offline ciphertext retrieval on resource constrained devices. The Computer Journal, bxv099.
7. Waters, B. (2011). Ciphertext-policy attribute-based encryption: An expressive, efficient, and provably secure realization. In Public Key Cryptography (pp. 53-70).
8. Padhya, M., & Jinwala, D. (2014). A novel approach for searchable CP-ABE with hidden ciphertext-policy. In International Conference on Information Systems Security (pp. 167-184).
9. Sahai, A., & Waters, B. (2005). Fuzzy identity-based encryption. In Annual International Conference on the Theory and Applications of Cryptographic Techniques (pp. 457-473).
10. Zheng, Q., Xu, S., & Ateniese, G. (2014). VABKS: verifiable attribute-based keyword search over outsourced encrypted data. In IEEE INFOCOM 2014-IEEE Conference on Computer Communications (pp. 522-530).

Large Universe Revocable Fine-Grained Encryption with Public Auditing

Xuewei Yan, Hua Ma, Zhenhua Liu and Ting Peng

Abstract Attribute-based encryption (ABE) allows for scalable and fine-grained data sharing in a cloud computing environment. However, most of existing ABE schemes with user revocation are not satisfactory on the efficiency side. In addition, since the data are stored on remote servers in the cloud storage environment, data owners do not know whether data is integrated in a timely manner. In this paper, we propose a novel large universe revocable fine-grained encryption with public auditing (LRA-FE) scheme based on prime-order bilinear groups. In this construction, we utilize extended proxy-assisted approach and appending redundancy approach, which weakens the trust of the cloud server. Furthermore, the proposed system introduces an auditor to inspect the integrity of data stored in the cloud. The size of attribute space can be exponentially large because our construction supports large universe. After comprehensive comparisons with the state-of-the-art works, the LRA-FE scheme features lightweight computation at the user side such that users can use resource-constrained devices to access cloud data.

Key words: Attribute-Based Encryption; Large Universe; Revocation; Audit.

Xuewei Yan
Xuewei Yan, School of Mathematics and Statistics, Xidian University, Xi'an 710071, P.R. China, e-mail: xweiyans@163.com

Hua Ma
Hua Ma, School of Mathematics and Statistics, Xidian University, Xi'an 710071, P.R. China. State Key Laboratory of Information Security (Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093), e-mail: hma@mail.xidian.edu.cn

Zhenhua Liu
Zhenhua Liu, School of Mathematics and Statistics, Xidian University, Xi'an 710071, P.R. China, e-mail: zhualiu@hotmail.com

Ting Peng
Ting Peng, School of Mathematics and Statistics, Xidian University, Xi'an 710071, P.R. China, e-mail: tpeeng@163.com

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_84

1 Introduction

With the rapid development of cloud computing, more and more individuals and enterprises outsource their sensitive data into the cloud server for alleviating the burden of maintaining huge data in local, while enjoying high quality data services. To ensure the privacy of user data, all the data is uploaded in encrypted form. In practice, one-to-many data sharing is very common. To achieve the flexibility and scalability of data sharing, a fine-grained access control is required.

This encryption notion, called ABE, was introduced by Sahai and Waters [6]. ABE has advantage over the traditional public key encryption as it achieves flexible one-to-many encryption instead of one-to-one. In ABE, a ciphertext can be decrypted using the corresponding decryption key only if the two match. Till now, there are two kinds of ABE: key-policy ABE (KP-ABE) and ciphertext-policy ABE (CP-ABE) [1]. In KP-ABE, the access control policy is embedded into users secret keys. Whereas, CP-ABE binds access policies with ciphertext.

However, there still exists some issues that hinder the application of ABE in the cloud computing, one of which is user revocation. User revocation is a critical requirement, particularly when there is a large number of users. When a user leaves the organization or is no longer involved in the project. User revocation would allow the data owner to revoke a user's ability to decipher the data. A series of approaches have been proposed to solve the above problem. For example, An approach is the key-update based revocation[3]. Nevertheless, this approach limits scalability as all data must be re-encrypted, and all non-revoked legitimate user keys need to be either updated or re-distributed. In order to solve existing some problems in above articles, proxy-assisted user revocation was introduced in [9]. In this approach, a cloud is regarded as a proxy, and the decryption is divided into two parts. The first part is cloud-side decryption by using user's cloud-side key. The final part is user-side decryption by utilizing user-side key. For user revocation, the cloud server will delete the cloud-side key associated with user who is revoked. To weaken the trust of the cloud server, Yang et al. [10] proposed an extended proxy-assisted approach. In their construction, the first partial decryption also requires the private key of cloud server. Accordingly, unless the dishonest cloud server is willing to disclose its private key, the leakage of a user's cloud-side key does not lead to the success of decryption. Nevertheless, the data owner plays the role of the distribution of the proxy keys in the scheme introduced in [10], which is a high requirement for the data owners and does not apply to the majority of general scenarios. The proposed system asks a attribute authority to act as the role for proxy key generation. In addition, we also utilize appending redundancy approach, the dishonest decryption of cloud server can be easily detected in our construction.

Another common problem is whether a system is a "small universe" or "large universe" construction. In "small universe" constructions, the size of the attribute space is polynomially bounded in the security parameter and the attributes were fixed at setup. In "large universe" schemes, The number of attributes is not polynomially bounded and any string can be used as an attribute, which is a desirable feature. The first large universe construction in the standard model were proposed in the work of

Lewko and Waters [2]. This is a large universe KP-ABE construction. Subsequently, Rouselakis and Waters [5] proposed a large universe ABE in the standard model and given a new proof techniques.
In addition, since the data are stored on remote servers in the cloud storage environment, data owners do not know whether data is intact intuitively. Some schemes [8, 7] are proposed to audit the integrity of data. In these schemes, The outsourced file is split into multiple blocks and each block related to an authenticator for integrity verifying. A third party auditor is introduced to execute the auditing tasks.
The main contributions of this paper can be summarized as follows:

1. In this paper, we propose a novel LRA-FE scheme on prime-order bilinear groups that simultaneously supports (1) immediate user revocation, (2) data auditing, (3) large attribute universe, (4) mitigation against cloud-user collusion, (5) key escrow-free, (6) no key update and data re-encryption. the proposed scheme features lightweight computation at the user side such that users can use resource-constrained devices to access cloud data.
2. The appending redundancy approach make that the dishonest decryption of cloud servers can be easily detected in our system.
3. We build our construction on the efficient prime order bilinear group

The rest of the paper is organized as follows. We formalize the notion and model of LRA-FE in Section 2. Our construction is detailed in Section 3. Security results together with performance comparisons are presented in Section 4. Finally, we conclude this paper in Section 5.

2 Problem Formulation

In this section, we give the formal definition of LRA-FE and define the system model. The definition of Bilinear Pairings, Access Structures and Linear Secret Sharing Schemes used in this scheme can be found in [4].

2.1 Definition of LRA-FE

Setup(1^λ): The Setup algorithm takes as input a security parameter 1^λ, and it outputs a master key *msk* and public parameters *pp*.

UKGen(u, pp): The user key generation algorithm takes as input a user identity u, the public parameters pp, and it outputs a key pair (pk_u, sk_u). cloud server(CS) and data owner(DO) also uses this algorithm to generate their key pairs,(pk_{CS}, sk_{CS}) and (pk_{DO}, sk_{DO}). Every system entity runs this algorithm to generate a key pair.

PxKGen$(msk, pk_{CS}, pk_u, S_u, pp)$: The proxy key generation algorithm takes as input the master key msk, the user public key pk_u, the CS public key pk_{CS}, a set of attributes S_u, the public parameters pp, and it outputs an proxy key PxK_u. Attribute authority(AA) runs this algorithm to authorize a user based on the user's attributes. The proxy key will be sent to CS who adds a new entry in its Proxy Key list $\mathscr{L}$.

Encrypt$(m, \mathbb{A}, pp)$: DO firstly appends the message m to be encrypted with a redundancy 0^k to obtain $m' = m \parallel 0^k$ where $\parallel$ is the concatenation of string. The encryption algorithm takes as input an access structure $\mathbb{A}$, a message, m', and the public parameters pp, and it outputs a ciphertext c.

AuthGen$(sk_{DO}, (c_\tau, id_\tau), pp)$: The authenticator generation algorithm takes as input the DO private key sk_{DO}, the public parameters pp, and (c_τ, id_τ), where id_τ is the identity of ciphertext block $c_\tau(\tau = 1, ..., n)$. and it outputs an authenticator σ_τ.

PxDec(c, sk_{CS}, PxK_u, pp): The proxy decryption algorithm takes as input a ciphertext c, the CS private key sk_{CS}, The proxy key PxK_u, the public parameters pp, and it outputs an intermediate v. CS runs this algorithm to help a user, u, partially decrypt a ciphertext.

UDec(sk_u, v): The user decryption algorithm takes as input an intermediate v, the user private key sk_u, and it outputs a message m'. The user continues to check whether a redundancy 0^k is appended with m'. If so (i.e.,$m' = m \parallel 0^k$), m is obtained through truncation; otherwise, a dishonest action of CS is detected.

Audit$(\sigma_\tau, pp, pk_{DO}, pp)$: The auditing algorithm takes as input the public parameters pp, the DO public key pk_{DO}, an authenticator σ_τ, and if the integrity proof for shared data pass the check, it outputs 1. Otherwise, it outputs 0. Third party auditor(TPA) runs this algorithm to audit the integrity proof for shared data.

Revoke$(u, \mathscr{L})$: The Revocation algorithm takes as input a user identity u, and Proxy key list $\mathscr{L}$, the algorithm revokes u's decryption capability by updating and outputting an updated Proxy key list, $\mathscr{L}'$.

2.2 System model

As depicted in Fig.1, our LRA-FE framework consists of five parties as follows:

- **Attribute Authority**: AA generates users' proxy keys. It sends users' proxy keys to CS.
- **Data Owner**: The users who outsource their encrypt data to CS for sharing with users. The data owner generates authenticators that are used to audit the integrity of data. Then they send authenticators to CS.
- **Cloud Server**: A party which provides storage service for cloud users and maintains a Proxy Key list, with each entry containing a users identity and the corresponding proxy key. When a user requests to retrieve a data record from the cloud, CS executes a proxy decryption operation.
- **Users**: The users who receive an intermediate value from CS and then decrypt it with his private key.

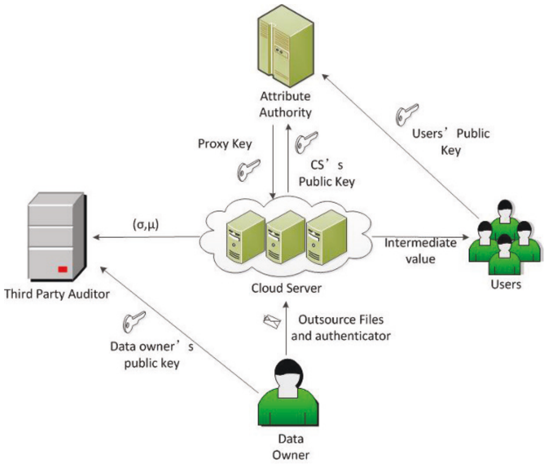


Fig. 1: System Model

- **Third Party Auditor**: An authority who performs data auditing process. It receives (σ,μ) and pk_{DO} from CS and DO respectively. Then it audits the integrity of data shared in the cloud.

3 Our Construction

Our system is based on the large universe CP-ABE scheme in [5] and uses the extended proxy-assisted approach in [10] and the data auditing approach in [7]. The new scheme is described as follows:

Setup(1^λ): On input a security parameter 1^λ, the algorithm gets the descriptions of the groups and the bilinear mapping $D=(p,G,G_T,e)$, where p is the prime order of the groups G and G_T. Selects a cryptographic hash function, $H:Z_p^*\to G$. The attribute universe is $\mathscr{U}=Z_p$. Then the algorithm picks the random terms $g,u,h,w,v\in G$ and $\alpha\in Z_p$, and sets $pp=(D,g,u,h,w,v,H,e(g,g)^\alpha)$ and $msk=(\alpha)$.

UKGen(u,pp): On input a user identity u, the algorithm chooses $x_u\in Z_p$ and sets $(pk_u=g^{x_u},sk_u=x_u)$. It can be seen that (pk_u,sk_u) is a standard ElGamal type key pair. CS and DO also use this algorithm to generate their key pairs, $(pk_{CS}=g^{x_{CS}},sk_{CS}=x_{CS})$ and $(pk_{DO}=g^{x_{DO}},sk_{DO}=x_{DO})$.

PxKGen$(msk,pk_{CS},pk_u,S_u=\{A_1,A_2,\ldots,A_k\},pp)$: On input the master key $msk=(\alpha)$, CS public key $pk_{CS}=g^{x_{CS}}$, a user public key $pk_u=g^{x_u}$ and a set of attributes S_u, the algorithm chooses $r,r',r_1,r_2,\ldots,r_k\in Z_p$, and sets $PxK_u=(K_0=(pk_{CS})^r(pk_u)^\alpha w^{r'},K_1=g^r,K_2=g^{r'},\forall i\in[k]:\{K_{i,1}=g^{r_i},K_{i,2}=(u^{A_i}h)^{r_i}v^{-r'}\})$.

Encrypt$(m,(M,\rho),pp)$: DO firstly appends the message m to be encrypted with a redundancy 0^k to obtain $m'=m\parallel 0^k$ where $\parallel$ is the concatenation of string. On input the message m' and the access structure encoded in an LSSS policy, with $M \in Z_p^{l\times n}$ and $\rho : [l] \to Z_p$. First, it selects $\mathbf{y} = (s, y_2, \ldots, y_n)^\top \in Z_p^{n\times 1}$, where s is the random secret to be shared among the shares. The vector of the shares is $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_l)^\top = M\mathbf{y}$. It then picks l random exponents $t_1, t_2, \ldots, t_l \in Z_p$ and sets the ciphertext $c = (C = m'e(g,g)^{\alpha s}, C_0 = g^s, \forall j \in [l] : C_{j,1} = w^{\lambda_j} v^{t_j}, C_{j,2} = (u^{\rho_j} h)^{-t_j}, C_{j,3} = g^{t_j}, C_{j,4} = g^{\lambda_j})$.

AuthGen$(sk_{DO}, (c_\tau, id_\tau), pp)$: On input $sk_{DO} = x_{DO}$ and (c_τ, id_τ), where id_τ is the identity of ciphertext block $c_\tau (\tau = 1, \ldots, n)$. DO computes an authenticator as follows:

$$\sigma_\tau = (H(id_\tau) \cdot u^{c_\tau})^{x_{DO}}$$

Finally, DO sends the data blocks along with the authenticators to CS.

PxDec(c, sk_{CS}, PxK_u, pp): On input CS private key $sk_{CS} = x_{CS}$, and the proxy key $PxK_u = (K_0, K_1, K_2, \forall i \in [k] : \{K_{i,1}, K_{i,2}\})$ associating with a set of attributes, S_u, and the ciphertext, $c = (C, C_0, \forall j \in [l] : \{C_{j,1}, C_{j,2}, C_{j,3}, C_{j,4}\})$, the proxy decryption algorithm calculates the set of rows in M that provide a share to attributes in S, i.e. $I = \{j : \rho(j) \in S\}$. Then it computes the constants $\{\omega_j \in Z_p\}_{j\in I}$ such that $\sum_{j\in I} \omega_j \mathbf{M}_j = (1, 0, \ldots, 0)$, where $\mathbf{M}_j$ is the j-th row of the matrix M. These constants exist if the set S is an authorized set of the policy. Then it computes

$$\frac{e(C_0, K_0)}{\prod_{j\in I}(e(C_{j,1}, K_2)e(C_{j,2}, K_{i,1})e(C_{j,3}, K_{i,2})e(C_{j,4}, K_1)^{x_{CS}})^{\omega_i}} = e(pk_u, g)^{s\alpha}$$

where i is the attribute $\rho(i)$'s index in S (it depends on i). Finally, it sets $v = (C, e(pk_u, g)^{s\alpha})$.

UDec(sk_u, v): On input a user private key, $sk_u = x_u$, and $v = (C, e(pk_u, g)^{s\alpha})$, the user decryption algorithm computes

$$\frac{C}{(e(pk_u, g)^{s\alpha})^{x_u^{-1}}} = m'$$

The user continues to check whether a redundancy 0^k is appended with m'. If so (i.e.,$m' = m \parallel 0^k$), m is obtained through truncation; otherwise, a dishonest action of CS is detected.

Audit$(\sigma_\tau, pp, pk_{DO}, pp)$: TPA randomly selects a challenge $ch = \{(id_\tau, f_\tau)\}_{\tau\in D}$ (where $D = \{s_1, s_2, \ldots, s_c\}$ is a c-element subset of set $[1, n]$ and $f_\tau \in Z_p^*$) and sends it to CS. CS first calculates an aggregated authenticator $\sigma = \prod_{\tau\in D} \sigma_\tau^{f_\tau}$. It also computes the linear combination of sampled blocks $\mu = \sum_{\tau\in D} f_\tau c_\tau$. Then CS sends $P = (\sigma, \mu)$ to TPA as the integrity proof of data storage.When TPA receives the (σ, μ), it verifies whether the following equation holds:

$$e(\prod_{\tau\in D} H(id_\tau)^{f_\tau} \cdot u^\mu, pk_{OD}) = e(g, \sigma)$$

If the equation holds, returns 1, otherwise returns 0.

Revoke$(u, \mathscr{L})$: On input a identity, u, and the Proxy Key list, $\mathscr{L}$, the user revoking algorithm deletes the entry corresponding to u from the list. i.e. $\mathscr{L}' = \mathscr{L} \setminus \{u, PxK_u\}$.

4 Discussion

In this subsection, Table 1 give the comparison between our work and several related work in terms of features and efficiency.

Table 1: Comparison of other mechanisms[1]

Schemes	[5]	[7]	[10]	[4]	Ours
Large Universe	$\surd$	$\times$	$\times$	$\surd$	$\surd$
Data Auditing	$\times$	$\surd$	$\times$	$\times$	$\surd$
Immediate User Revocation	$\times$	$\times$	$\surd$	$\times$	$\surd$
Mitigation Against Cloud-User Collusion	$\times$	$\times$	$\surd$	$\times$	$\surd$
Checkability	$\times$	$\times$	$\times$	$\surd$	$\surd$
PP	6	–	3	6	7
SK	$2\|S\|+2$	–	$2\|S\|+2$	$2\|S\|+2$	$2\|S\|+3$
CH	$3l+2$	–	$2l+3$	$5l+1$	$4l+2$
Dec(User)	$(3\|I\|+1)P+$ $\|I\|Exp$	–	Exp	3Exp	Exp

In [10], the system only supports user revocation and mitigation against cloud-user collusion. In [7], the system only supports the data auditing. In [5], the system only supports large universe. In [4], the construction supports largre universe and checkability. Our scheme achieves user revocation, data auditing, mitigation against cloud-user collusion and large universe simultaneously. In addition, the dishonest decryption of CS can be easily detected in our construction.

As shown in Table 1, compared with [5] and [4], our new system only needs an exponentiation in decryption phase. This makes our new systems more practical for applications. When compared with [10], our construction sacrifices tiny size of the public parameters, private key and ciphertext to achieve large universe, checkability and data auditing. Compared with [4], our scheme has shorter size of the ciphertext.

[1] Exp denotes one exponentiation in G and P denotes one pairing operation of bilinear map $\hat{e}$. And let PP stand for public parameter size, SK stand for decryption key size, CH stand for ciphertext size, Dec stand for pairing and exponentiation computations in user side decryption, $|S|$ be the size of the attribute set of a private key, $|I|$ be the number of attributes in a decryption key that satisfies a ciphertext's access policy, and l be the row size of access structures. Note that multiplication is negligible in this scheme.

5 Conclusions

In the paper, we propose a large universe revocable R-FE with public auditing in prime-order bilinear groups. It supports user revocation, data auditing and large universe simultaneously. Our system overcomes the limitation that data owner has to trust that the cloud does not disclose users' proxy keys. In addition, through appending redundancy, the dishonest decryption of CS can be easily detected.

6 Acknowledgements

We are grateful to the anonymous referees for their invaluable suggestions. This work is supported by the National Natural Science Foundation of China under Grants No.61472470, the Science and Technology Bureau Project of Weiyang District of Xi'an City under Grants No.201403 and the National Natural Science Foundation of Shaanxi Province under Grants No.2014JM2-6091 and 2015JQ1007.

References

1. Bethencourt, J., Sahai, A., Waters, B.: Ciphertext-policy attribute-based encryption. In: Proceedings of IEEE S and P, pp. 321-334, 2007
2. Lewko A, Waters B.: Unbounded HIBE and attribute-based encryption. Advances in Cryptology-EUROCRYPT 2011. Springer Berlin Heidelberg, PP. 547-567, 2011.
3. Liang, K., Liu, J.K., Wong, D.S., Susilo, W.: GO-ABE: an efficient cloud-based revocable identity-based proxy re-encryption scheme for public clouds data sharing. In: Proceedings of ESORICS 2014, pp. 257-272, 2014.
4. Ma H, Zhang R, Wan Z, et al.: Verifiable and Exculpable Outsourced Attribute-Based Encryption for Access Control in Cloud Computing. IEEE Transactions on Dependable and Secure Computing. DOI 10.1109/TDSC.2015.2499755.
5. Rouselakis Y, Waters B.: Practical constructions and new proof methods for large universe attribute-based encryption. Proceedings of the 2013 ACM SIGSAC conference on Computer and communications security. ACM, pp. 463-474, 2013.
6. Sahai A, Waters B.: Fuzzy identity-based encryption. Advances in Cryptology-EUROCRYPT 2005. Springer Berlin Heidelberg, pp. 457-473, 2005.
7. Yang G, Yu J, Shen W, et al.: Enabling public auditing for shared data in cloud storage supporting identity privacy and traceability. Journal of Systems and Software, 113: pp. 130-139, 2016.
8. Yang J J, Li J Q, Niu Y.: A hybrid solution for privacy preserving medical data sharing in the cloud environment. Future Generation Computer Systems, 43: pp. 74-86, 2015.
9. Yang, Y., Lu, H., Weng, J., Zhang, Y., Sakurai, K.: Fine-grained conditional proxy re-encryption and application. In: Chow, S.S.M., Liu, J.K., Hui, L.C.K., Yiu, S.M.(eds.) ProvSec 2014. LNCS, vol. 8782, pp. 206-222. Springer, Heidelberg 2014. Extended version to appear: Pervasive and Mobile Computing, ELSEVIER
10. Yang Y, Liu J K, Liang K, et al.: Extended Proxy-Assisted Approach: Achieving Revocable Fine-Grained Encryption of Cloud Data. Computer Security-ESORICS 2015. Springer International Publishing, pp. 146-166, 2015.

A privacy-preserving personal health record with searchability and revocability using attribute-based encryption

Shuhong Duan, Zhenhua Liu, and Peilin Zhou

Abstract Online personal health record (PHR) enables patients to store their health records at a cloud server and selectively share them with doctors for convenient medical care, and greatly facilitates the storage and sharing of the data. However, there are some challenges, such as risks of privacy exposure, flexible access, rapid ciphertext retrieval and efficient user revocation, impeding the development of PHR. As for solving the above problems, we propose a privacy-preserving personal health record with search and revocation. In our scheme, we obfuscates the attributes exposed in the access policy. Consequently, this avoids the risk that an attacker may guess what kind of disease the patient gets according to the attributes exposed in the access policy. Instead of an exhaustive search with cryptographic calculations, a simple comparison algorithm is adopted to improve the searching efficiency. Furthermore, our scheme supports dynamic user revocation through the cloud re-encrypts normal and unnormal ciphertexts according to the revocation list.

1 Introduction

With the development of cloud computing, cloud providers can provide more and more service projects, including Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service(SaaS). Among many cloud services,

Shuhong Duan
School of Mathematics and Statistics, Xidian University, Xi'an 710071,P.R. China. e-mail: 0801duanshuhong@163.com

Zhenhua Liu
School of Mathematics and Statistics, Xidian University, Xi'an 710071,P.R. China. e-mail: zhualiu@hotmail.com

Peilin Zhou
School of Mathematics and Statistics, Xidian University, Xi'an 710071,P.R. China. e-mail: plzhou1224@163.com

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_85

PHR service allows a patient to create, manage, control, and share his PHR in one place through the web, and has made the storage, retrieval, and sharing of the medical information more efficient. While it is exciting to have convenient PHR services in the cloud for everyone, there are many security and privacy risks, which could impede its wide application. Therefore, it is still essential to ensure the confidentiality of PHRs and proper access to PHRs.

A promising way to realize secure and privacy-preserving PHR system is to encrypt the data before storing it to the cloud server. Attribute-based encryption (ABE) just enables fine-grained access control on encrypted sensitive data. Thus, the concept of attribute-based encryption (ABE) can be used to encrypt personal data in PHR systems.

ABE is first introduced by Sahai and Waters [1]. In ABE schemes, only users whose attributes satisfy the specified access structure can decrypt the ciphertext. However, traditional ABE schemes can not satisfy many practical requirements. To solve these problems, several ABE schemes with different functionalities were proposed. In this paper, we focus on ABE with searchablity and user revocationablity.

In 2004, Bonech et al. first proposed the concept of Public Key Encryption with Keyword Search (PEKS)[2] to address the problem of rapid search on encrypted data. Then, for fine-grained access control, a series of schemes [3, 4] supporting attribute-based encryption with keyword search appear.

In addition, as for satisfying the practical requirement that a user's access right should be revoked if she/he is compromised or leaves the system, many revocable ABE schemes [5, 6, 7] come into being.

In this paper, we propose a privacy-preserving personal health record using attribute-based encryption with search and direct user revocation(PSR-CPABE). The main techniques and contributions are summarized as follows:

- In previous related schemes, the attributes in the access policy are exposed. However, our construction obfuscates the attributes exposed in the access policy and avoids the risk that an attacker may guess what kind of disease that a patient gets from the attributes exposed in the access policy.
- The proposed scheme improves the searching efficiency through adopting a simple comparison algorithm between an index and a trapdoor in the retrieval phase. While most previous schemes adopt an exhaustive search with cryptographic calculations.
- Our scheme simultaneously supports the properties of privacy protection of patients, rapid retrieval for the encrypted PHR and direct revocation of users, which make our scheme more applicable to the PHR system.

2 The security requirements

The proposed scheme satisfies the following security properties[8]:

- **Data confidentiality**: If a user's attributes don't satisfy the access policy, it is not allowed to access the encrypted PHR.
- **Anonymity**: Once the PHR is encrypted and stored to cloud server, no one can identify the actual identities of a PHR owner and user.
- **Controlled searching**: The cloud server can not search without user's authorization, and it learns nothing more than the search result about the ciphertext.
- **Collusion resistance**: the unauthorized users, whose attributes do not satisfy the access policy, can not decrypt the encrypted PHR. Even though multiple unauthorized users collaborate by combining their attributes such that these attributes satisfy the access policy, they still can not decrypt the encrypted PHR.

3 Our Construction

In this section, a privacy-preserving personal health record with search and revocation functions using attribute-based encryption is proposed. Now, we describe our detail construction as follows.

Setup(k): The algorithm takes as input a security parameter k. Let $\mathbb{G}$ and $\mathbb{G}_T$ be a bilinear group of prime order p, g be a generator of $\mathbb{G}$ and $e : \mathbb{G} \times \mathbb{G} \to \mathbb{G}_T$ be a bilinear map. Then, the algorithm randomly chooses $\alpha_1, \alpha_2, \beta \in \mathbb{Z}_p$ and sets $\alpha = \alpha_1 + \alpha_2, h = g^{\beta}$. Finally, the algorithm selects a hash function $H : \{0,1\}^* \to \mathbb{G}$. The public parameters are published as:

$$PP =< \mathbb{G}, \mathbb{G}_T, g, h, e(g,g)^{\alpha}, H > .$$

The master secret key is set as:

$$MSK =< \beta, g^{\alpha}, \alpha_1, \alpha_2 > .$$

KeyGen$_u(MSK, PP, S)$: The algorithm takes as input the master secret key MSK, the public parameters PP and an attribute set S belonging to a user u. Given a user u, the algorithm randomly chooses $r \in \mathbb{Z}_p$. For each attribute $\lambda_j \in S$, the algorithm randomly chooses $r_j \in \mathbb{Z}_p$. The private key SK is set as:

$$SK =< D = g^{\frac{\alpha_1 + r}{\beta}}, \{D_j = g^r H(\lambda_j)^{r_j}, D'_j = g^{r_j}, D''_j = H(\lambda_j)^{\beta}\}_{\lambda_j \in S} > .$$

KeyGen$_c(MSK, PP)$: The algorithm is used to generate a delegation key for the cloud server using the mast secret key MSK. The delegation key is set as:

$$SK_c =< D_c = g^{\alpha_2/\beta} > .$$

KeyGen$_o(MSK, PP, ID_0)$: The algorithm is used to generate an anonymous key which works for obfuscating attributes exposed in the access tree. Given an identity ID_0 of a PHR owner, on input the mast secret key MSK and the public parameters PP, the algorithm computes the anonymous key as follows:

$$A_0 = H(ID_0)^{\beta}.$$

Encrypt(PP, m, T): m denotes PHR data and T is an access tree. Let Y be the set of leaf nodes in the access tree T. For each node x (including the leaves) in the tree T, the algorithm chooses a polynomial q_x same as that in the Bethencourt et al.'s scheme [9]. Then, the algorithm chooses a random $t \in \mathbb{Z}_p$ and generates a pseudonym $P_0 = H(ID_0)^t$ for a PHR owner, where ID_0 is the PHR owner's identity. Finally, the ciphertext is constructed as follows:

$$CT =< \mathrm{T}, C = m \cdot e(g,g)^{\alpha s}, C' = h^s, C'' = P_0,$$

$$\{C_y = g^{q_y(0)}, C'_y = H(attr_y)^{q_y(0)}\}_{y \in Y} > .$$

AttScm(CT, A_0, PP): To obfuscate all attribute values in the access tree T, the algorithm takes the ciphertext CT, the anonymous key A_0 and the parameters PP as input, and computes the obfuscated attribute set as follows:

$$K_{o,\mathbf{Y}} = \{e(A_0^t, H(\lambda_j))\}_{\lambda_j \in \mathbf{Y}} = \{e(H(ID_0), H(\lambda_j)^{\beta t}\}_{\lambda_j \in \mathbf{Y}},$$

where t is chosen in the encryption algorithm.

Then, as depicted in Fig.2, the algorithm replaces the attribute λ_j exposed in the access tree T with the obfuscated attribute value $scm_{attr_x} \in K_{o,\mathbf{Y}}$, where λ_j and $attr_y$ can be exchanged each other since they both represent the same attribute corresponding to a leaf node in the access tree T.

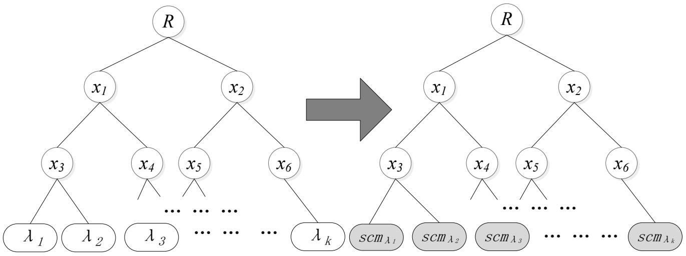


Fig. 1 Attribute Scrambling

Finally, replaced the access tree T in ciphertext CT with the new access tree T$'$, the new ciphertext is set as:

$$CT' =< \mathrm{T}', C, C', C'', \{C_y, C'_y\}_{y \in Y} >,$$

where Y is the set of original leaf nodes in the access tree T.

Then, the PHR owner uploads CT' to the cloud server.

Re-encrypt(RL): Let $\mathbf{Y} = \{\lambda_i, \ldots \lambda_k\}$ be the attribute set corresponding to the leaf nodes in the access tree T, where $1 \le i \le k \le |\mathbb{L}|$.
Suppose that the revocation list is $RL = \{ID_1, ID_2, \ldots, ID_n\}$, where ID_i is the identity of user u_i. The cloud server re-encrypts the ciphertext CT' as follows:

- If a user's identity $ID \in RL$, it randomly selects k and $k' \in \mathbb{Z}_p$, and computes CT'' as:

$$CT'' =< \mathrm{T}', C = m \cdot e(g,g)^{\alpha sk'}, C' = h^{sk}, D'_c = D^k_c, C'' = P_0,$$

$$\{C_y = g^{q_y(0)k}, C'_y = H(attr_y)^{q_y(0)k}\}_{y \in Y} > .$$

- If a user's identity $ID \notin RL$, it randomly selects $k \in \mathbb{Z}_p$ and computes CT'' as:

$$CT'' =< \mathrm{T}', C = m \cdot e(g,g)^{\alpha sk}, C' = h^{sk}, D'_c = D^k_c, C'' = P_0,$$

$$\{C_y = g^{q_y(0)k}, C'_y = H(attr_y)^{q_y(0)k}\}_{y \in Y} > .$$

Query(PP, SK, P_0): The algorithm takes the public parameters PP, a private key SK and a PHR owner's pseudonym P_0 as input. Since the cloud server can extract a PHR owner's pseudonym $P_0 \in CT'$ according to the stored ciphertext CT', without learning her/his any actual identity, a user can acquire a pseudonym of a PHR owner from the cloud server or the PHR owner. Once a user determines to retrieve and access the ciphertext of some PHR owner whose pseudonym is $C'' = P_0$, the algorithm generates a trapdoor corresponding attributes as follows:

$$K_{o,S} = \{e(D''_j, C'')\}_{j \in S} = \{e(H(\lambda_j), H(ID_0)^{\beta t}\}_{\lambda_j \in S}.$$

Then, the user sends the trapdoor to the cloud server as a query for data request.

Retrieve$(K_{o,S})$: On input a trapdoor $K_{o,S}$ with a set of scrambled attributes, the cloud server first checks whether the requested ciphertext is stored in the cloud and which one satisfies with the requested index terms according to a simple comparison algorithm $\mathbf{C}(\mathrm{T}', K_{o,S})$, which returns boolean value: "true" or "false". Let T'_x be a subtree of the access tree T' rooted at the node x and X' be a set of children whose parent is the node x, that is $X' = \{x | parent(x') = x\}$. We compute the boolean value of the comparison algorithm $\mathbf{C}(\mathrm{T}', K_{o,S})$ recursively as follows:

- If x is a leaf node, $\mathbf{C}(\mathrm{T}'_x, K_{o,S})$ returns "true" if and only if $scm_{attr_x} \in K_{o,S}$.
- If x is a non-leaf node in T', $\mathbf{C}(\mathrm{T}'_x, K_{o,S})$ returns "true" if and only if $|\{x' | x' \in X'$ and $\mathbf{C}(\mathrm{T}'_x, K_{o,S}) = \text{“}true\text{”}\}| \ge k_x$. In other words, $\mathbf{C}(\mathrm{T}'_x, K_{o,S})$ returns "true" if and only if at least k_x children return "true".

For each re-encrypted ciphertext CT'', the cloud server simply follows the access tree T' and determines whether $\mathbf{C}(\mathrm{T}'_x, K_{o,S})$="true" or not. If the boolean value returns "true", which means the ciphertext a user required exists. Then, the cloud server sends the corresponding ciphertext to the user. Otherwise, it outputs "false".

Decrypt(CT'', SK): The first part of the decryption proceeds the same as in the Bethencourt et al.'s scheme [9]. If the user's attributes satisfy the access tree, the component of decryption can be recovered as follows:

$$A = DecryptNode(CT', SK, R) = e(g,g)^{rks}.$$

The second part of the decryption proceeds as the following: If the user is not in the revoked list, the message m can be recovered by:

$$m = \frac{C}{(e(C', D) \cdot e(C, D'_c))/A}.$$

4 Security Requirements

Now, we show that the proposed scheme satisfies the security properties described in section 3.3.

- **Data confidentiality**: Firstly, if a user's attributes don't satisfy the access tree T, the user will not recover the value $e(g,g)^{ras}$, which leads the ciphertext not to be decrypted. Secondly, when a user is revoked from the system, the cloud server chooses two different random numbers k and k' to re-encrypt the ciphertext, which leads the re-encrypted ciphertext not to be decrypted. Note that we don't consider the cloud server colludes with the revoked users.
- **Anonymity**: The identity of a PHR owner in the ciphertext is replaced by a pseudonym $P_0 = H(ID_0)^t$. In addition, a user's identity is not contained in any algorithms of the proposed scheme. Therefore, no one knows the actual identities of the PHR owner and a user.
- **Controlled searching**: In our construction, the cloud server searches the requested ciphertext using a trapdoor $K_{o,S}$. Therefore, without the trapdoor $K_{o,S}$ that a user authorizes, the cloud server can not search. Furthermore, On input a trapdoor $K_{o,S}$, it searches the requested ciphertext according to a simple comparison algorithm $\mathbf{C}(\mathrm{T}', K_{o,S})$, which returns the boolean value: "true" or "false". Thus, the cloud server learns nothing more than the search result about the ciphertext.
- **Collusion resistance**: In the phase of generating secret key, the attribute authority chooses different values of r for different users, which leads that unauthorized users can't derive the message m. To decrypt a ciphertext, we needs to recover $e(g,g)^{\alpha s}$. Thus, the attacker must pair C_x from the ciphertext with D_i from the other colluding user's secret key for an attribute i that the attacker does not hold. However, every user's private key is uniquely generated by a random r. Thus, even if the colluding users are all valid, the attacker can not recover $e(g,g)^{\alpha s}$.

5 Efficiency

In this section, Table 1 and 2 give the comparison between our work and several related works in terms of functionalities (i.e. Search, Revocation, etc.) and performance. For convenience, E and P respectively represents exponent operation and pairing operation; $|S|$ stands for the size of a user's attribute set in the private key. $|Y|$ is the number of the leaf nodes in the access tree T.

Table 1 Functionality comparison

Schemes	Search	Revocation	Attributes obfuscating in the access tree
[4]	$\surd$	$\times$	$\surd$
[7]	$\times$	$\surd$	$\times$
Ours	$\surd$	$\surd$	$\surd$

Table 2 Efficiency comparison

Schemes	Encryption (with AttriScm)	Query	Re-encryption	Decryption
[4]	$(3+2\|Y\|)E$ $+\|Y\|P$	$\|S\|E$	–	$O(\|S\|E)$ $+(1+2\|S\|)P$
[7]	$(2+2\|Y\|)E$	–	$(3+2\|Y\|)E$	$O(\|S\|E)$ $+(2+2\|S\|)P$
Ours	$(3+2\|Y\|)E$ $+\|Y\|P$	$\|S\|E$	$(3+2\|Y\|)E$	$O(\|S\|E)$ $+(2+2\|S\|)P$

In Table 1, our scheme has advantages over the existing schemes [4] and [7] in that our scheme simultaneously supports search, revocation and attributes scrambling in the access tree. While Koo et al.'s scheme [4] doesn't support the revocability and Xu et al.'s scheme [7] doesn't support the search and attributes scrambling in the access tree.

Table 2 shows that the efficiencies of the proposed scheme for encryption and decryption are almost same as that of Koo et al's scheme [4] and Xu et al's scheme [7]. In addition, the proposed scheme compares with the schemes [4] and [7] respectively, there is the cost of the query and re-encryption. However, our scheme only sacrifices very low cost of the query and re-encryption to achieve search and revocation. In terms of the practicability and application, it is worth to construct a privacy-preserving attribute-based encryption scheme supporting search and revocation.

6 Conclusion

In this paper, we have proposed a privacy-preserving personal health record with search and revocation functions using attribute-based encryption. Compared with the existing schemes, our scheme has the following advantages: (1) Obfuscates the attributes exposed in the access policy and prevents the risk that an attacker may guess what kind of disease that the patient gets from the attributes exposed in the access policy; (2) Improves the searching efficiency through adopting a simple comparison algorithm in the matching verification phase; (3) Supports dynamic user revocation.

7 Acknowledgments

We would like to thank the anonymous reviewers for their valuable comments and suggestions. This work is supported by the National Natural Science Foundation of China under Grants No.61472470, the National Natural Science Foundation of Shaanxi Province under Grants No.2014JM2-6091 and 2015JQ1007.

References

1. Sahai, A., & Waters, B. (2005, May). Fuzzy identity-based encryption. In Annual International Conference on the Theory and Applications of Cryptographic Techniques (pp. 457-473). Springer Berlin Heidelberg.
2. Boneh, D., Di Crescenzo, G., Ostrovsky, R., & Persiano, G. (2004, May). Public key encryption with keyword search. In International Conference on the Theory and Applications of Cryptographic Techniques (pp. 506-522). Springer Berlin Heidelberg.
3. Dong, Q., Guan, Z., & Chen, Z. (2015, December). Attribute-based Keyword Search Efficiency Enhancement Via an Online/Offline Approach. In Parallel and Distributed Systems (ICPADS), 2015 IEEE 21st International Conference on (pp. 298-305). IEEE.
4. Koo, D., Hur, J., & Yoon, H. (2013). Secure and efficient data retrieval over encrypted data using attribute-based encryption in cloud storage. Computers & Electrical Engineering, 39(1), 34-46.
5. Zu, L., Liu, Z., & Li, J. (2014, September). New Ciphertext-Policy Attribute-Based Encryption with Efficient Revocation. In Computer and Information Technology (CIT), 2014 IEEE International Conference on (pp. 281-287). IEEE.
6. Xie, X., Ma, H., Li, J., & Chen, X. (2013). An Efficient Ciphertext-Policy Attribute-Based Access Control towards Revocation in Cloud Computing. J. UCS, 19(16), 2349-2367.
7. Xu, Z., & Martin, K. M. (2012, June). Dynamic user revocation and key refreshing for attribute-based encryption in cloud storage. In 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications (pp. 844-849). IEEE.
8. Li, J., & Zhang, L. (2014). Attribute-based keyword search and data access control in cloud. In Computational Intelligence and Security (CIS)(pp. 382-386).
9. Bethencourt, J., Sahai, A., & Waters, B. (2007, May). Ciphertext-policy attribute-based encryption. In 2007 IEEE symposium on security and privacy (SP'07) (pp. 321-334). IEEE.

Security Analysis and Improvement of A Collusion-Resistant Identity-Based Proxy Re-Encryption Scheme

Linchao Zhang, Hua Ma, Zhenhua Liu, and Enting Dong

Abstract An identity-based proxy re-encryption scheme allows a semi-trusted proxy to convert an encryption under Alice's identity into the encryption under Bob's identity. The proxy does not know the secret key of Alice or Bob, neither does the plaintext during the conversion. In an identity-based proxy re-encryption scheme, the collusion of the proxy and a delegatee may decrypt the ciphertext for a delegator if the ciphertext is re-encrypted. So it is important to resist the collusion attack in the identity-based proxy re-encryption scheme. In 2015, Qiu et al. proposed an identity-based proxy re-encryption without random oracles, and claimed that their scheme can resist against the collusion attack. However, we analyze the security of Qiu et al.'s scheme and show that the claim is incorrect. In this paper, we propose an improved scheme, which is secure against collusion attack and chosen ciphertext attack in the standard model.

Linchao Zhang
School of Mathematics and Statistics, Xidian University, Xi'an, Shaanxi 710071, P.R.China. e-mail: linchao _zhang00@163.com

Hua Ma
School of Mathematics and Statistics, Xidian University, Xi'an, Shaanxi 710071, P.R.China. e-mail: ma_hua@126.com
State Key Laboratory of Information Security (Institute of Information Engineering, Chinese A-cademy of Sciences, Beijing 100093)

Zhenhua Liu
School of Mathematics and Statistics, Xidian University, Xi'an, Shaanxi 710071, P.R.China. e-mail: zhualiu@hotmail.com

Enting Dong
School of Mathematics and Statistics, Xidian University, Xi'an, Shaanxi 710071, P.R.China. e-mail: dongenting@126.com

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_86

1 Introduction

In 1998, Blaze et al. [5] first proposed the primitive of proxy re-encryption , in which a proxy with re-encryption key can translate a ciphertext for the original decryptor, Alice, to another ciphertext with the same plaintext for the delegated decryptor, Bob. However, The proxy cannot access the plaintext. PRE schemes have many applications, such as digital rights management, user revocation, private remote data integrity checking, secure could email [1-4].

The first identity-based proxy re-encryption scheme was proposed by Green and Ateniese [9] in 2007. Their scheme allows a proxy to convert the ciphertext under Alice's identity into the ciphertext under Bob's identity. The above scheme is based on Boneh-Franklin's identity-based encryption scheme [8], which was shown to be secure in the random oracles model. Subsequently, Chu and Tzeng [6] proposed two identity-based proxy re-encryption schemes without random oracles. Both of them satisfy the properties of unidirectionality, non-interactivity and multi-use. But this two schemes are vulnerable to collusion attack.

In the process of an identity-based proxy re-encryption scheme, the security properties of proxy must be carefully considered. Evidently, the collusion of the proxy and a malicious delegatee may decrypt any other ciphertext for a delegator if the ciphertext is re-encrypted. So it is important to resist the collusion attack in the identity-based proxy re-encryption scheme. In 2015, Qiu et al. [7] proposed an identity-based proxy re-encryption scheme without random oracles. they claimed their scheme is secure against the collusion attack.

In this paper, we give a security analysis of Qiu et al.'s scheme and show that their scheme is still vulnerable to collusion attack. To address the issue, we propose an improved scheme. In the improved scheme, we get the security by changing the re-encryption key. What is more, although we reduce secret parameter, our scheme is still secure against chosen ciphertext attack in the standard model.

Organization

This paper is organized as follows. In Section 2, we give some preliminaries, including bilinear groups and the definition of identity-based proxy re-encryption. In Section 3, we briefly review Qiu et al.'s scheme and give its security analysis. An improved scheme was proposed in Section 4. In Section 5, we analyze the improved scheme's security and give the security proof. Finally, the conclusion is presented in Section 6.

2 Preliminaries

In this section, we will give some preliminaries.

2.1 Bilinear groups

Let $\mathbb{G}$ and $\mathbb{G}_T$ be multiplicative cyclic bilinear group of prime order p and g be a generator of $\mathbb{G}$. A bilinear map e is a map function $e : \mathbb{G} \times \mathbb{G} \to \mathbb{G}_T$ with the following properties:

1. Bilinearity: For all $u, v \in \mathbb{G}$, and all $a, b \in \mathbb{Z}_p$, $e(u^a, v^b) = e(u, v)^{ab}$.
2. Non-degeneracy: $e(g, g) \neq 1$.
3. Computability: There exsits an efficient algorithm to compute the pairing.

We say that $\mathbb{G}$ is a bilinear group and e is a bilinear pairing in $\mathbb{G}$.

2.2 Identity-Based Proxy Re-encryption

Definition 1. An identity-based proxy re-encryption scheme consists of six algorithms as follows[10]:

- ***Setup***(1^λ): On input a security parameter, the public parameter u and master secret key msk are outputted.
- ***KeyGen***(u, msk, id): On input the master secret key msk and an identity id, output the decryption key sk_{id}.
- ***Encryption***(u, id, m): On input an identity id and a message m, output the ciphertext C_{id}.
- ***RkGen***$(u, sk_{id_1}, id_1, id_2)$: On input a decryption key sk_{id_1} and identities id_1 ,id_2 output the re-encryption key $rk_{id_1 \to id_2}$.
- ***Reencryption***$(u, rk_{id_1 \to id_2}, C_{id_1})$: On input a re-encryption key $rk_{id_1 \to id_2}$ and a ciphertext C_{id_1}, output the re-encrypted ciphertext C_{id_2}.
- ***Decryption***(u, sk_{id}, C_{id}): On input a private key sk_{id} and a ciphertext C_{id}, output the plaintext m or $\perp$.

Correctness. Suppose(u, msk) properly generated by ***Setup***(1^λ), for all identity id, ***KeyGen***$(u, mk, id) \to sk_{id}$, ***RkGen***$(u, id, m) \to rk_{id_1 \to id_2}$. Let C_{id} be the ciphertext output from ***Encryption***(u, id, m) or ***Reencryption***$(u, rk_{id_1 \to id_2}, C_{id})$. Then the following equations hold:

$$m = \textbf{\textit{Decryption}}(u, sk_{id_1}, C_{id_1}),$$
$$m = \textbf{\textit{Decryption}}(u, sk_{id_2}, C_{id_2}).$$

2.3 Decisional Bilinear Diffie-Hellman Assumption

The challenger chooses $a, b, c, z \in_R Z_p$ at random and then flips a fair binary coin β. If $\beta = 1$, it outputs the tuple $(g, g^a, g^b, g^c, Z = e(g, g)^{abc})$. Otherwise, if $\beta = 0$, the challenger outputs the tuple $(g, g^a, g^b, g^c, Z = e(g,g)^z)$. The adversary must then output a guess β' of β.

An adversary, $\mathscr{A}$, has at least an ε advantage in solving the DBDH problem if

$$|\Pr[(\mathscr{A}(g, g^a, g^b, g^c, Z = e(g,g)^{abc}) = 0]$$
$$- \Pr[\mathscr{A}(g, g^a, g^b, g^c, Z = e(g,g)^z) = 0]| \geq 2\varepsilon$$

where the probability is over the randomly chosen a, b, c, z and the random bits consumed by $\mathscr{A}$.

Definition 2. The (t, ε)-DBDH assumption holds if no t time adversary has at least ε advantage in solving the game [10].

3 Analysis of Collusion Attack of Qiu et al.'s Scheme

We simply review Qiu et al.'s scheme. In Qiu et al.'s scheme, *Setup* algorithm generates the public parameter $u = (g, g_1, g_2, F_1(), F_2(), (\mathbb{G}, Sign, Vrfy))$ and master secret key $msk = (g_2^\alpha, g_2^\beta)$, where g is generator of $\mathbb{G}$, $g_1 = g^\alpha \cdot g^\beta$, g_2 randomly. *KeyGen* algorithm generates the private decryption key $sk_{id} = (g_2^\alpha F(id)^r, g_2^\beta F(id)^r, g^r) = (sk_1, sk_2, sk_3)$, where id is an identity and $r \in_R Z_p$. *RkGen* algorithm generates the re-encryption key $rk_{id_1 \to id_2} = ((sk_1 \cdot sk_2)K^{-1}, sk_3, R)$, where $K = E_2(k) \in \mathbb{G}$, $R \leftarrow Encryption(u, id_2, k)$.

Suppose the proxy and a malicious delegatee conspire.

- First of all, the malicious delegatee's identity id_2 is known, u is a public parameter, and R is the ciphertext which is encrypted k under delegatee's identity. So k's value can be easily obtained by decrypting R using delegatee's secret key.
- Then by computing $E_2(k)$, K's value can be computed.
- And then the re-encryption key $rk_{id_1 \to id_2}$ is known to the proxy, which also means that sk_3 is known to the proxy. Therefore $(sk_1 \cdot sk_2)$ can be obtained through the equation $rk_{id_1 \to id_2} = ((sk_1 \cdot sk_2)K^{-1}, sk_3, R)$.
- At last, the original ciphertext

$$C = (M \cdot e(g_1, g_2)^t, g^t, F_1(id)^t, F_2(vk)^t) = (c_1, c_2, c_3, c_4)$$

is known to the proxy, so the proxy and a malicious delegatee conspire, M can be recovered by the following derivation methods:

$$M = c_1 \cdot \frac{e(sk_3, c_3) \cdot e(sk_3, c_3)}{e(sk_1 \cdot sk_2, c_2)}$$

That is to say, the collusion of the proxy and a delegatee can decrypt any other ciphertext for a delegator if the ciphertext is re-encrypted. So Qiu et al.'s scheme cannot resist collusion attack.

4 The Improved Identity-Based Proxy Re-encryption Scheme Algorithms

To resist the collusion attack , we improve Qiu et al.'s scheme, The improved scheme also consists of six algorithms as follows:

- ***Setup***(1^λ) : On input a security parameter 1^λ, randomly choose two groups $\mathbb{G}$ and $\mathbb{G}_T$ with prime order p, a bilinear map e and a generator g defined above. Let $L \leq |p| - 2$, $E_1 : \{0,1\}^{l+1} \rightarrow \mathbb{G}_T$, and $E_2 : \{0,1\}^l \rightarrow \mathbb{G}$ be two encodings. Let $\alpha \in Z_p$ be a randomly chosen secret. And Set the value $g_1 = g^\alpha$ and choose the value $g_2 \in \mathbb{G}$ randomly. Let v, w be two n-bit string and V, W be the set of all i for which i-th bit of v and w is one. Define two functions $F_1(v) = u'_1 \prod_{i \in v} u_{1,i}$, $F_2(w) = u'_2 \prod_{i \in v} u_{2,i}$, where $u'_1, u_{1,1}, u_{1,2}, ..., u_{1,n}$ and $u'_2, u_{2,0}, u_{2,1}, ..., u_{2,n}$ are randomly chosen from $\mathbb{G}$. Let $(\mathbb{G}, Sign, Vrfy)$ as a one-time signature scheme. Then the master secret key is set as $msk = g_2^\alpha$ and public parameter is
$$u = (g, g_1, g_2, F_1(), F_2(), (\mathbb{G}, Sign, Vrfy)).$$
- ***KeyGen***(u, msk, id): On input the master secret key msk and an identity id, output the decryption key $sk_{id} = (g_2^\alpha F_1(id)^r, g^r)$, where $r \in_R Z_p$.
- ***Encryption***(u, id, m): Perform $\mathbb{G}(1^{\lambda'})$ to get verification key (vk, sk). For an identity id and a message $m \in \{0,1\}^l$, compute
$$\tilde{C} = (M \cdot e(g_1, g_2)^t, g^t, F_1(id)^t, F_2(vk)^t),$$
where $t \in_R Z_p$ and $M = E_1(m||0)$. Then compute $\sigma = Sign_{sk}(\tilde{C})$. Output the ciphertext $C_{id} = (\tilde{C}, vk, \sigma)$.
- ***RkGen***$(u, sk_{id_1}, id_1, id_2)$: Let $sk_{id_1} = (sk_1, sk_2)$. Compute the re-encryption key for id_2 as
$$rk_{id_1 \rightarrow id_2} = (sk_1 \omega K^{-1}, sk_2, R, Q),$$
where $k \in_R \{0,1\}^l$, $K = E_2(k) \in \mathbb{G}$, $R \leftarrow Encryption'(u, id_2, k)$ and
$$Q \leftarrow Encryption'(u, id_2, (e(\omega, g)||e(\omega, g^t))),$$
where $\omega \in \mathbb{G}$, $e(\omega, g), e(\omega, g^t) \in \mathbb{G}_T$.
We define $Encryption'$ the same as $Encryption$ except that it appends '1' to the message m.
- ***Reencryption***$(u, rk_{id_1 \rightarrow id_2}, C_{id_1})$: Let $rk_{id_1 \rightarrow id_2} = (sk'_1, sk_2, R, Q)$ and $C_{id_1} = (c_1, c_2, c_3, c_4, vk, \sigma)$. Check if
$$Vrfy_{vk}((c_1, c_2, c_3, c_4), \sigma) \stackrel{?}{=} 1.$$
If not, output $\perp$. Otherwise, compute
$$C_{id_2} = (C_{id_1}, R, Q, {sk_1}' F_2(vk)^{r'}, sk_2, g^{r'})$$

as the second-level ciphertext, where $r' \in_R Z_p$.

– ***Decryption***(u, sk_{id}, C_{id}): The decryption is proceeded as follows:

(a) If C_{id} is a regular encryption, let $sk_{id_1} = (sk_1, sk_2)$ and $C_{id} = (c_1, c_2, c_3, c_4, vk, \sigma)$. Check if

$$Vrfy_{vk}((c_1, c_2, c_3, c_4), \sigma) \overset{?}{=} 1.$$

If not, output $\perp$. Otherwise, compute $sk_1'' = sk_1 F_2(vk)^{r'}$, $sk_2' = g^{r'}$, where $r' \in_R Z_p$. Then compute

$$M = c_1 \frac{e(sk_2, c_3) e(sk_2', c_4)}{e(sk_1'', c_2)}.$$

(b) If C_{id} is a re-encryption ciphertext, let $C_{id} = (C_{id_1}, R, Q, sk_1{}'', sk_2, sk_2')$ and $C_{id_1} = (c_1, c_2, c_3, c_4, vk, \sigma)$. Check if

- $Vrfy_{vk}((c_1, c_2, c_3, c_4), \sigma) \overset{?}{=} 1.$
- $e(sk_1'' K, g) \overset{?}{=} e(g_1, g_2) e(F_1(id_1), sk_2) e(\omega, g) e(F_2(vk), sk_2').$

If not, output$\perp$. Otherwise, compute

$$M = c_1 \frac{e(sk_2, c_3) e(sk_2', c_4) e(\omega, g^t)}{e(sk_1'' K, c_2)}.$$

where $K = E_2(k), k \leftarrow Decryption'(u, id_2, R)$ and

$$(e(\omega, g) || e(\omega, g^t)) \leftarrow Decryption'(u, id_2, Q).$$

Compute $m||b = E_1(M)^{-1}$. If $b = 0$, output m. Otherwise, output $\perp$. We define *Decryption'* the same as *Decryption* except that it outputs m if the decrypted message ends with '1' and outputs $\perp$ if it end with '0'.

5 Security

5.1 Analysis of Collusion Attack of The Improved Scheme

Suppose the proxy and a malicious delegatee conspire.

- To begin with, the malicious delegatee's identity id_2 is known, u is a public parameter, R is the ciphertext which is encrypted k under delegatee's identity, and Q is the ciphertext which is encrypted $(e(\omega, g)||e(\omega, g^t))$ under delegatee's identity. So k's and $(e(\omega, g)||e(\omega, g^t))$'s value can be easily obtained by decrypting R and Q using delegatee's secret key, respectively. But ω can not be obtained, according to the nature of bilinear map e.
- Next by computing $E_2(k)$, K's value can be computed.

- Then the re-encryption key $rk_{id_1 \to id_2}$ is known to the proxy, which also means that sk_2 is known to the proxy. Therefore $(sk_1\omega)$ can be obtained through the equation $rk_{id_1 \to id_2} = (sk_1\omega K^{-1}, sk_2, R, Q)$.
- Last but not least, the original ciphertext

  $$C = (M \cdot e(g_1,g_2)^t, g^t, F_1(id)^t, F_2(vk)^t) = (c_1,c_2,c_3,c_4)$$

  is known to the proxy, so if the proxy and a malicious delegatee conspire, the equation below can be obtained.

  $$c_1 \cdot \frac{e(sk_2,c_3)}{e(sk_1\omega,c_2)} = \frac{M}{e(\omega,g^t)}$$

On the one hand, t is randomly chosen, in the encryption stage, so g^t is also random. On the other hand, ω is unknown, so it is impossible that $e(\omega,g^t)$ can be directly computed. Meanwhile, the malicious delegate only knows that the delegator gives him $e(\omega,g^t)$, but does not know others. Accordingly, they also can not recover M. Our scheme can resist collusion attack.

5.2 The Security Proof of The Improved Scheme

Our scheme can be proved secure according to [7]. For limited space, we will not show the details.

5.3 Comparison

Table 1 Security Comparison

Schemes	Security	Standard model	collusion resistance
[7,8]	CCA	YES	NO
[10]	CCA	NO	NO
Our	CCA	YES	YES

Table 1 shows the security comparison between [7,8,10] and ours. From the table we can see the four schemes are secure against CCA. Compared to [7,8], although the two schemes and ours can be proved CCA secure in the standard mode, our scheme is secure against collusion attack. As for [10], our scheme not only achieves CCA security without random oracles, but also resists collusion attack.

6 Conclusion

In this paper, we have analyzed the security of Qiu et al.'s scheme and have shown that the scheme is not secure against the collusion attack. Aiming at the limitation of Qiu et al.'s scheme, we have proposed an improved scheme to overcome the security weakness. As a result of security analysis, the improved scheme is secure against the collusion attack and chosen ciphertext attack in the standard model.

Acknowledgments

We would like to thank the anonymous reviewers for their valuable comments and suggestions. This work is supported by the National Natural Science Foundation of China under Grants No.61472470, the Science and Technology Bureau Project of Weiyang District of Xian city under Grants No.201403 and the National Natural Science Foundation of Shaanxi Province under Grants No.2014JM2-6091 and 2015JQ1007.

References

1. Kim, H. T., Kang, H. G., Ahn,C. J., & Cho, S. H. (2013). A Study on the Automated Compatibility Standard Test System for eBook DRM, The Journal of the Institute of Webcasting, Internet and Telecommunication, 13(2), 127-136.
2. Liang, K. Liu, J. K., Wong, D. S., & Susilo. W.(2014). An efficient cloud-based revocable identity-based proxy re-encryption scheme for public clouds data sharing, In Computer Security ESORICS , (pp.257-272). Springer International Publishing.
3. Wang, H., He, D., & Tang, S. (2016). Identity-Based Proxy-Oriented Data Uploading and Remote Data Integrity Checking in Public Cloud. IEEE Transactions on Information Forensics and Security, 11(6), 1165-1176.
4. Xu, P., Jiao, T., Wu, Q., Wang, W., & Jin, H. (2016). Conditional identity-based broadcast proxy re-encryption and its application to cloud email. IEEE Transactions on Computers, 65(1), 66-79.
5. Blaze, M., Bleumer, G., & Strauss, M. (1998, May). Divertible protocols and atomic proxy cryptography. In International Conference on the Theory and Applications of Cryptographic Techniques (pp. 127-144). Springer Berlin Heidelberg.
6. Chu, C. K., & Tzeng, W. G. (2007, October). Identity-based proxy re-encryption without random oracles. In International Conference on Information Security (pp. 189-202). Springer Berlin Heidelberg.
7. Qiu, J., Jo, J., & Lee, H. (2015). Collusion-Resistant Identity-Based Proxy Re-Encryption Without Random Oracles. International Journal of Security and Its Applications, 9(9), 337-344.
8. Boneh, D., & Franklin, M. (2001, August). Identity-based encryption from the Weil pairing. In Annual International Cryptology Conference (pp. 213-229). Springer Berlin Heidelberg.
9. Green, M., & Ateniese, G. (2007). Identity-based proxy re-encryption. In Applied Cryptography and Network Security (pp. 288-306). Springer Berlin Heidelberg.

Part VIII
The 3-rd International Workshop on Large Scale Networks and Applications (LSNA-2016)

A Provably Secure Two-Factor Authenticated Key Exchange Protocol for Wireless Sensor Networks Based on Authenticated Encryption

Fushan Wei[1], Ruijie Zhang[1], Jian Shen[2]
[1] State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou, China
[2] School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, China
emails: weifs831020@163.com,rjz_wonder@163.com, s_shenjian@126.com

Abstract. Two-factor authenticated key exchange (TFAKE) protocols are widely used in wireless sensor networks (WSNs) to provide user authentication and data confidentiality. However, many existing TFAKE protocols are found to be insecure against different attacks. In this paper, we investigate how to design provably secure TFAKE protocols using asymmetric cryptology mechanisms. Our main technique tool is robust authenticated encryption schemes and fuzzy verifiers. We first present a formal security model for TFAKE protocol in WSNs and then propose a novel TFAKE protocol based on authenticated encryption schemes. We prove the security of the proposed protocol in the random oracle model. The performance comparison result shows that our protocol not only enjoys provable security but also has high efficiency. protocols, our protocol is more efficient and enjoys provable security.

1 Introduction

Wireless sensor networks (WSNs), which are usually composed of millions of resource-constrained sensor nodes, are gaining more and more attention from the researchers all over the world. Numerous sensor nodes can be deployed in unattended environments to collect valuable data of interest. Due to its ubiquitous nature, WSNs are widely used in healthcare monitoring, wine production, natural disaster prevention and data logging. The primary concern in WSNs is the security of the collected data. In order to protect the security of the data, users should be authenticated by the gateway node before accessing the sensor nodes. Moreover, a session key should be established between the user and the sensor node. Besides the authentication requirement, the privacy of the user should also be ensured. Because of interest conflict, a user does not want its counterpart knows what kind of data he is interested in collecting in WSNs.

In 2009, Das[1] proposed the first two-factor user authentication scheme for WSNs using smart card and password. Unfortunately, Khan et al.[2] soon found that Das' protocol is insecure against the gateway node bypassing attack and the privileged-insider attack. Moreover, some desirable attributes, such as mutual authentication and free password updating, are missing. At the same time, He et al.[3] also pointed out

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_87

that Das's protocol is vulnerable to the insider attack and the impersonation attack. They also proposed an improved protocol to overcome the shortcomings. In 2012, Vaidya et al.[4] demonstrated that Das's protocol and its derivatives still have some security pitfalls and don't achieve session key establishment. They proposed a novel user authentication scheme with key agreement for WSN. In 2013, Sun et al.[5] showed that Khan et al.'s protocol[2] suffers from the gateway node impersonation attack, the gateway node bypassing attack, and the privileged-insider attack. They put forward an improved protocol and proves its security in the Bellare-Rogaway security model. He et al. [6]provided a robust anonymous two-factor authentication scheme. Their scheme can achieve anonymity, authentication, and perfect forward secrecy. Meanwhile, Yuan[7] also demonstrated that Khan et al.'s protocol[2] fails to achieve non-repudiation and suffer from the smart card lost attack. He also proposed an enhanced two factor authentication scheme. Unfortunately, his improved protocol is very inefficient because it heavily relies on computation-expensive public encryption schemes. In 2015, Jiang et al.[8] proposed an enhanced two-factor authentication scheme with unlinkability. Very recently, Wei et al.[9] found several loopholes of Yuan's protocol[7] and presented their improvement.

Until now, there is no satisfactory two-factor authentication scheme for wireless sensor networks. Existing protocol either are vulnerable to various attacks or fail to achieve some desirable attributes. Only few protocols have rigorous security proofs. What's worse, researcher pay much attention to find the shortcomings of existing protocols and make their improvements. Little attention has been paid to the principles and formal security models for two-factor authentication schemes for WSNs. To the best of our knowledge, only Wang et al.[10,11] take first step to find out the security requirements, design principles and formal security models of two-factor authentication schemes for WSNs. In this paper, we pursue the research line of Wang et al. to propose a provably secure two-factor authenticated key exchange protocol for WSNs. Our main technical tools are fuzzy verifiers[10] and robust authenticated encryption schemes[12]. Fuzzy verifiers can balance usability and security of two-factor authenticated key exchange protocols for WSNs. Robust authenticated encryption scheme enables us to formally prove the security of the proposed protocol. Our protocol not only enjoys provable security but also maintains high computation and communication efficiency. Consequently, we believe that our protocol is more suitable for applications in WSNs.

The rest of the paper is organized as follows. In the next section, we recall the security model and we present some building blocks in section 3. In section 4, we describes our proposed protocol and proves its security within the security model present in section 2. Finally, the final section concludes the paper.

2 Security Model

In this section, we first briefly review the security model presented in [11]. The participants in two-factor authenticated key exchange protocols are users, gateway nodes and sensor nodes. For simplicity, we assume there is only one gateway node in the system. The users register to the gateway node. The gateway will issue a smart card which stores some private information for authentication. Besides the smart card, a user also remembers a low-entropy password. The gateway holds a high-entropy private key which in unknown to other participants. There are many pre-deployed sensor nodes in the system. Usually, the sensor node share a common secret key with the gateway node which is derived from the master key of the gateway node. The communication model for two-factor authenticated key exchange protocol in WSNs is shown in Figure 1.

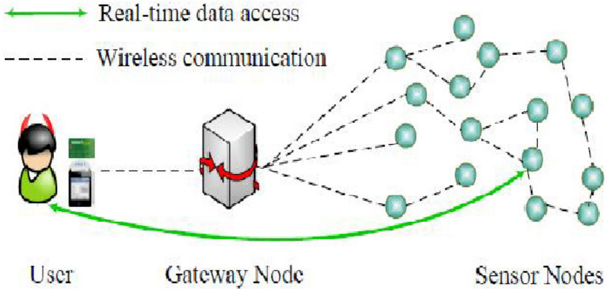


Figure 1. Communication model for TFAKE protocols in WSNs

The adversary controls all the communications among the participants. It can eavesdrops, modify, delay and cut the messages at will. The adversary can also compromise some sensor nodes, steal the smart card of the user and compromise the password of the user. However, the adversary cannot steal the smart card and compromise the password of a user at the same time, otherwise the user is fully corrupted. The abilities of the adversary are modeled using oracle queries. We use $\prod_i$ to denote the i-th session of participant $\prod_i$. The adversary is allowed to ask the following queries:

$Execute(U_i, G_j, SN_k)$: This query models the adversary's passive attack ability. Through the query, the adversary will get the transcript of the session among the user instance U_i, the gateway instance G_j and the sensor node instance SN_k.

$Send(\prod_i, m)$: This query models the adversary's active attack ability. Through the query, the adversary can impersonate a participant and send a forged message to the instance $\prod_i$. The adversary will get the message generated by the instance $\prod_i$ upon receiving the message m.

$\mathrm{Re}\,veal(\prod_i)$: This query models misuse of session key. The adversary can only send this query to a user instance or a sensor node instance. The victim instance will reveal the session key to the adversary.

$Corrupt(U,1)$: This query models the smart card lost attack. The adversary can get the smart card of the user U and extract the stored information using side channel attacks.

$Corrupt(U,2)$: This query models the adversary's ability to compromise the password of user U.

$Corrupt(SN)$: This query models the adversary's ability to compromise the sensor node SN . The adversary will control the compromised sensor node and get the secret key of the victim sensor node.

$Test(\prod_i)$: This query is used to measure the semantic security of the session key held by the instance $\prod_i$. It is not a real attack ability of the adversary. In order to answer the query, a coin is flipped; If the result is 1, then the simulator send back the real session to the adversary; otherwise, the simulator will send back a random key to the adversary. The adversary should distinguish the random session key from the session key. If the adversary succeeds, we say the semantic security of the session is violated.

In order to prevent the adversary from trivially winning the attack game, the adversary is only allowed to ask Test query to fresh instance. By the terminology fresh, we mean that the session key held by the instance $\prod_i$ is unknown to the adversary. The adversary can corrupt the participant $\prod$ or reveal the session key of instance $\prod_i$ to get the session key.

Let $\mathcal{P}$ be a TFAKE protocol and $\mathcal{A}$ be a probabilistic polynomial time adversary against the semantic security of $\mathcal{P}$. If the adversary wins the attack game by correctly guessing the coin flipping, we denote this event by $Succ$.

The advantage of an adversary $\mathcal{A}$ in breaking the semantic security of $\mathcal{P}$ is defined as $Adv_{\mathcal{P}}^{TFAKE}(\mathcal{A}) = 2\Pr[Succ]-1$. If for all probabilistic polynomial time adversary, the advantage is negligible, we say the protocol achieves semantic security.

3 Robust Authenticated Encryption

In this section, we briefly introduce the definition of robust authenticated encryption. For more details, refer to [12]. The robust authenticated encryption ensures that the adversary will not be able to generate a valid ciphertext without the knowledge of the secret key, nor can he get any information of the message from a ciphetext.

Given an alphabet $\sum$, an robust authenticated encryption scheme is a triple $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$. The key space $\mathcal{K}$ is a set of strings with an associated distribution; The encryption procedure $\mathcal{E}$ is deterministic and maps a tuple (K, N, A, λ, M) to a string $C = \mathcal{E}_K^{N,A}(M)$ of length $|M| + \lambda$, where K is the secret key, N is the nonce, A is the associated data, λ is the ciphertext expanding parameter and M is the message. We require that an robust authenticated encryption scheme can always encrypt any M using any K, N, A, λ . The decryption algorithm $\mathcal{D}$ is also deterministic and maps a tuple (K, N, A, λ, C) to a message $\mathcal{D}_K^{N,A,\lambda}(C) \in \sum^* \cup \{\perp\}$, which means the decryption algorithm either decrypts the correct plaintext or send a note of invalidation if the ciphertext is generated by the adversary.

The security of an robust authenticated encryption scheme can be defined using the games $Real$ and $Random$. The adversary has two oracles, namely an encryption oracle and a decryption oracle. For game $Real$, the queries are answered by the actual encryption and decryption algorithms. For game *Random* , the queries by the adversary are answered according to the family of random injections. The advantage of the adversary is defined to be the probability to distinguish these two games. The security of an robust authenticated encryption scheme demands that for all PPT adversary, the advantage is negligible.

4 The Proposed Protocol

In this section, we describe our scheme in detail. Our scheme has three phases: registration phase, authentication phase and password updating phase. In our system, the gateway node G has a long-term secret key K , and each sensor node SN_j shares a secret key $x_j = h(K, SN_j)$, where $h()$ is a hash function.

In the registration phase, a user U_i registers himself to the gateway node G . First, U_i chooses its identity ID_i , password and PW_i a random number b . Then U_i sends $PW_i^* = h(ID_i, PW_i)$ to the gateway node. Upon receiving the message, the gateway node first computes $A_i = h(h(ID_i) \oplus PW_i^*) \bmod n$, where n is a medium integer $2^4 \le n \le 2^8$. Then G compute $V_i = h(ID_i, K)$ and $N_i = V_i \oplus h(ID_i, PW_i^*)$. Finally, G generates a smart card with parameters $(N_i, A_i, n, h())$ and sends it to the user. Upon receiving the smart card, the user U_i writes b to the smart card.

When a user U_i wants to get real-time data from a sensor node SN_j , he execute the authentication phase. The participants execute the following steps:

1. the user U_i inserts the smart card to a card-reader and types his identity ID_i and password PW_i . The smart card first computes $h(h(ID_i) \oplus h(ID_i, PW_i)) \bmod n$ and checks whether this value is equal to A_i . If the verification is successful, then the smart card recovers $V_i = N_i \oplus h(ID_i, PW_i^*)$ and computes an encryption key $k_1 = h(V_i, T_1)$ for the robust authenticated encryption scheme, where T_1 is the timestamp. The smart card then randomly chooses a number R_1 and encrypts it $C_1 = \mathcal{E}_{k_1}^{T_1, ID_i}(R_1)$, where the timestamp T_1 is used as the nonce and the identity ID_i is used as the associated data. Finally, the smart card sends the message (ID_i, SN_j, T_1, C_1) to the gateway node.

2. Upon receiving the message (ID_i, SN_j, T_1, C_1) , the gateway node checks whether the timestamp T_1 is within the time-bound. If this is the case, the gateway node computes the encryption key $k_1 = h(h(ID_i, K)_i, T_1)$ and decrypts the ciphertext C_1 to get the random number R_1 . If the decryption is successful, then the user passes the verification of the gateway node. The gateway node computes an encryption key $k_2 = h(ID_i, G, SN_j, x_j, T_2)$ and computes $C_2 = \mathcal{E}_{k_2}^{T_2, ID}(R_1)$, where $ID = (ID_i, G, SN_j)$. Finally, the gateway node sends the message (ID, T_2, C_2) to the sensor node SN_j .

3. Upon receiving the message (ID, T_2, C_2), the sensor node SN_j checks whether the timestamp T_2 is within the time-bound. If this is the case, the sensor node SN_j computes the encryption key $k_2 = h(ID_i, G, SN_j, x_j, T_2)$ and decrypts the ciphertext C_2 to get the random number R_1. If the decryption is successful, SN_j also chooses a random number R_2 and computes an encryption key $k_3 = h(ID_i, R_1, T_3)$, where T_3 is the current timestamp in the system. SN_j calculates $C_3 = \mathcal{E}_{k_3}^{T_3, ID}(R_2)$ and finally sends the message (T_3, C_3) to the user. SN_j also computes the session key $sk = h(ID_i, T_3, R_1, R_2)$

4. Upon receiving the message (T_3, C_3), the user U_i checks whether the timestamp T_3 is within the time-bound. If this is the case, the user U_i computes the encryption key $k_3 = h(ID_i, R_1, T_3)$ and decrypt C_3 to get R_2. If the decryption is successful, then U_i accepts the session and computes the session key $sk = h(ID_i, T_3, R_1, R_2)$.

The password updating phase is similar with [9], we omit it for simplicity. The following theorem presents the security analysis to our proposed protocol. Due to lack of space, we omit the security proof. The detailed security proof will be presented in the full version.

Theorem 1. If the encryption scheme used in our protocol is an robust authenticated encryption scheme, and the hash function is an random oracle. For any PPT adversary, the advantage of the adversary in breaking the semantic security of our protocol is negligible.

5 Conclusions

In this paper, we proposed a provably secure two-factor authenticated key exchange protocol for WSNs. For the first time, we introduce the concept of robust authenticated encryption to provide rigorous security proof. We also use the fuzzy verifier to provide resilience to the smart card lost attack. Compared with existing protocols, our protocol has high efficiency while enjoy formal security proof. As a result, it is more suitable for applications in WSNs.

Acknowledgments. This work is supported by the National Natural Science Foundation of China (Nos. 61309016,61379150,61501515), Postdoctoral Science Foundation of China (Grant No. 2014M562493), Postdoctoral Science Foundation of Shanxi Province, and Key Scientific Technological Project of Henan Province (Grant Nos. 122102210126, 092101210502).

References

1. Das, M. L.: Two-Factor User Authentication in Wireless Sensor Networks. IEEE Transactions on Wireless Communications, vol. 8, no. 3, 2009, pp. 1086-1090.
2. Khan, M. K., Alghathbar, K.: Cryptanalysis And Security Improvements of ‘Two-Factor User Authentication in Wireless Sensor Networks’. Sensors, 10(3), 2010, pp. 2450-2459.
3. He, D., Gao, Y., Chan, S., Chen, C., & Bu, J.: An Enhanced Two-factor User Authentication Scheme in Wireless Sensor Networks. Ad Hoc & Sensor Wireless Networks, 10(4), 2010, pp. 361-371.
4. Vaidya, B., Makrakis, D., Mouftah, H.: (2012). Two-Factor Mutual Authentication with Key Agreement in Wireless Sensor Networks. Security and Communication Networks, 9(2), 2016, pp. 171-183.
5. Sun, D. Z., Li, J. X., Feng, Z. Y., Cao, Z. F., & Xu, G. Q: On the Security And Improvement of a Two-Factor User Authentication Scheme in Wireless Sensor Networks. Personal and Ubiquitous Computing, 17(5), 2013, pp. 895-905.
6. He, D., Kumar, N., Khan, M. K., & Lee, J. H.: Anonymous Two-Factor Authentication for Consumer Roaming Service in Global Mobility Networks. IEEE Transactions on Consumer Electronics, 59(4), 2013, pp. 811-817.
7. Yuan, J. J.: An Enhanced Two-Factor User Authentication in Wireless Sensor Networks. Telecommunication Systems, 55(1), 2014, pp. 105-113.
8. Jiang, Q., Ma, J., Lu, X., Tian, Y.: An Efficient Two-Factor User Authentication Scheme with Unlinkability for Wireless Sensor Networks. Peer-to-Peer Networking and Applications, 8(6), 2014, pp. 1070-1081.
9. Wei, F., Ma, J., Jiang, Q., Shen, J., & Ma, C.: Cryptanalysis and Improvement of an Enhanced Two-Factor User Authentication Scheme in Wireless Sensor Networks. Information Technology And Control, 45(1), 2016, pp. 62-70.
10. Wang, D., & Wang, P.: On the Anonymity of Two-Factor Authentication Schemes for Wireless Sensor Networks: Attacks, Principle and Solutions. Computer Networks, 73, 2014, pp. 41-57.
11. Wang, D., & Wang, P.: Understanding Security Failures of Two-Factor Authentication Schemes for Real-Time Applications in Hierarchical Wireless Sensor Networks. Ad Hoc Networks, 20, 2014, pp. 1-15.
12. Hoang V T, Krovetz T, Rogaway P: Robust Authenticated-Encryption AEZ and the Problem That it Solves. Advances in Cryptology-EUROCRYPT 2015. Springer Berlin Heidelberg, 2015, pp. 15-44.

Three elliptic curve cryptography-based RFID authentication protocols for Internet of Things

Rui An 1, Hui Feng 1 , Qin Liu 2, Li Li 3

1 School of Mathematics and Statistics, Wuhan University, Wuhan, China

ruia.whu@qq.com hfeng.math@whu.edu.cn

2 School of Computer, Wuhan University, Wuhan, China

csqliu@qq.com

3 International School of Software, Wuhan University, Wuhan, China

lli@whu.edu.cn

Abstract: With the development of information technology, the Internet of Thing (IoT) is extensively employed in many fields such as logistics, medical healthcare, food safety and intelligent transportation. The Radio Frequency Identification (RFID) technology is an important building block of the IoT. Therefore, how to address security problem in RFID system is a crucial issue for the security of the IoT. The RFID authentication protocol is a key cryptographic protocol ensuring communication security because it could provide authentication between the tag and the server. Recently, elliptic curve cryptography (ECC)-based RFID authentication protocols were studied widely because they could provide better security attributes compared with traditional RFID authentications. Lv et al. proposed three ECC-based RFID protocols and claimed their protocols could overcome weaknesses in previous protocols. Unfortunately, in this paper, we show that Lv et al.'s protocols cannot withstand the man-in-the-middle attack. To solve security problems in their protocols, we propose three improved ECC-based RFID authentication protocols.

Key words: Radio-frequency identification; Authentication protocol; Elliptic curve cryptography; Man-in-the-middle attack;

1. Introduction

The Internet of Things (IoT) is an emerging paradigm based on the modern wireless communication technology. Using embedded intelligence, the IoT could provide interconnections among different things including physical objects, cyber objects, and social objects [1]. With the development of many related technologies,

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_88

such as communication technology, electrical production technology and system integration technology, the IoT has been extensively used in many fields including logistic management, supply chain management, electronic commerce, electronic government and industrial manufacturing. According to a recent study [2], about 50 to 100 billion things will be connected to the Internet through the IoT by 2020. Due to wireless communication, the IoT is more vulnerable to different attacks compared with the traditional networks. Therefore, how to solve the security problem in the Iot become a very important issue in practical applications.

To expand the application of the IoT, many technologies and network devices such as the Radio Frequency Identification (RFID), wireless sensor networks and cloud computation have been used in the IoT. As an important building block of the IoT, the RFID technology attracted worldwide attentions from different fields. As an important automatic identification and data capture technology, the RFID technology is introduced during the Second World War. It could identify different objects such as goods and animal using radio waves. Compared with the traditional barcode technology, the RFID technology has many advantages: 1). Providing both read capability and write capability; 2). Providing the function of reading many tags synchronously; 3). Requiring no line-of-sight contact. Therefore, it could be applied in many environments and considered as the best replacement of the traditional barcode technology. According to a recently study [3], the market value of the RFID technology will gross over USD 25 billion in 2018.

RFID authentication protocol is an important security protocol for ensuring secure communication in RFID systems because it could provide authentication between the tag and the server. Due to the limited computing power and storage of the tag, it is difficult to design authentication protocols for RFID systems. Many RFID authentication protocols [4-13] using XOR operations or hash function operations or pseudo-random number generator have been proposed. According to Lee et al.'s study [14], the Elliptic Curve Cryptography (ECC) is also suitable for the design of RFID authentication protocol. Several ECC-based RFID authentication protocols [14-17] have been proposed to support mutual authentication between the tag and the server. The authentication process of those protocols is very complicated. In many applications such as logistic management and supply chain management, only the function that the server could authenticate the tag is needed. Compared with ECC-based RFID authentication protocols supporting mutual authentication, ECC-based RFID authentication protocols supporting single authentication have better performance.

Lee et al. [18] proposed an Elliptic Curve Discrete Logarithm (ECDL) problem based randomized access control (EC-RAC) protocols for single authentication in RFID systems. They demonstrated that their protocols were provably secure in the generic group model. Unfortunately, Bringer et al. [19] and Deursen et al. [20] pointed out that Lee et al.'s EC-RAC authentication protocols cannot withstand tracking attacks and replay attacks. To solve those security problems, Lee et al. [21] proposed three improved EC-RAC protocols. However, Deursen and Radomirovic [22] pointed out that Lee et al. improved EC-RAC protocols were still vulnerable to the tracking attacks. Lv et al. [23] also pointed out Lee et al.' protocols [21] were vulnerable to tracking attacks. To withstand tracking attacks, Lv et al. proposed three improved EC-RAC protocols. In this paper, we analyze the security of Lv et al.'s EC-RAC protocols. We demonstrate that their protocols cannot withstand the man-in-the-middle attacks. Afterwards, we proposed three improved EC-RAC protocols by modifying Lv et al.'s protocols slightly.

The organization of the paper is sketched as follows. Section 2 reviews Lv et al.'s EC-RAC protocols briefly. Section 3 analyzes the security of Lv et al.'s EC-RAC protocols. Section 4 proposes the improved EC-RAC protocols to solve problems in Lv et al.'s protocols. Security analysis and performance analysis are proposed in Section 5 and Section 6 respectively. At last, Section 7 gives some conclusions of the paper.

2. Review of Lv et al.'s protocols

To enhance security, Lv et al. proposed three ECC-based RFID authentication protocols, i.e. Lv et al.'s EC-RAC 1 protocol, Lv et al.'s EC-RAC 2 protocol and Lv et al.'s EC-RAC 3 protocol. For convenience, some notations used in the paper are defined as follows.

- $F(q)$: a finite filed;
- n: a large prime number;
- $E(F(q))$: an elliptic curve defined in $F(q)$;

- P: a point on $E(F(p))$ with order ;
- G: the group generated by the point P;
- (y, Y): the p n rivate/public key pair of the server, where $Y = yP$;
- (x_i, X_i): the secret information of the tag, where $X_i = x_i P, i = 1, 2$;

2.1. Lv et al.'s EC-RAC 1 protocol

This protocol is a kind of secure identity transfer scheme. In the protocol, the server could authenticate the tag by checking whether the received identity verifier is stored in its database. At the beginning, the server chooses system parameters $params = \{F(q), E(F(q)), n, P, Y\}$. It also stores (X_1) and (x_1, Y) in its database and the tag's memory separately. As shown in Fig. 1, the following steps will be executed between the tag and the server.

1). The tag generates a random number r_{t1}, computes $T_1 = r_{t1}P$ and sends the message $\{T_1\}$ the server.

2). Upon receiving the message $\{T_1\}$, the server generates a random number r_{s1}, and sends the message $\{r_{s1}\}$ to the tag.

3). Upon receiving the message $\{r_{s1}\}$, the tag computes $T_2 = (r_{t1} + r_{s1}x_1 r_{t1})Y$ and sends the message $\{T_2\}$ to the server.

4). Upon receiving the message $\{T_2\}$, the server computes $U = r_{s1}^{-1}(y^{-1}T_2 - T_1)$. The server checks whether U and x_1T_1 are equal. If they are not equal, the server rejects the session; otherwise, the tag is authenticated.

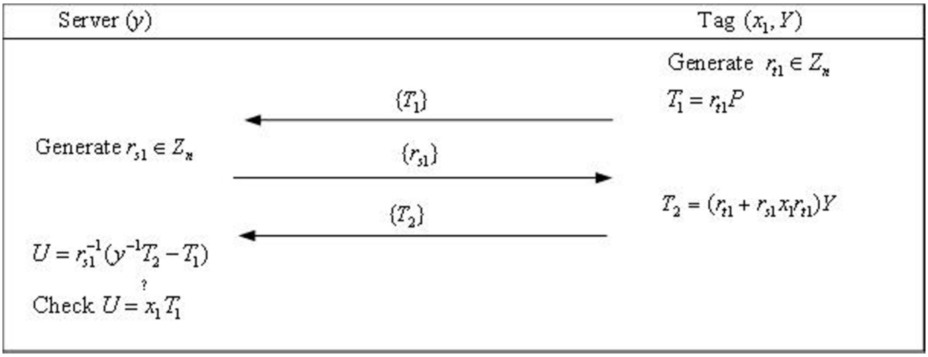

Server (y)
Tag (x_1, Y)
Generate $r_{t1} \in Z_n$
$T_1 = r_{t1}P$
$\{T_1\}$
Generate $r_{s1} \in Z_n$
$\{r_{s1}\}$
$\{T_2\}$
$T_2 = (r_{t1} + r_{s1}x_1r_{t1})Y$
$U = r_{s1}^{-1}(y^{-1}T_2 - T_1)$
Check $U \overset{?}{=} x_1 T_1$

Fig. 1. Lv et al.'s Modified EC-RAC 1 protocol

2.2. Lv et al.'s EC-RAC 2 protocol

This protocol is a kind of secure identity transfer scheme and secure password transfer scheme. In the protocol, the server could authenticate the tag by checking whether the received identity verifier is stored in its database and the corresponding password is correct. At the beginning, the server chooses system parameters $params = \{F(q), E(F(q)), n, P, Y\}$. It also stores (x_1, X_1, x_2, X_2) and (x_1, x_2, Y) in its database and the tag's memory separately. As shown in Fig. 2, the following steps will be executed between the tag and the server.

1). The tag generates a random number r_{t1}, computes $T_1 = r_{t1}P$ and sends the message $\{T_1\}$ the server.

2). Upon receiving the message $\{T_1\}$, the server generates a random number r_{s1}, and sends the message $\{r_{s1}\}$ to the tag.

3). Upon receiving the message $\{r_{s1}\}$, the tag computes $T_2 = (r_{t1} + r_{s1}x_1r_{t1})Y$, $T_3 = (r_{t1}x_1 + r_{s1}x_2r_{t1})Y$ and sends the message $\{T_2, T_3\}$ to the server.

4). Upon receiving the message $\{T_2, T_3\}$, the server computes $W = r_{s1}^{-1}(y^{-1}T_2 - T_1)$ and $V = r_{s1}^{-1}(y^{-1}T_3 - x_1T_1)$. The server checks whether both equations $W = x_1T_1$ and $V = x_2T_1$ hold. If either of them does not hold, the server stops the session; otherwise, the tag is authenticated.

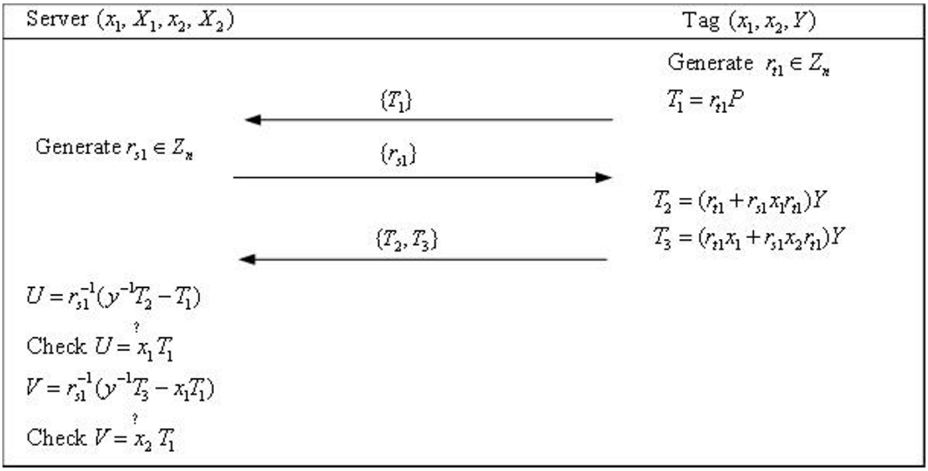

Server (x_1, X_1, x_2, X_2)
Tag (x_1, x_2, Y)
Generate $r_{t1} \in Z_n$
$T_1 = r_{t1}P$
$\{T_1\}$
Generate $r_{s1} \in Z_n$
$\{r_{s1}\}$
$T_2 = (r_{t1} + r_{s1}x_1r_{t1})Y$
$T_3 = (r_{t1}x_1 + r_{s1}x_2r_{t1})Y$
$\{T_2, T_3\}$
$U = r_{s1}^{-1}(y^{-1}T_2 - T_1)$
Check $U \overset{?}{=} x_1 T_1$
$V = r_{s1}^{-1}(y^{-1}T_3 - x_1T_1)$
Check $V \overset{?}{=} x_2 T_1$

Fig. 2. Lv et al.'s Modified EC-RAC 2 protocol

2.3. Lv et al.'s EC-RAC 3 protocol

This protocol is a kind of secure identity transfer scheme and secure password transfer scheme. In the protocol, the server could authenticate the tag by checking whether the received identity verifier is stored in its database and the corresponding password is correct. At the beginning, the server chooses system parameters $params = \{F(q), E(F(q)), n, P, Y\}$. It also stores (x_1, X_1, x_2, X_2) and (x_1, x_2, Y) in its database and the tag's memory separately. As shown in Fig. 2, the following steps will be executed between the tag and the server.

1). The tag generates two random numbers r_{t1}, r_{t2}, computes $T_1 = r_{t1}P$, $T_2 = r_{t2}P$ and sends the message $\{T_1, T_2\}$ the server.

2). Upon receiving the message $\{T_1, T_2\}$, the server generates a random number r_{s1}, and sends the message $\{r_{s1}\}$ to the tag.

3). Upon receiving the message $\{r_{s1}\}$, the tag computes $T_3 = (r_{t1} + r_{s1}x_1r_{t1})Y$, $T_4 = (r_{t2}x_1 + r_{s1}x_2r_{t2})Y$ and sends the message $\{T_3, T_4\}$ to the server.

4). Upon receiving the message $\{T_3, T_4\}$, the server computes $U = r_{s1}^{-1}(y^{-1}T_3 - T_1)$ and $V = r_{s1}^{-1}(y^{-1}T_4 - x_1T_2)$. The server checks whether both equations $U = x_1T_1$ and $V = x_2T_2$ hold. If either of them does not hold, the server stops the session; otherwise, the tag is authenticated.

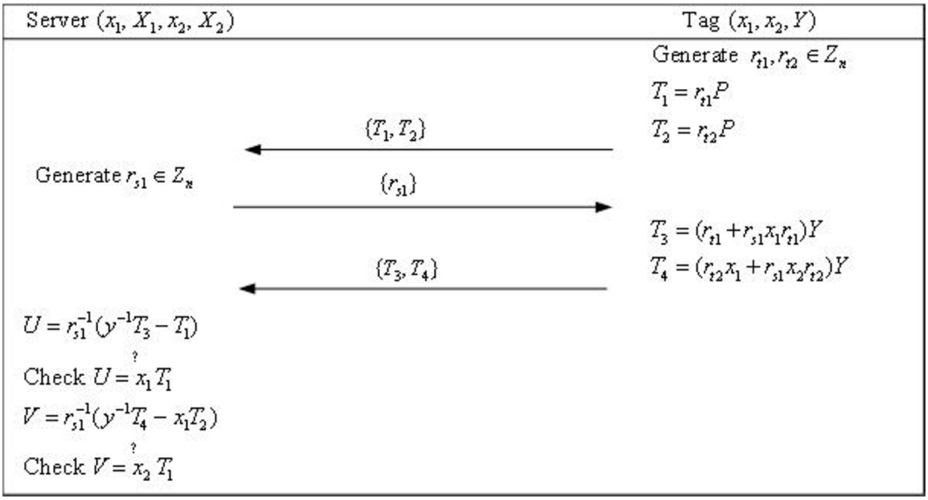

Server (x_1, X_1, x_2, X_2)
Tag (x_1, x_2, Y)
Generate $r_{t1}, r_{t2} \in Z_n$
$T_1 = r_{t1}P$
$T_2 = r_{t2}P$
$\{T_1, T_2\}$
Generate $r_{s1} \in Z_n$
$\{r_{s1}\}$
$T_3 = (r_{t1} + r_{s1}x_1r_{t1})Y$
$T_4 = (r_{t2}x_1 + r_{s1}x_2r_{t2})Y$
$\{T_3, T_4\}$
$U = r_{s1}^{-1}(y^{-1}T_3 - T_1)$
Check $U \overset{?}{=} x_1 T_1$
$V = r_{s1}^{-1}(y^{-1}T_4 - x_1T_2)$
Check $V \overset{?}{=} x_2 T_1$

Fig. 3. Lv et al.’s Modified EC-RAC 3 protocol

3. Security analysis of Lv et al.’s protocols

With the development of the cryptographic theory, several security model for RFID authentication protocols have been proposed . According to Vaudenay’s work, attackers against the RFID authentication protocols could be divided into wide (or narrow) attackers and strong (or weak) attackers. A wide (narrow) attacker is the one who could (not) get the verification result of the server. A strong (weak) attacker is the one who could (not) extract a tag’s secret and reuse it. It is easy to say the wide–strong attacker is the most powerful. We call a RFID authentication protocol is wide-strong privacy-preserving if it is untraceable against the wide–strong attacker.

Lv et al. claimed that all their three protocols are wide-strong privacy-preserving against the wide–strong attacker. Unfortunately, we will show their protocols are not secure against the wide–strong attacker through proposing three concrete attacks.

3.1. Security analysis of Lv et al.’s EC-RAC 1 protocol

In this subsection, we analyze the security of Lv et al.’s EC-RAC 1 protocol. As show in Fig. 4, the man-in-the middle attack is described as follows.

1). The tag generates a random number r_{t1}, computes $T_1 = r_{t1}P$ and sends the message $\{T_1\}$ the server.

2). Upon intercepting the message $\{T_1\}$, the adversary generates a random number r_a, computes $T_1' = r_a T_1$ and sends message $\{T_1'\}$ to the server.

3). Upon receiving the message $\{T_1'\}$, the server generates a random number r_{s1}, and sends the message $\{r_{s1}\}$ to the adversary.

4). Upon receiving the message $\{r_{s1}\}$, the adversary sends it to the tag directly.

5). Upon receiving the message $\{r_{s1}\}$, the tag computes $T_2 = (r_{t1} + r_{s1}x_1 r_{t1})Y$ and sends the message $\{T_2\}$ to the server.

6). Upon intercepting the message $\{T_2\}$, the adversary generates a random number r_a, computes $T_2' = r_a T_2$ and sends message $\{T_2'\}$ to the server.

7). Upon receiving the message $\{T_2'\}$, the server computes $U = r_{s1}^{-1}(y^{-1}T_2' - T_1')$. The server checks whether U and $x_1 T_1'$ are equal. If they are not equal, the server rejects the session; otherwise, the tag is authenticated.

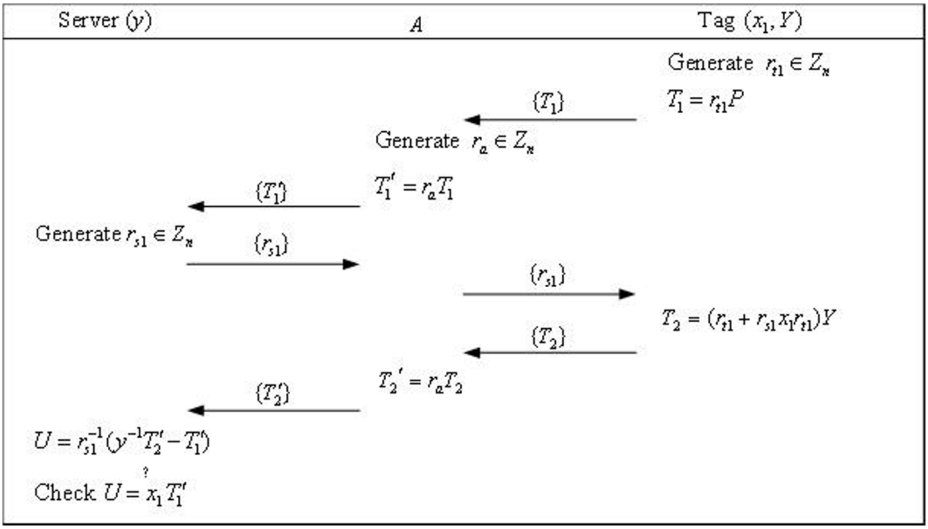


Fig. 4. Attack against Lv et al.'s Modified EC-RAC 1 protocol

Since $T_1 = r_{t1}P$, $T_1' = r_aT_1$, $T_2 = (r_{t1} + r_{s1}x_1r_{t1})Y$ and $T_2' = r_aT_2$, then we could get that

$$\begin{aligned} U &= r_{s1}^{-1}(y^{-1}T_2' - T_1') = r_{s1}^{-1}(y^{-1}r_aT_2 - r_aT_1) \\ &= r_{s1}^{-1}(y^{-1}r_a(r_{t1} + r_{s1}x_1r_{t1})Y - r_ar_{t1}P) \\ &= r_{s1}^{-1}(y^{-1}r_a(r_{t1} + r_{s1}x_1r_{t1})yP - r_ar_{t1}P) \\ &= r_{s1}^{-1}(r_a(r_{t1} + r_{s1}x_1r_{t1})P - r_ar_{t1}P) \\ &= r_{s1}^{-1}(r_ar_{t1}P + r_ar_{s1}x_1r_{t1}P - r_ar_{t1}P) \\ &= r_{s1}^{-1}r_ar_{s1}x_1r_{t1}P = x_1r_ar_{t1}P = x_1T_1' \end{aligned} \tag{1}$$

Thus, the message $\{T_1'\}$ and $\{T_2'\}$ could pass the verification of the server. Therefore, we can conclude that Lv et al.'s EC-RAC 1 protocol cannot withstand the man-in-the-middle attack.

3.2. Security analysis of Lv et al.'s EC-RAC 2 protocol

In this subsection, we analyze the security of Lv et al.'s EC-RAC 2 protocol. As show in Fig. 5, the man-in-the middle attack is described as follows.

1). The tag generates a random number r_{t1}, computes $T_1 = r_{t1}P$ and sends the message $\{T_1\}$ the server.

2). Upon intercepting the message $\{T_1\}$, the adversary generates a random number r_a, computes $T_1' = r_aT_1$ and sends message $\{T_1'\}$ to the server.

3). Upon receiving the message $\{T_1'\}$, the server generates a random number r_{s1}, and sends the message $\{r_{s1}\}$ to the adversary.

4). Upon receiving the message $\{r_{s1}\}$, the adversary sends it to the server directly.

5). Upon receiving the message $\{r_{s1}\}$, the tag computes $T_2 = (r_{t1} + r_{s1}x_1r_{t1})Y$, $T_3 = (r_{t1}x_1 + r_{s1}x_2r_{t1})Y$ and sends the message $\{T_2, T_3\}$ to the server.

6). Upon intercepting the message $\{T_2, T_3\}$, the adversary generates a random number r_a, computes $T_2' = r_a T_2$, $T_3' = r_a T_3$ and sends message $\{T_2', T_3'\}$ to the server.

7). Upon receiving the message $\{T_2', T_3'\}$, the server computes $W = r_{s1}^{-1}(y^{-1}T_2' - T_1')$ and $V = r_{s1}^{-1}(y^{-1}T_3' - x_1 T_1')$. The server checks whether both equations $W = x_1 T_1'$ and $V = x_2 T_1'$ hold. If either of them does not hold, the server stops the session; otherwise, the tag is authenticated.

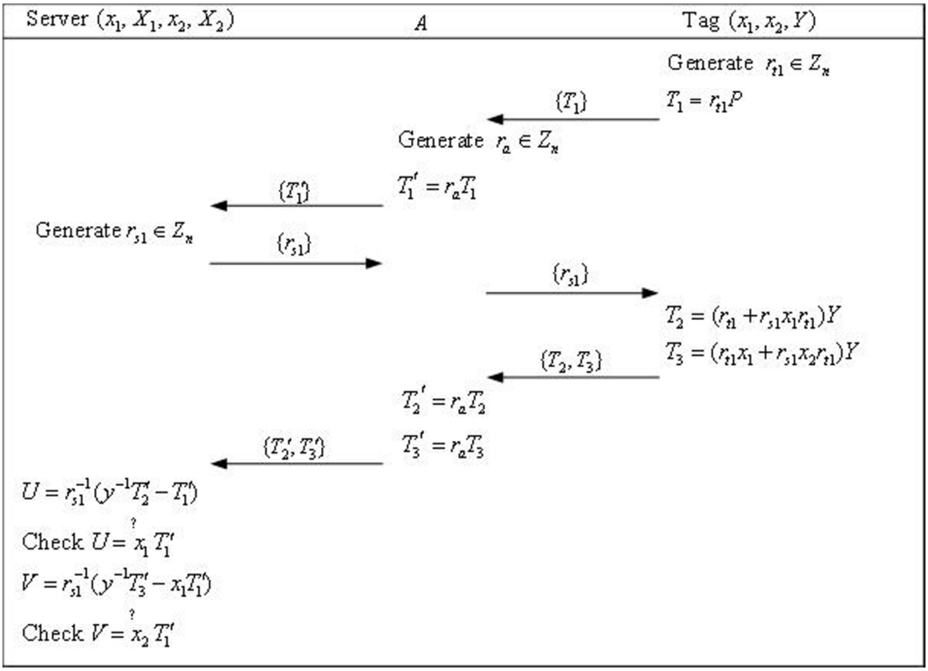

Server (x_1, X_1, x_2, X_2)
A
Tag (x_1, x_2, Y)
Generate $r_{t1} \in Z_n$
$T_1 = r_{t1}P$
$\{T_1\}$
Generate $r_a \in Z_n$
$T_1' = r_a T_1$
$\{T_1'\}$
Generate $r_{s1} \in Z_n$
$\{r_{s1}\}$
$\{r_{s1}\}$
$T_2 = (r_{t1} + r_{s1}x_1 r_{t1})Y$
$T_3 = (r_{t1}x_1 + r_{s1}x_2 r_{t1})Y$
$\{T_2, T_3\}$
$T_2' = r_a T_2$
$T_3' = r_a T_3$
$\{T_2', T_3'\}$
$U = r_{s1}^{-1}(y^{-1}T_2' - T_1')$
Check $U \overset{?}{=} x_1 T_1'$
$V = r_{s1}^{-1}(y^{-1}T_3' - x_1 T_1')$
Check $V \overset{?}{=} x_2 T_1'$

Fig. 5. Attack against Lv et al.'s Modified EC-RAC 2 protocol

Since $T_1 = r_{t1}P$, $T_1' = r_a T_1$, $T_2 = (r_{t1} + r_{s1}x_1 r_{t1})Y$, $T_3 = (r_{t1}x_1 + r_{s1}x_2 r_{t1})Y$, $T_2' = r_a T_2$ and $T_3' = r_a T_3$, then we could get that

$$
\begin{aligned}
U &= r_{s1}^{-1}(y^{-1}T_2' - T_1') = r_{s1}^{-1}(y^{-1}r_aT_2 - r_aT_1) \\
&= r_{s1}^{-1}(y^{-1}r_a(r_{t1} + r_{s1}x_1r_{t1})Y - r_ar_{t1}P) \\
&= r_{s1}^{-1}(y^{-1}r_a(r_{t1} + r_{s1}x_1r_{t1})yP - r_ar_{t1}P) \\
&= r_{s1}^{-1}(r_a(r_{t1} + r_{s1}x_1r_{t1})P - r_ar_{t1}P) \\
&= r_{s1}^{-1}(r_ar_{t1}P + r_ar_{s1}x_1r_{t1}P - r_ar_{t1}P) \\
&= r_{s1}^{-1}r_ar_{s1}x_1r_{t1}P = x_1r_ar_{t1}P = x_1T_1'
\end{aligned}
\tag{2}
$$

and

$$
\begin{aligned}
V &= r_{s1}^{-1}(y^{-1}T_3' - x_1T_1') = r_{s1}^{-1}(y^{-1}r_aT_3 - x_1r_aT_1) \\
&= r_{s1}^{-1}(y^{-1}r_a(r_{t1}x_1 + r_{s1}x_2r_{t1})Y - x_1r_ar_{t1}P) \\
&= r_{s1}^{-1}(y^{-1}r_a(r_{t1}x_1 + r_{s1}x_2r_{t1})yP - x_1r_ar_{t1}P) \\
&= r_{s1}^{-1}(r_a(r_{t1}x_1 + r_{s1}x_2r_{t1})P - x_1r_ar_{t1}P) \\
&= r_{s1}^{-1}(r_ar_{t1}x_1P + r_{s1}x_2r_{t1}P - x_1r_ar_{t1}P) \\
&= r_{s1}^{-1}r_{s1}x_2r_{t1}P = x_2r_{t1}P = x_2T_1'
\end{aligned}
\tag{3}
$$

Thus, the message $\{T_1'\}$ and $\{T_2', T_3'\}$ could pass the verification of the server. Therefore, we can conclude that Lv et al.'s EC-RAC 2 protocol cannot withstand the man-in-the-middle attack.

3.3. Security analysis of Lv et al.'s EC-RAC 3 protocol

In this subsection, we analyze the security of Lv et al.'s EC-RAC 2 protocol. As show in Fig. 5, the man-in-the middle attack is described as follows.

1). The tag generates two random numbers r_{t1}, r_{t2}, computes $T_1 = r_{t1}P$, $T_2 = r_{t2}P$ and sends the message $\{T_1, T_2\}$ the server.

2). Upon intercepting the message $\{T_1, T_2\}$, the adversary generates a random number r_a, computes $T_1' = r_aT_1$, $T_2' = r_aT_2$ and sends message $\{T_1', T_2'\}$ to the server.

3). Upon receiving the message $\{T_1', T_2'\}$, the server generates a random number r_{s1}, and sends the message $\{r_{s1}\}$ to the tag.

4). Upon receiving the message $\{r_{s1}\}$, the adversary sends it to the server directly.

5). Upon receiving the message $\{r_{s1}\}$, the tag computes $T_3 = (r_{t1} + r_{s1}x_1r_{t1})Y$, $T_4 = (r_{t2}x_1 + r_{s1}x_2r_{t2})Y$ and sends the message $\{T_3, T_4\}$ to the server.

6). Upon intercepting the message $\{T_3, T_4\}$, the adversary generates a random number r_a, computes $T_3' = r_aT_3$, $T_4' = r_aT_4$ and sends message $\{T_3', T_4'\}$ to the server.

7). Upon receiving the message $\{T_3', T_4'\}$, the server computes $U = r_{s1}^{-1}(y^{-1}T_3' - T_1')$ and $V = r_{s1}^{-1}(y^{-1}T_4' - x_1T_2')$. The server checks whether both equations $U = x_1T_1'$ and $V = x_2T_2'$ hold. If either of them does not hold, the server stops the session; otherwise, the tag is authenticated.

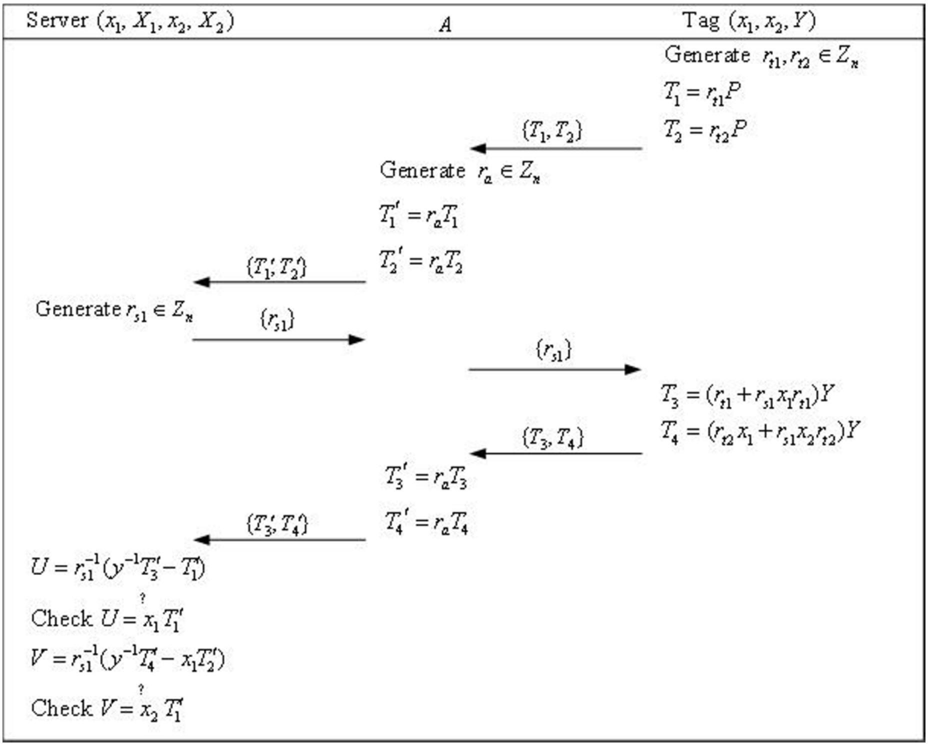

Server (x_1, X_1, x_2, X_2)
A
Tag (x_1, x_2, Y)
Generate $r_{t1}, r_{t2} \in Z_n$
$T_1 = r_{t1}P$
$T_2 = r_{t2}P$
$\{T_1, T_2\}$
Generate $r_a \in Z_n$
$T_1' = r_aT_1$
$T_2' = r_aT_2$
$\{T_1', T_2'\}$
Generate $r_{s1} \in Z_n$
$\{r_{s1}\}$
$\{r_{s1}\}$
$T_3 = (r_{t1} + r_{s1}x_1r_{t1})Y$
$T_4 = (r_{t2}x_1 + r_{s1}x_2r_{t2})Y$
$\{T_3, T_4\}$
$T_3' = r_aT_3$
$T_4' = r_aT_4$
$\{T_3', T_4'\}$
$U = r_{s1}^{-1}(y^{-1}T_3' - T_1')$
Check $U \overset{?}{=} x_1T_1'$
$V = r_{s1}^{-1}(y^{-1}T_4' - x_1T_2')$
Check $V \overset{?}{=} x_2T_1'$

Fig. 6. Attack against Lv et al.'s Modified EC-RAC 3 protocol

Since $T_1 = r_{t1}P$, $T_2 = r_{t2}P$, $T_1' = r_aT_1$, $T_2' = r_aT_2$, $T_3 = (r_{t1} + r_{s1}x_1r_{t1})Y$, $T_4 = (r_{t2}x_1 + r_{s1}x_2r_{t2})Y$, $T_3' = r_aT_3$ and $T_4' = r_aT_4$, then we could get that

$$\begin{aligned} U &= r_{s1}^{-1}(y^{-1}T_3' - T_1') = r_{s1}^{-1}(y^{-1}r_aT_3 - r_aT_1) \\ &= r_{s1}^{-1}(y^{-1}r_a(r_{t1} + r_{s1}x_1r_{t1})Y - r_ar_{t1}P) \\ &= r_{s1}^{-1}(y^{-1}r_a(r_{t1} + r_{s1}x_1r_{t1})yP - r_ar_{t1}P) \\ &= r_{s1}^{-1}(r_a(r_{t1} + r_{s1}x_1r_{t1})P - r_ar_{t1}P) \\ &= r_{s1}^{-1}(r_ar_{t1}P + r_ar_{s1}x_1r_{t1}P - r_ar_{t1}P) \\ &= r_{s1}^{-1}r_ar_{s1}x_1r_{t1}P = x_1r_ar_{t1}P = x_1T_1' \end{aligned} \tag{2}$$

and

$$\begin{aligned} V &= r_{s1}^{-1}(y^{-1}T_4' - x_1T_2') = r_{s1}^{-1}(y^{-1}r_aT_4 - x_1r_aT_2) \\ &= r_{s1}^{-1}(y^{-1}r_a(r_{t2}x_1 + r_{s1}x_2r_{t2})Y - x_1r_ar_{t2}P) \\ &= r_{s1}^{-1}(y^{-1}r_a(r_{t2}x_1 + r_{s1}x_2r_{t2})yP - x_1r_ar_{t2}P) \\ &= r_{s1}^{-1}(r_a(r_{t2}x_1 + r_{s1}x_2r_{t2})P - x_1r_ar_{t2}P) \\ &= r_{s1}^{-1}(r_ar_{t2}x_1P + r_{s1}x_2r_{t2}P - x_1r_ar_{t2}P) \\ &= r_{s1}^{-1}r_{s1}x_2r_{t2}P = x_2r_{t2}P = x_2T_2' \end{aligned} \tag{3}$$

Thus, the message $\{T_1', T_2'\}$ and $\{T_3', T_4'\}$ could pass the verification of the server. Therefore, we can conclude that Lv et al.'s EC-RAC 3 protocol cannot withstand the man-in-the-middle attack.

4. The proposed protocols

From the description of Lv et al.'s protocols, we know that there is linear relation between two messages sent by the tag. The linear relation could be used by the adversary to carry out the man-in-the-middle attacks. Subsequently, breaking the linear relation is the simplest way to withstand those attacks. Based on such thought, our protocols are described as follows.

4.1. Our EC-RAC 1 protocol

This protocol is a kind of secure identity transfer scheme. In the protocol, the server could authenticate the tag by checking whether the received identity verifier is stored in its database. At the beginning, the server chooses system parameters $params = \{F(q), E(F(q)), n, P, Y\}$. It also stores (X_1) and (x_1, Y) in its database and the tag's memory separately. As shown in Fig. 7, the following steps will be executed between the tag and the server.

1). The tag generates a random number r_{t1}, computes $T_1 = r_{t1}P$ and sends the message $\{T_1\}$ the server.

2). Upon receiving the message $\{T_1\}$, the server generates a random number r_{s1}, and sends the message $\{r_{s1}\}$ to the tag.

3). Upon receiving the message $\{r_{s1}\}$, the tag computes $T_2 = (r_{t1}x(T_1) + r_{s1}x_1r_{t1})Y$ and sends the message $\{T_2\}$ to the server, where $x(T_1)$ denotes the x-coordinate of the elliptic curve point T_1.

4). Upon receiving the message $\{T_2\}$, the server computes $U = r_{s1}^{-1}(y^{-1}T_2 - x(T_1)T_1)$. The server checks whether U and x_1T_1 are equal. If they are not equal, the server rejects the session; otherwise, the tag is authenticated.

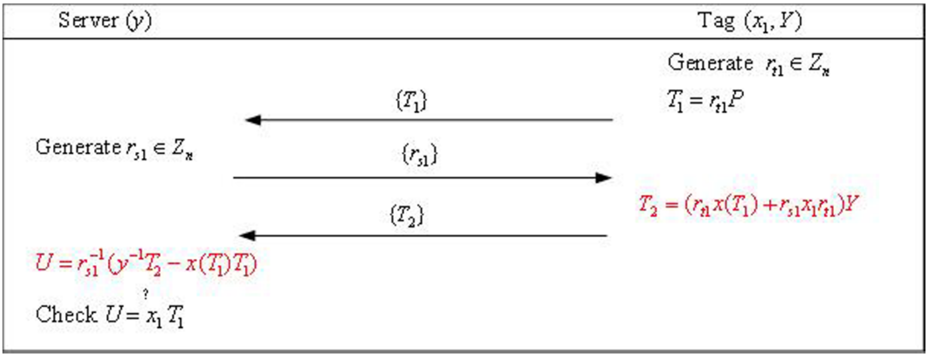

Server (y)
Tag (x_1, Y)
Generate $r_{t1} \in Z_n$
$T_1 = r_{t1}P$
$\{T_1\}$
Generate $r_{s1} \in Z_n$
$\{r_{s1}\}$
$T_2 = (r_{t1}x(T_1) + r_{s1}x_1r_{t1})Y$
$\{T_2\}$
$U = r_{s1}^{-1}(y^{-1}T_2 - x(T_1)T_1)$
Check $U \overset{?}{=} x_1 T_1$

Fig. 7. Lv et al.'s Modified EC-RAC 1 protocol

4.2. Our EC-RAC 2 protocol

This protocol is a kind of secure identity transfer scheme and secure password transfer scheme. In the protocol, the server could authenticate the tag by checking whether the received identity verifier is stored in its database and the corresponding password is correct. At the beginning, the server chooses system parameters $params = \{F(q), E(F(q)), n, P, Y\}$. It also stores (x_1, X_1, x_2, X_2) and (x_1, x_2, Y) in its database and the tag's memory separately. As shown in Fig. 8, the following steps will be executed between the tag and the server.

1). The tag generates a random number r_{t1}, computes $T_1 = r_{t1}P$ and sends the message $\{T_1\}$ the server.

2). Upon receiving the message $\{T_1\}$, the server generates a random number r_{s1}, and sends the message $\{r_{s1}\}$ to the tag.

3). Upon receiving the message $\{r_{s1}\}$, the tag computes $T_2 = (r_{t1}x(T_1) + r_{s1}x_1r_{t1})Y$, $T_3 = (r_{t1}x_1x(T_1) + r_{s1}x_2r_{t1})Y$ and sends the message $\{T_2, T_3\}$ to the server, where $x(T_1)$ denotes the x-coordinate of the elliptic curve point T_1.

4). Upon receiving the message $\{T_2, T_3\}$, the server computes $W = r_{s1}^{-1}(y^{-1}T_2 - x(T_1)T_1)$ and $V = r_{s1}^{-1}(y^{-1}T_3 - x_1x(T_1)T_1)$. The server checks whether both equations $W = x_1T_1$ and $V = x_2T_1$ hold. If either of them does not hold, the server stops the session; otherwise, the tag is authenticated.

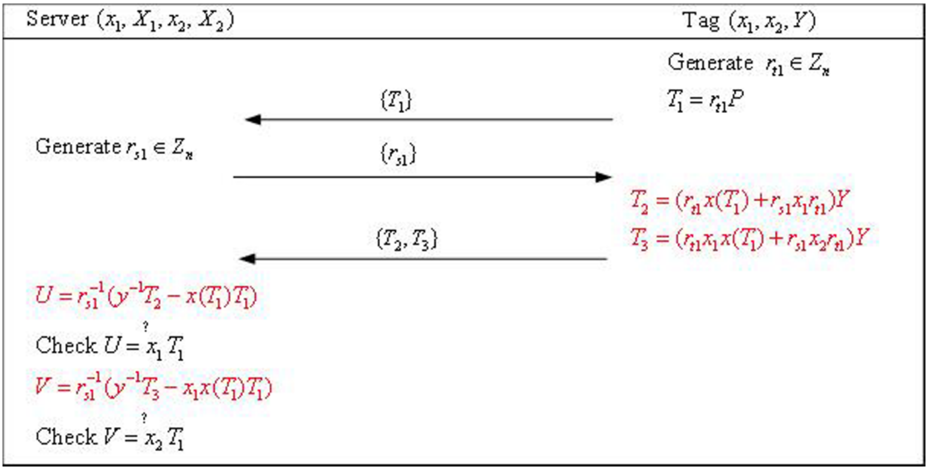

Server (x_1, X_1, x_2, X_2)
Tag (x_1, x_2, Y)
Generate $r_{t1} \in Z_n$
$T_1 = r_{t1}P$
$\{T_1\}$
Generate $r_{s1} \in Z_n$
$\{r_{s1}\}$
$T_2 = (r_{t1}x(T_1) + r_{s1}x_1r_{t1})Y$
$T_3 = (r_{t1}x_1x(T_1) + r_{s1}x_2r_{t1})Y$
$\{T_2, T_3\}$
$U = r_{s1}^{-1}(y^{-1}T_2 - x(T_1)T_1)$
Check $U \overset{?}{=} x_1 T_1$
$V = r_{s1}^{-1}(y^{-1}T_3 - x_1x(T_1)T_1)$
Check $V \overset{?}{=} x_2 T_1$

Fig. 8. Lv et al.'s Modified EC-RAC 2 protocol

4.3. Our EC-RAC 3 protocol

This protocol is a kind of secure identity transfer scheme and secure password transfer scheme. In the protocol, the server could authenticate the tag by checking whether the received identity verifier is stored in its database and the corresponding password is correct. At the beginning, the server chooses system parameters $params = \{F(q), E(F(q)), n, P, Y\}$. It also stores (x_1, X_1, x_2, X_2) and (x_1, x_2, Y) in its database and the tag's memory separately. As shown in Fig. 9, the following steps will be executed between the tag and the server.

1). The tag generates two random numbers r_{t1}, r_{t2}, computes $T_1 = r_{t1}P$, $T_2 = r_{t2}P$ and sends the message $\{T_1, T_2\}$ the server.

2). Upon receiving the message $\{T_1, T_2\}$, the server generates a random number r_{s1}, and sends the message $\{r_{s1}\}$ to the tag.

3). Upon receiving the message $\{r_{s1}\}$, the tag computes $T_3 = (r_{t1}x(T_1) + r_{s1}x_1r_{t1})Y$, $T_4 = (r_{t2}x_1x(T_2) + r_{s1}x_2r_{t2})Y$ and sends the message $\{T_3, T_4\}$ to the server, where $x(T_1)$ and $x(T_2)$ denote the x-coordinate of the elliptic curve points T_1 and T_2 respectively.

4). Upon receiving the message $\{T_3, T_4\}$, the server computes $U = r_{s1}^{-1}(y^{-1}T_3 - x(T_1)T_1)$ and $V = r_{s1}^{-1}(y^{-1}T_4 - x_1x(T_2)T_2)$. The server checks whether both equations $U = x_1T_1$ and $V = x_2T_2$ hold. If either of them does not hold, the server stops the session; otherwise, the tag is authenticated.

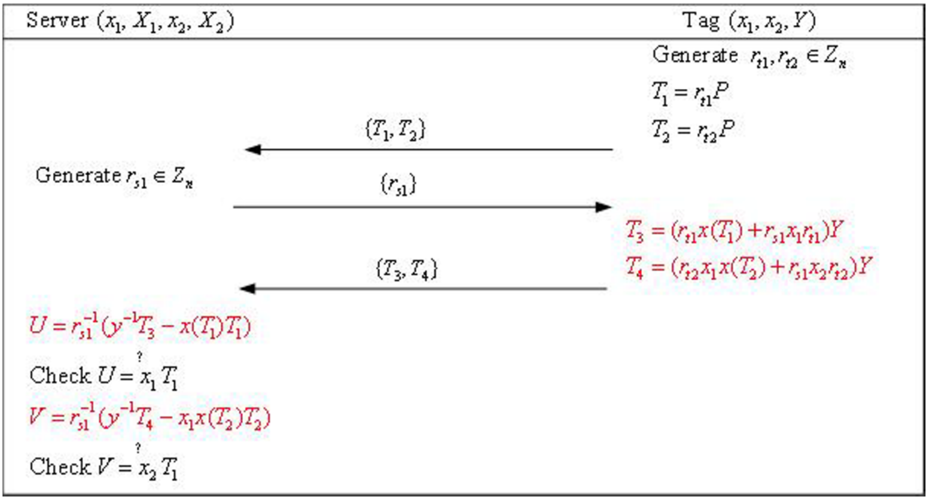

Server (x_1, X_1, x_2, X_2)
Tag (x_1, x_2, Y)
Generate $r_{t1}, r_{t2} \in Z_n$
$T_1 = r_{t1}P$
$T_2 = r_{t2}P$
$\{T_1, T_2\}$
Generate $r_{s1} \in Z_n$
$\{r_{s1}\}$
$T_3 = (r_{t1}x(T_1) + r_{s1}x_1r_{t1})Y$
$T_4 = (r_{t2}x_1x(T_2) + r_{s1}x_2r_{t2})Y$
$\{T_3, T_4\}$
$U = r_{s1}^{-1}(y^{-1}T_3 - x(T_1)T_1)$
Check $U \overset{?}{=} x_1 T_1$
$V = r_{s1}^{-1}(y^{-1}T_4 - x_1x(T_2)T_2)$
Check $V \overset{?}{=} x_2 T_1$

Fig. 9. Lv et al.'s Modified EC-RAC 3 protocol

5. Security analysis

In this section, we just analyze the security of our EC-RAC 1 protocol because security analysis of the other two protocols is similar. We demonstrate that our EC-RAC 1 protocol could provide security properties and withstand various attacks.

Authentication: According to the description of our EC-RAC 1 protocol, it is impossible to generate $T_2 = (r_{t1}x(T_1) + r_{s1}x_1r_{t1})Y$ without the secret key x_1 because the adversary faces the ECDL problem. Thus, the server is able to authenticate the tag by checking if $U = r_{s1}^{-1}(y^{-1}T_2 - x(T_1)T_1)$ and x_1T_1 are equal in step 4 of our EC-RAC 1 protocol.

Anonymity: The adversary may intercepts messages $\{T_1\}$, $\{r_{s1}\}$ and $\{T_2\}$ transmitted between the tag and the server, where $T_1 = r_{t1}P$ and $T_2 = (r_{t1}x(T_1) + r_{s1}x_1r_{t1})Y$. Due to the hardness of the ECDL problem, the adversary cannot get any information about x_1 from T_2 because he does not know the server's secret key y. Thus, our EC-RAC 1 could provide anonymity.

Man-in-the-middle attack: Upon receiving the message $\{T_1\}$ generated by the tag, the adversary generates a random number r_a, computes $T_1' = r_a T_1$ and sends message $\{T_1'\}$ to the server, where $T_1 = r_{t1}P$. Upon receiving the message $\{T_1'\}$, the server generates a random number r_{s1}, and sends the message $\{r_{s1}\}$ to the adversary. Upon receiving the message $\{r_{s1}\}$, the adversary sends it to the tag directly. Upon receiving the message $\{r_{s1}\}$, the tag computes $T_2 = (r_{t1}x(T_1) + r_{s1}x_1 r_{t1})Y$ and sends the message $\{T_2\}$ to the server. Upon intercepting the message $\{T_2\}$, the adversary generates a random number r_a, computes $T_2' = r_a T_2$ and sends message $\{T_2'\}$ to the server. Upon receiving the message $\{T_2'\}$, the server computes $U = r_{s1}^{-1}(y^{-1}T_2' - T_1')$. The server checks whether U and $x_1 T_1'$ are equal. It is easy to check that U and $x_1 T_1'$ are not equal. Then, the server could find the attack. Thus, our EC-RAC 1 protocol could withstand the man-in-the-middle attack.

Impersonation attack: The adversary generates a random number r_{t1}, computes $T_1 = r_{t1}P$ and sends the message $\{T_1\}$ the server. Upon receiving $\{T_1\}$, the server generates a random number r_{s1}, and sends the message $\{r_{s1}\}$ to the adversary. However, the adversary cannot generate $T_2 = (r_{t1}x(T_1) + r_{s1}x_1 r_{t1})Y$ because he does not the secret key x_1. The server could find the attack by checking whether $U = r_{s1}^{-1}(y^{-1}T_2 - x(T_1)T_1)$ and $x_1 T_1$ are equal. Thus, our EC-RAC 1 protocol could withstand the impersonation attack.

Replay attack: Suppose the adversary intercepts messages $\{T_1\}$ and $\{T_2\}$ sent by the tag, where $T_1 = r_{t1}P$ and $T_2 = (r_{t1}x(T_1) + r_{s1}x_1 r_{t1})Y$. The adversary replays $\{T_1\}$ to the server. Upon receiving $\{T_1\}$, the server generates a random number r_{s1}, and sends the message $\{r_{s1}\}$ to the adversary. Then, the adversary replays $\{T_2\}$ to the server. However, the server could find the attack by checking whether $U = r_{s1}^{-1}(y^{-1}T_2 - x(T_1)T_1)$ and x_1T_1 are equal because the server generates a new random number r_{s1} for each session. Thus, our EC-RAC 1 protocol could withstand the replay attack.

Tracking attack: The adversary may intercepts messages $\{T_1\}$, $\{r_{s1}\}$ and $\{T_2\}$ transmitted between the tag and the server, where $T_1 = r_{t1}P$ and $T_2 = (r_{t1}x(T_1) + r_{s1}x_1r_{t1})Y$. However, he cannot get information about tag's identity from those messages because he does not the server's secret key y. Thus, our EC-RAC protocol could withstand the tracking attack.

6. Performance analysis

In this section, we give performance analysis of our three EC-RAC protocols. We also compare the performance of our protocol with that of Lee et al.'s three EC-RAC protocols [21] and Lv et al.'s three EC-RAC protocols [23]. Some notations used in our analysis are defined as follows.

- T_{ma} : the running time of a modular addition operation;
- T_{mm}: the running time of a modular multiplication operation;
- T_{inv} : the running time of a modular inversion operation;
- T_{eca} : the running time an elliptic curve point addition operation;
- T_{ecm}: the running time an elliptic curve point multiplication operation;

Table 1. Computation cost comparison

	The server	The tag
Lee et al.'s EC-RAC 1	$2T_{inv}+1T_{eca}+3T_{ecm}$	$1T_{ma}+1T_{mm}+2T_{ecm}$
Lv et al.'s EC-RAC 1	$2T_{inv}+1T_{eca}+3T_{ecm}$	$1T_{ma}+2T_{mm}+2T_{ecm}$
Our EC-RAC 1	$2T_{inv}+1T_{eca}+4T_{ecm}$	$1T_{ma}+3T_{mm}+2T_{ecm}$
Lee et al.'s EC-RAC 2	$2T_{inv}+2T_{eca}+4T_{ecm}$	$2T_{ma}+3T_{mm}+3T_{ecm}$

Lv et al.’s EC-RAC 2	$2T_{inv}+2T_{eca}+6T_{ecm}$	$2T_{ma}+5T_{mm}+3T_{ecm}$
Our EC-RAC 2	$2T_{inv}+2T_{eca}+7T_{ecm}$	$2T_{ma}+7T_{mm}+3T_{ecm}$
Lee et al.’s EC-RAC 3	$2T_{inv}+2T_{eca}+4T_{ecm}$	$2T_{ma}+3T_{mm}+4T_{ecm}$
Lv et al.’s EC-RAC 3	$2T_{inv}+2T_{eca}+7T_{ecm}$	$2T_{ma}+5T_{mm}+4T_{ecm}$
Our EC-RAC 3	$2T_{inv}+2T_{eca}+8T_{ecm}$	$2T_{ma}+7T_{mm}+4T_{ecm}$

The computational cost comparison of our three EC-RAC protocols, Lee et al.’s three EC-RAC protocols [21] and Lv et al.’s three EC-RAC protocols [23] is demonstrated in Table 1. According to Table 1, the Lee et al.’s EC-RAC 1/2/3 protocol and Lv et al.’s 1/2/3 protocol has better performance than our EC-RAC 1/2/3 protocol. However, Lee et al.’s three EC-RAC protocols [21] suffer from the tracking attack and Lv et al.’s EC-RAC protocols [23] suffer from the man-in-the-middle attack. As a cryptographic protocol, the security is the first important factor in the design of RFID authentication protocol. Our three EC-RAC protocols sacrifice performance slightly to solve the security problems in Lee et al.’s protocols and Lv et al.’s protocol. Therefore, our EC-RAC protocols are more suitable for RFID systems.

7. Conclusions

With the widespread use of the RFID system in our daily life, the design secure RFID authentication protocols attract extensive attention. Recently, ECC-based RFID authentication protocols were studied widely because they could provide better security. Based on Lee et al.’s work, Lv et al. proposed three EC-RAC protocols for authentication in RFID systems. We first demonstrate that Lv et al.’s protocol suffer from the man-in-the-middle attacks. Subsequently, we proposed three security enhanced EC-RAC protocols to solve security problems in Lv et al.’s protocol. Analysis shows that our protocols are more suitable for RFID systems.

Acknowledges

The authors thank the editor and anonymous reviewers for their valuable comments. This study was supported by the National Science foundation of China (Nos. 61402339, 61572370).

References

1. B. Guo, D. Zhang, Z. Yu, Y. Liang, Z. Wang, and X. Zhou, From the Internet of Things to Embedded Intelligence, *World Wide Web Journal*, vol. 16, no. 4, pp. 399-420, 2013.
2. M. Feki, F. Kawsar, M Boussard, and L Trappeniers, The Internet of Things: The Next Technological Revolution, *Computer*, vol. 46, no. 2, pp. 24-25, 2013.
3. R. Das, Rfid market projections 2008–2018, IDTechEx, 2008.
4. Y. Tian, G. Chen, J. Li, A New Ultralightweight RFID Authentication Protocol with Permutation, *IEEE Communication Letters*, vol. 16, no. 5, pp. 702-705, 2012.
5. H. Lee, T. Yi, J. Hyun, Secure and Lightweight Authentication Protocol for Mobile RFID Privacy, *Applied Mathematics & Information Sciences*, vol. 7, no. 1, pp. 421-426, 2013.
6. Y. Lee, Y. Park, A New Privacy-preserving Path Authentication Scheme using RFID for Supply Chain Management, *Advances in Electrical and Computer Engineering*, vol. 13, no. 1, pp. 23-26, 2013.
7. Z. Wu, L. Chen, J. Wu, A Reliable RFID Mutual Authentication Scheme for Healthcare Environments, *Journal of Medical Systems*, vol. 37, no. 2, Article ID: 9917, 2013.
8. C. Yen, M. Lo, N. Lo, Authentication with low-cost RFID tags in mobile networks, *Security and Communication Networks*, vol. 6. no. 8, pp. 1021-1027, 2013.
9. G. Deng, H. Li, Y, Zhang, Tree-LSHB plus : An LPN-Based Lightweight Mutual Authentication RFID Protocol, *Wireless Personal Communication*, vol. 72, no. 1, pp. 159-174, 2013.
10. G. Avoine, M. Bingol, X. Carpent, Privacy-Friendly Authentication in RFID Systems: On Sublinear Protocols Based on Symmetric-Key Cryptography, *IEEE Transactions on Mobile Computing*, vol. 12, no. 10, pp. 2037-2049, 2013.
11. S. Kaul, A. K. Awasthi, RFID Authentication Protocol to Enhance Patient Medication Safety, *Journal of Medical Systems*, vol. 37, no. 6, Article ID: 9979, 2013.
12. M. Dehkordi, Y. Farzaneh, Improvement of the Hash-Based RFID Mutual Authentication Protocol, *Wireless Personal Communication*, vol. 75, no. 1, pp. 219-232, 2014.
13. L. Gao, M. Ma, Y. Shu, An ultralightweight RFID authentication protocol with CRC and permutation, *Journal of Network and Computer Applications*, vol. 41, no. 1, pp. 37-46, 2014.

14. Y. Lee, K. Sakiyama, L. Batina, I. Verbauwhede, Elliptic curve-based security processor for RFID, *IEEE Transactions on Computers*, vol. 57, no. 11, pp. 1514-1527, 2008.
15. Y. Liao, C. Hsiao, A secure ECC-based RFID authentication scheme integrated with ID-verifier transfer protocol, *Ad Hoc Networks*, vol. 18, no. 1, pp. 133-146, 2014.
16. Z. Zhang, Q. Qi, An Efficient RFID Authentication Protocol to Enhance Patient Medication Safety Using Elliptic Curve Cryptography, *Journal of Medical Systems*, vol. 38, no. 5, Article ID: 47, 2014.
17. Z. Zhao, A Secure RFID Authentication Protocol for Healthcare Environments Using Elliptic Curve Cryptosystem, *Journal of Medical Systems*, vol. 38, no. 5, Article ID: 46, 2014.
18. Y. Lee, L. Batina, I. Verbauwhede, EC-RAC (ECDLP based randomized access control): provably secure RFID authentication protocol, In: *IEEE International Conference on RFID 2008*, pp. 97-104, 2008.
19. J. Bringer, H. Chabanne, T. Icart, Cryptanalysis of EC-RAC, a RFID identification protocol. In: *7th International Conference on Cryptology And Network Security-CANS'08*, pp. 149-161, 2008.
20. T. Deursen, S. Radomirovic, Attacks on RFID protocols (version 1.1), *Technical Report*, University of Luxembourg, 2009.
21. Y. Lee, L. Batina, I. Verbauwhede, Untraceable RFID authentication protocols: revision of EC-RAC, In: *IEEE International Conference on RFID 2009*, pp. 178–185, 2009.
22. T. Deursen, S. Radomirovic, Untraceable RFID protocols are not trivially composable: attacks on the revision of EC-RAC. *Technical Report*, University of Luxembourg, 2009.
23. C. Lv, H. Li, J. Ma, Y. Zhang, Vulnerability analysis of elliptic curve cryptography-based RFID authentication protocols, *Transactions on Emerging Telecommunications Technologies*, vol. 23, no. 7, pp. 618–624.
24. S. Vaudenay, On privacy models for RFID, In: *Advances in Cryptology - Asiacrypt 2007*, pp. 68–87, 2007.
25. A. Juels, S. Weis, Defining strong privacy for RFID, *ACM Transactions on Information and System Security*, vol. 13, no. 1, pp. 1–23, 2009.
26. G. Avoine, Adversarial model for radio frequency identification, *Cryptology ePrint Archive*, Report 2005/049, 2005.
27. C. Ng, W. Susilo, Y. Mu, R. Safavi-Naini, RFID privacy models revisited. In: *Proceedings of the 13th European Symposium on Research in Computer Securit*y, pp. 251–266, 2008.

On the Security of Three-factor Authentication Scheme for Telecare Medical Information Systems

Qi Jiang, Bingyan Li, Jianfeng Ma
School of Cyber Engineering, Xidian University, Xi'an, China
jiangqixdu@gmail.com

Abstract. Although a number of three-factor authentication schemes have been developed to ensure that sensitive medical information are only available to legal users in telecare medical information system, most of them are found to be flawed. Understanding security and privacy failures of authentication protocols is a prerequisite to both fixing existing protocols and designing future ones. In this paper, we analyze an enhanced three-factor authentication scheme of Lu et al., and reveal that it cannot achieve the claimed security and privacy goals. (1) It fails to provide anonymity and untraceability, and is susceptible to the following attacks targeting user privacy: identity revelation attack and tracking attack. (2) It is also susceptible to offline password guessing attack, user impersonation attack, and server impersonation attack.

1 Introduction

The provision of medical services using information and communication technologies has given rise to the emergence of telecare medical information system (TMIS) [1], which enables the public to access medical services or medical information at remote sites. In TMIS, the medical server maintains highly sensitive medical information of registered users, such as electronic medical record (EMR), which is being shared and accessed via public channel by the doctors, hospitals, medical institute and academia to enhance their decision.

However, such an adoption also brings about a series of challenges, especially, how to ensure the security and privacy of medical information from various attacks over public networks, such as eavesdropping and tampering [2-5]. Among them, mutual authentication between the user and medical server is indispensable to ensure that sensitive medical information is only available to legal users. Additionally, there is a growing demand to protect the privacy of user identity. The medical information may be of interest to different types of users ranging from patients to doctors, hospitals, and insurance companies. The activities of these users may be of great sensitiveness to the outsiders. Therefore, it is desirable to design a privacy preserving three-factor authentication scheme for TMIS.

Awasthi and Srivastava [6] presented a biometric authentication scheme based on nonce for TMIS in 2013. Later, Tan et al. [7] revealed that Awasthi and Srivastava's scheme is vulnerable to reflection attack and cannot provide three factor security and user anonymity. Then, Tan et al. [7] put forward a three factor authentication scheme and claimed that their scheme is immune from various known attacks. Mishra et al. [8]

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_89

also found that demonstrate Awasthi and Srivastava's scheme suffers from online and offline password guessing attack. They further proposed an enhanced three-factor authentication scheme for TIMS.

More recently, Arshad and Nikooghadam [9] pointed out that Tan et al.'s scheme cannot resist DoS and replay attacks. Then they presented an enhanced elliptic curve cryptosystem (ECC)-based scheme to prevent the flaws. Das [10] revealed that both Tan's scheme [7] and Arshad and Nikooghadam's scheme [9] are insecure. Lu et al. [11] also demonstrated Arshad and Nikooghadam's scheme [9] fails to protect against off-line password guessing attack and user impersonation attack. Furthermore, they put forward a new biometric based authentication scheme for TMIS.

In this paper, we use Lu et al.'s scheme [11] as a case study and demonstrate the subtleties and challenges in designing a practical three-factor scheme preserving user privacy. We demonstrate it fails to protect user privacy, and is susceptible to the following attacks targeting user privacy: identity revelation attack and tracking attack. Moreover, although Lu et al. claimed that their scheme can defend offline password guessing attack and impersonation attack, we identify that their scheme is susceptible to offline password guessing attack, user impersonation attack, and server impersonation attack.

The rest of the paper is organized as follows. Next section defines the adversarial model for three-factor authentication. Section 3 briefly reviews Lu et al.'s scheme. Section 4 elaborates on the weaknesses of their scheme. Finally, the final section concludes the paper.

2 Adversarial Model

The adversarial model is one of the key factors in evaluating a three-factor authentication protocol. There have been dozens of papers concerning three-factor authentication in recent years, yet only a few ones explicitly define the adversarial model in their work. The capabilities of the adversary are summarized as follows, which are based on those for two-factor authentication refined in [12].

(1) The adversary A has full control of the communication channel between the protocol participants. Any message transmitted through the channels may be intercepted, inserted, deleted, or modified. (2) The adversary A can enumerate offline all the items in the Cartesian product $D_{ID} * D_{PW}$ within polynomial time, where D_{ID} and D_{PW} denote the identity space and password space, respectively. (3) The adversary may either (i) compromise the password and the biometric template, or (ii) compromise the password and the smart card, or (iii) compromise the biometric template and the smart card, but cannot compromise all the three factors. When a smart card is compromised, such as being lost/stolen, the secret information in it can be extracted by side channel attacks. (4) The adversary has the capability of learning the long term private key(s) only when evaluating perfect forward secrecy. (5) The adversary may learn the previous session keys.

Note that a truly three-factor authentication protocol should still be secure even any two of three factors are compromised.

3 Review of Lu et al.'s Scheme

In this section, we now briefly review Lu et al.'s scheme [11], which is consists of three phases, namely registration, login and authentication, password updating. We only review the first two phases due to limited space.

3.1 Registration Phase

(1) A user U_i selects an identity ID_i, a password PW_i, and imprints the biometrics B_i. Then he computes $MP_i = PW_i \oplus H(B_i)$ and sends $\{ID_i, MP_i\}$ to the remote server for the TMIS S.

(2) Upon receiving $\{ID_i, MP_i\}$, S computes $AID_i = ID_i \oplus h_2(x)$, $V_i = h_1(ID_i \| MP_i)$, and S issues a smart card SC_i containing $\{V_i, AID_i, h_1(\cdot), h_2(\cdot), H(\cdot)\}$ to U_i, where x is S's secret key, $H(\cdot)$ is biohash function, $h_1(\cdot)$ and $h_2(\cdot)$ are hash function.

3.2 Login and Authentication Phase

(1) U_i attaches SC_i to a card reader, inputs ID_i, PW_i, and imprints B_i. SC_i computes $h_1(ID_i \| PW_i \oplus H(B_i))$, and verifies whether it is equal to the value V_i. If it holds, SC_i selects a random number d_u, computes $K = h_1(ID_i \| ID_i \oplus AID_i)$, $M_1 = K \oplus d_u P$, $M_2 = h_1(ID_i \| d_u P \| T_1)$, and transmits the login message $\{AID_i, M_1, M_2, T_1\}$ to S.

(2) On receiving the login message, S first checks whether $|T_1 - T_c| < \Delta T$, where T_c is the current timestamp of S. If it is invalid, S rejects the request. Otherwise, S derives ID_i by computing $AID_i \oplus h_2(x)$, computes $K = h_1(ID_i \| h_2(x))$, $d_u P = K \oplus M_1$, and then compares whether $h_1(ID_i \| d_u P \| T_1)$ is equal to M_2. If they are equal, S chooses a random number d_s, and computes $SK = d_s d_u P$, $M_3 = K \oplus d_s P$, $M_4 = h_1(K \| d_u P \| SK \| T_2)$. Finally, S sends the second message $\{M_3, M_4, T_2\}$ to U_i.

(3) On receiving the second message, SC_i verifies the validity of T_2. If it is valid, U_i computes $d_s P = M_3 \oplus K$, $SK = d_u d_s P$, and checks whether $h_1(K \| d_u P \| SK \| T_2)$ is equal to M_4 in the received message. If it holds, U_i computes $M_5 = h_1(K \| d_s P \| SK \| T_3)$ and then sends the third message $\{M_5, T_3\}$ to S.

(4) Upon receiving $\{M_5, T_3\}$, S confirms whether $|T_3 - T_c| < \Delta T$ and $M_5 \overset{?}{=} h_1(K \| d_s P \| SK \| T_3)$. If both hold, S accepts U_i as a legal user.

4 Cryptanalysis of Lu et al.'s Scheme

Although Lu et al.'s scheme is claimed to be secure against various known attacks, we observe that the scheme is prone to the following attacks.

4.1 Identity Revelation Attack

A malicious user A can reveal the identity through capturing the login request message. Suppose that A gets a legal smart card from the server first. Then he can manipulate the smart card to carry out the identity revelation attack as follows.

(1) A inserts his smart card into a card reader and inputs ID_A, PW_A, and B_A, and follows the protocol procedure. Then the card sends a request message $\{AID_A, M_1, M_2, T_1\}$ to S.

(2) A extracts AID_A by capturing the message $\{AID_A, M_1, M_2, T_1\}$, where $AID_A = ID_A \oplus h_2(x)$. Next, A picks one message $\{AID_i, M_1, M_2, T_1\}$, where $AID_i = ID_i \oplus h_2(x)$. Then A can reveal the identity of U_i by computing $ID_i = AID_i \oplus h_2(x) = AID_i \oplus (AID_A \oplus ID_A)$, for $h_2(x) = ID_A \oplus AID_A$.

4.2 Tracking Attack

AID_i in the login message $\{AID_i, M_1, M_2, T_1\}$ is fixed for a specific smart card, since it is determined by the user's identity ID_i and the static secret of the server S.

To mount tracking attack, A first eavesdrops a legal user U_i's login message $\{AID_i, M_1, M_2, T_1\}$, then retrieves and stores AID_i. With the knowledge of AID_i, A can endanger user privacy. A might track the user and try to collect the user's personal information, access patterns, TMIS usage habit, etc. Therefore, Lu et al.'s scheme [40] cannot withstand tracking attack and fails to achieve user untraceability.

4.3 Offline Password Guessing Attack

In [11], Lu et al. claimed that the scheme is secure against offline password guessing attack based on the assumption that the user identity and biometrics are unknown. As is indicated by Wang et al. [12], it is more practical to regard user identity as a known value. Suppose that a smart card is stolen or lost, and then the adversary extracts the secret information $V_i = h_1(ID_i \| MP_i) = h_1(ID_i \| PW_i \oplus H(B_i))$ from the card. With the secret parameter $V_i = h_1(ID_i \| MP_i)$ and the biometric B_i of the user, the adversary can mount offline password guessing attack by picking ID_i' a password PW_i' and checking whether $V_i = h_1(ID_i' \| PW_i' \oplus H(B_i))$ holds. Therefore, their scheme is vulnerable to stolen smart card attack.

4.4 User Impersonation Attack

Suppose an adversary A captures the message $\{AID_i, M_1, M_2, T_1\}$ transmitted between U_i and S. Then A can obtain the identity of U_i through the identity revelation attack described above. After that, A can construct a legal request message. The details are given as follows.

(1) A selects a random number d_a, computes $K = h_1(ID_i \| ID_i \oplus AID_i)$, $M_1^* = K \oplus d_a P$, $M_2^* = h_1(ID_i \| d_a P \| T_1^*)$, and transmits $\{AID_i, M_1^*, M_2^*, T_1^*\}$ to S.

(2) Upon receiving the request message, derives ID_i by computing $AID_i \oplus h_2(x)$, computes $d_a P = h_1(ID_i \| h_2(x)) \oplus M_1$, and then compares whether $h_1(ID_i \| d_a P \| T_1^*)$ is equal to M_2^*. It is obvious that $h_1(ID_i \| d_a P \| T_1^*)$ computed by S is equal to M_2^* received. That is, A has passed the verification of S.

(3) A follows the specified procedure to perform the following steps and computes $M_5 = h_1(K \| d_s P \| SK \| T_3)$.

(4) S follows the specified procedure, it is easy to see that M_5 will pass the verification of S, as A and S share the same key SK.

The root cause of the above attack is that the only secret information in the request message is the value ID_i. This proves that this scheme fails to fulfill three-factor security.

4.5 Server Impersonation Attack

Suppose A captures the message $\{AID_i, M_1, M_2, T_1\}$ transmitted between U_i and S. Then A can obtain the identity of U_i through the identity revelation attack or offline password guessing attack described above. After that, A can lunch server impersonation attack as follows.

(1) U_i follows the specified procedure and sends $\{AID_i, M_1, M_2, T_1\}$ to S.

(2) A intercepts the message $\{AID_i, M_1, M_2, T_1\}$, and computes $h_2(x) = ID_i \oplus AID_i$, $K = h_1(ID_i \| h_2(x))$, $d_u P = K \oplus M_1$. Then A chooses a random number d_a, and computes $SK = d_a d_u P$, $M_3 = K \oplus d_a P$, $M_4 = h_1(K \| d_u P \| SK \| T_2)$. Finally, A sends $\{M_3, M_4, T_2\}$ to U_i.

(3) Upon receiving the response message, U_i computes $d_a P = M_3 \oplus K$, $SK = d_u d_a P$, and checks whether $h_1(K \| d_a P \| SK \| T_2)$ is equal to M_4 in the received message. It is easy to see that the they are equal. That is, A has passed the verification performed by U_i. U_i computes $M_5 = h_1(K \| d_s P \| SK \| T_3)$ and then sends the message $\{M_5, T_3\}$ to S.

(4) A intercepts the message $\{M_5, T_3\}$, and follows the specified procedure to perform the following steps. Finally, A and U_i share the same key SK.

5 Conclusions

In this paper, we have analyzed an enhanced three-factor authentication scheme of Lu et al., and have revealed that it cannot achieve the claimed security and privacy goals. (1) It fails to provide anonymity and untraceability, and is susceptible to identity revelation attack and tracking attack. (2) It also fails to accomplish mutual authentication, and is susceptible to offline password guessing attack, user impersonation attack, and server impersonation attack. Our cryptanalysis results discourage any practical deployment of this scheme and indicate that it still remains as a hard problem to develop a privacy-preserving three-factor protocol.

Acknowledgments. This work is supported by National Natural Science Foundation of China (Program Nos. 61672413, U1405255, U1536202,), Natural Science Basic Research Plan in Shaanxi Province of China (Program No. 2016JM6005), Fundamental Research Funds for the Central Universities (Program No. JB161501).

References

1. Li, S. H., Wang, C. Y., Lu, W. H., Lin, Y. Y., & Yen, D. C.: Design and Implementation of a Telecare Information Platform. J. Med. Syst. 36(3) (2012) 1629-1650
2. Li, H., Yang, Y., Luan, T., Liang, X., Zhou, L., Shen, X.: Enabling Fine-grained Multi-keyword Search Supporting Classified Sub-dictionaries over Encrypted Cloud Data. IEEE Transactions on Dependable and Secure Computing 13(3) (2016) 312-325
3. Jiang Q., Ma Z., Ma J., Li G.: Security Enhancement of a Robust User Authentication framework for Wireless Sensor Networks. China Communications 9(10) (2012) 103-111
4. Jiang Q., Ma J., Li G., Yang L.: Robust Two-factor Authentication and Key Agreement Preserving User Privacy. International Journal of Network Security 16(3)(2014) 229-240
5. Jiang Q., Wei F., Fu S., Ma J., Li G., Alelaiwi A.: Robust Extended Chaotic Maps-based Three-factor Authentication Scheme Preserving Biometric Template Privacy. Nonlinear Dynamics 83(4) (2016) 2085-2101
6. Awasthi, A.K., Srivastava, K.: A Biometric Authentication Scheme for Telecare Medicine Information Systems with Nonce. J. Med. Syst. 37(5) (2013) 1–4
7. Tan, Z.: A User Anonymity Preserving Three-factor Authentication Scheme for Telecare Medicine Information Systems. J. Med. Syst. 38(3) (2014) 1–9
8. Mishra, D., Mukhopadhyay, S., Kumari, S., Khan, M.K., Chaturvedi, A.: Security Enhancement of a Biometric based Authentication Scheme for Telecare Medicine Information Systems with Nonce. J. Med. Syst. 38(5): (2014) 1–11
9. Arshad, H., Nikooghadam, M.: Three-factor Anonymous Authentication and Key Agreement Scheme for Telecare Medicine Information Systems. J. Med. Syst. 38(12) (2014) 1-12
10. Das, A.K.: A Secure User Anonymity-Preserving Three-Factor Remote User Authentication Scheme for the Telecare Medicine Information Systems. J. Med. Syst. 39(3) (2015) 1-20
11. Lu, Y., Li, L., Peng, H., Yang, Y.: An Enhanced Biometric-based Authentication Scheme for Telecare Medicine Information Systems Using Elliptic Curve Cryptosystem. J. Med. Syst. 39(3) (2015)
12. Wang D., He D., Wang P., Chu C.-H.: Anonymous Two-factor Authentication in Distributed Systems: Certain Goals are Beyond Attainment. IEEE Transactions on Dependable and Secure Computing 12(4) (2015)428-442.

Oblivious Transfer Protocols Based on Group Factoring Problem ⋆

Jing Li[1], Xiong Li[2], Licheng Wang*[1], Debiao He[3], and Xinxin Niu[1]

[1] State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China
[2] School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411201, China
[3] State Key Laboratory of Software Engineering, Computer School, Wuhan University, Wuhan, China

Abstract. In this paper, we propose 1-out-of-n oblivious transfer protocol by using the group of matrices over group ring $Z_q[S_m]$. The security of the proposal is on the basis of factorization problems of non-commutative algebraic structures. Meanwhile, some new intractable assumptions are defined based on the group factorization problem (GFP). Subsequently, we present a simpler 1-out-of-n oblivious transfer construction for underlying non-commutative group. Furthermore, to achieve the oblivious transfer for more challenged messages, an efficient k-out-of-n oblivious transfer protocol with fewer public parameters is designed based on the newly defined hard assumptions.

Keywords: Oblivious Transfer, Matrices over Group Rings, Group Factorization Problem

1 Introduction

Oblivious transfer (OT) introduced by Rabin [16] is a protocol to solve the problem of users' privacy preservation in electronic commerce, where one-party, referred as the sender, has some messages to send, while the other party, referred as the receiver, learns some information about his chosen message, but nothing more; and the sender could never figure out the information obtained by receiver. Similarly, 1-out-of-2 OT protocol [5], 1-out-of-n OT protocol [1] and k-out-of-n OT protocol [8] are designed for achieving different functionalities of OT.

During the past three decades, we have witnessed a variety of OT protocols, which were designed relying upon different mathematical platforms [2, 4, 11, 15, 16, 18]. For example, Rabin OT protocol [16] was designed based on large integers factorization problem (IFP); the protocols given in [2, 4] were constructed on the basis of discrete logarithm problem (DLP); the one reported in [15] was proposed using elliptic curves; and OT protocol in [11] was depending on bilinear pairing. In general, most oblivious transfer protocols are constructed using

⋆ This work was supported by the National Natural Science Foundation of China under Grant Nos. 61370194, 61411146001.

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_90

the intractability problems of IFP and DLP over a large finite field. However, Shor [17] put forward the quantum attack algorithms to solve the IFP problem and the DLP problem in a polynomial time. Fortunately, an vigorous branch of cryptography based on the hard problem of non-commutative algebraic structures comes to us [6, 7, 9, 10, 12, 13]. This branch is called the post-quantum since the cryptographic schemes based on the non-commutative algebraic structures are secure under quantum attack algorithms. The group factorization problem (GFP) is one typical hard problem.

Our contributions. In this work, we define new hard assumptions based on the GFP. Then a novel 1-out-of-n oblivious transfer protocol is proposed by using the group of matrices over group rings $Z_q[S_m]$. Based on this idea, we further simplify the above construction and obtain a more practical version. The security of the two proposed OT protocols depends on the newly defined intractable assumptions of matrices over the group ring $Z_q[S_m]$, as well as the non-commutativity of this algebraic structure. Furthermore, to achieve the oblivious transfer for more challenged messages, we present an efficient k-out-of-n OT protocol with fewer public parameters.

2 Preliminaries

2.1 Background of group rings

Now we review some definitions over group rings.

Definition 1 (Group Rings [9]). *Let G be a group and R be a ring. The group ring $R[G]$ is the set of all formal sums*

$$\sum_{g_i \in G} r_i g_i$$

where $r_i \in R$.

The addition of $R[G]$ [9] is defined as

$$\sum_{g_i \in G} a_i g_i + \sum_{g_i \in G} b_i g_i = \sum_{g_i \in G} (a_i + b_i) g_i.$$

The multiplication of $R[G]$ [9] is given as follows:

$$\sum_{g_i \in G} a_i g_i \cdot \sum_{g_i \in G} b_i g_i = \sum_{g_i \in G} (\sum_{g_j g_k = g_i} a_j b_k) g_i.$$

For example, $Z_q[S_m]$ is a group ring, where $R = Z_q$ and G is the m-degree symmetric group S_m.

Definition 2 (Matrices of Group Rings [9]). *A matrix over a group ring $R[G]$ is such a matrix in which the values of elements are determined by the group ring $R[G]$. $M_l(Z_q[S_m])$ is a set of $l \times l$ matrices over $Z_q[S_m]$.*

Definition 3 (Semidirect product [7]). *Let G_1, G_2 be two groups. Suppose that $Aut(G_1)$ is the group of automorphisms of G_1 and ρ: $G_2 \rightarrow Aut(G_1)$ is a homomorphism. Then the semidirect product of G_1 and G_2 is the set*

$$\varGamma = G_1 \times_\rho G_2 = \{(g_1, g_2) : g_1 \in G_1, g_2 \in G_2\}$$

with the group operation defined by

$$(g_{11}, g_{21})(g_{12}, g_{22}) = (g_{11}^{\rho(g_{22})} \cdot g_{12}, \quad g_{21} \cdot g_{22}). \tag{1}$$

Here, $g_{11}^{\rho(g_{22})}$ denotes the image of g_{11} under the automorphism $\rho(g_{22})$.

2.2 Intractability Assumptions

Let G be a non-communicative group of prime order p. The hard problems are defined as below.

- **Group Factorization Problem (GFP)[14]**: Given G_1, G_2, G and $g_1 \cdot g_2 \in G$ as input, output $g_1 \in G_1$ and $g_2 \in G_2$, where $g_1 \cdot g_2 \neq g_2 \cdot g_1$ and $G_1, G_2 \subset G$.

Based on the description of the GFP, we define some new hard problems.

- **Mid-Extract Problem (MEP)**: Given $g_2 \in G$ and $g_1 \cdot g_2 \cdot g_3 \in G$ as input, where $g_i \cdot g_j \neq g_j \cdot g_i$ for $1 \leq i \neq j \leq 3$, output $g_1 \cdot g_3$.

Note that, we need to remove the mid-element for solving the MEP. Maybe, g_1 and g_3 are unknown to us. Thus, the problem has lower complexity than the GFP. In addition, the corresponding **Decisional MEP** is given as follows:

- **Decisional Mid-Extract Problem (DMEP)**: Given a tuple (g_2, $g_1 g_2 g_3$) and a random element $T \in G$ as input, decide whether or not $T = g_1 \cdot g_3$.

Remark 1. Recently, quantum algorithm for solving discrete logarithm problem for group ring matrices was proposed in [13]. However, for the factorization problem over the semigroup of matrices over group rings, there doesn't exist an efficient algorithm either in classical or in quantum.

3 Warming up

3.1 1-out-of-*n* OT protocol using matrices over group ring $Z_q[S_m]$

Now we present 1-out-of-n OT protocols based on semidirect product $\varGamma$. Define a projection operator σ for $\varGamma$: $\sigma(X, Y) = X$, where $(X, Y) \in \varGamma$. The scheme is constructed as follows:

Setup: Suppose that λ is a secure parameter and messages $m_i \in \{0,1\}^\lambda$ ($i = 1, \ldots, n$). Let $h(\cdot)$: $GL_l(Z_7[S_5]) \rightarrow \{0,1\}^\lambda$ be a collision-resistance hash function. Alice and Bob randomly select invertible matrices M_i, $H_i \in GL_l(Z_7[S_5])$, where matrices M_i, H_i are public keys for $i = 1, \ldots, n$. Then Alice and Bob choose their private keys s_A, $s_B \in \{0,1\}^\lambda$ at random, respectively.

Transfer: Alice and Bob execute the following interactions:

- Alice computes $A_i \leftarrow \sigma((M_i, \varphi_{H_i})^{s_A})$. Then Alice computes ciphertext $c_i \leftarrow h(A_i) \oplus m_i$ for $i = 1, \ldots, n$.
- Bob makes his choice $v \in \{1, \ldots, n\}$ and computes $B_v \leftarrow \sigma((M_v, \varphi_{H_v})^{s_B})$. Then he sends B_v to Alice.
- Alice chooses an automorphism ϕ and computes $K_i \leftarrow \sigma((B_v, \phi)(A_i, \varphi_{H_i}^{s_A}))$ using Eq.(2). Then Alice sends K_i, c_i to Bob for $i = 1, \ldots, n$.
- Bob chooses K_v and c_v. Finally, he computes $A_v \leftarrow H_v^{s_B} K_v (H_v M_v)^{-s_B}$ and recovers message $m_v \leftarrow c_v \oplus h(A_v)$.

The correctness of the proposed scheme can be obtained immediately:

- Alice computes

$$\begin{aligned}(M_i, \varphi_{H_i})^{s_A} &= (H_i^{-s_A+1} M_i H_i^{s_A-1} \cdots H_i^{-2} M_i H_i^2 \cdot H_i^{-1} M_i H_i \cdot M_i, \varphi_{H_i}^{s_A}) && (2)\\ &= (H_i^{-s_A}(H_i M_i)^{s_A}, \varphi_{H_i}^{s_A}), && (3)\end{aligned}$$

  then $A_i = H_i^{-s_A}(H_i M_i)^{s_A}$ for $i = 1, \ldots, n$. Similarly, $B_v = H_v^{-s_B}(H_v M_v)^{s_B}$.
- Bob calculates $(B_v, \phi)(A_i, \varphi_{H_i}^{s_A})$ =$(\varphi_{H_i}^{s_A}(B_v)$ $\cdot A_i, \phi \cdot \varphi_{H_i}^{s_A})$, where

$$\begin{aligned}\varphi_{H_i}^{s_A}(B_v) \cdot A_i &= H_i^{-s_A} B_v H_i^{s_A} \cdot A_i && (4)\\ &= H_i^{-s_A}[H_v^{-s_B}(H_v M_v)^{s_B}] H_i^{s_A} \cdot [H_i^{-s_A}(H_i M_i)^{s_A}] && (5)\\ &= H_i^{-s_A} H_v^{-s_B} (H_v M_v)^{s_B} (H_i M_i)^{s_A}. && (6)\end{aligned}$$

  then $K_i = H_i^{-s_A} H_v^{-s_B} (H_v M_v)^{s_B} (H_i M_i)^{s_A}$ for $i = 1, \ldots, n$.
- Bob recovers A_v by computing

$$\begin{aligned}H_v^{s_B} K_v (H_v M_v)^{-s_B} &= H_v^{s_B}[H_v^{-(s_A+s_B)}(H_v M_v)^{s_A+s_B}](H_v M_v)^{-s_B} && (7)\\ &= H_v^{-s_A}(H_v M_v)^{s_A} && (8)\\ &= A_v. && (9)\end{aligned}$$

Remark 2. It is worth noting that in the protocol only the first component of each pair is sent between Alice and Bob. Thus, any discrete logarithm in the semigroup won't be revealed by the two parties. Hence, the protocol is secure against the quantum algorithm for solving discrete logarithms in a semigroup. Namely, any adversary cannot get secret-key s_A (or s_B) by solving discrete logarithms. (The algorithm of solving discrete logarithms in a semigroup has been reported in [3].)

3.2 Security analysis

A secure OT protocol ensures the privacy of both sender and receiver.

- **Privacy of Alice**: Keeping A_i $(1 \le i \ne v \le n)$ secret from Bob.
  In the proposal, Bob receives $K_i = (H_i^{-s_A})(H_v^{-s_B}(H_v M_v)^{s_B})(H_i M_i)^{s_A}$ $(i = 1, \ldots, n)$. Thus, Bob needs to solve the MEP if he wants to derive A_i $(1 \le i \ne v \le n)$ from K_i. Based on the MEP assumption, Bob cannot obtain either A_i or message m_i $(1 \le i \ne v \le n)$.

- **Privacy of Bob**: Preventing Alice from knowing Bob's choice v.
  (1) Alice fails to decompose $B_v = H_v^{-s_B}(H_v M_v)^{s_B}$ into $H_v^{-s_B}$, $(H_v M_v)^{s_B}$ based on the group factorization problem. Assume that Alice can solve the GFP, then the number v will be obtained by finding the matrix over set $\{H_i : \ i = 1, \ldots, n\}$, which is commutative with $H_v^{-s_B}$.
  (2) On the other hand, $B_v = H_v^{-s_B}(H_v M_v)^{s_B}$ and a random choosen matrix $T \in GL_l(Z_7[S_5])$ are indistinguishable for any uniformly s_B.
  Thus, it is difficult for Alice to establish a relationship between B_v and v.

3.3 A Simpler Construction

The above OT scheme is a prototype inspired by the key-exchange protocol in [7]. Actually, it can be further simplified and generalized.

Setup: Suppose that λ is a security parameter and messages $m_i \in \{0,1\}^\lambda$ $(i = 1, \ldots, n)$. Define a collision-resistance hash function $h(\cdot)$: $GL_l(Z_7[S_5]) \to \{0,1\}^\lambda$. Alice and Bob select invertible matrices M_i, $H_i \in GL_l(Z_7[S_5])$, where matrices M_i, H_i are public keys satisfying $\langle M_i \rangle \bigcap \langle H_i \rangle = \{I\}$ for $i = 1, \ldots, n$. Then Alice and Bob choose their private keys s_A, $s_B \in \{0,1\}^\lambda$ at random, respectively.

Transfer: Alice and Bob execute the following operations:

- Alice computes $A_i \leftarrow H_i^{s_A} M_i^{s_A}$. Then Alice computes ciphertext $c_i \leftarrow h(A_i) \oplus m_i$ $(i = 1, \ldots, n)$.
- Bob calculates $B_v \leftarrow H_v^{s_B} M_v^{s_B}$ for his choice v $(1 \leq v \leq n)$. Then Bob sends B_v to Alice.
- Alice computes $K_i \leftarrow H_i^{s_A} B_v M_i^{s_A}$. Then Alice sends K_i, c_i to Bob $(i = 1, \ldots, n)$.
- Bob selects K_v and c_v. Finally, he computes $A_v \leftarrow H_v^{-s_B} K_v M_v^{-s_B}$ and recovers message $m_v \leftarrow c_v \oplus h(A_v)$.

4 Efficient k-out-of-n OT protocol

Based on the above constructions based on group factorization problem, we propose an efficient k-out-of-n OT protocol.

4.1 Protocol Description

Setup: Suppose that λ is a secure parameter and messages $m_i \in \{0,1\}^\lambda$ $(i = 1, \ldots, n)$. Define a collision-resistance hash function h: $GL_l(Z_7[S_5]) \to \{0,1\}^\lambda$. Alice and Bob select matrices X_i $(i = 1, \ldots, n)$ and invertible matrices $M, H \in GL_l(Z_7[S_5])$, where matrices M, H and X_i are public keys. Then Alice and Bob chooses their private keys s_A and s_B, respectively.

Transfer: Alice and Bob execute the following operations:

- Alice computes $A_i \leftarrow H^{s_A} X_i M^{s_A}$. Then Alice calculates ciphertext $c_i \leftarrow h(A_i) \oplus m_i$ $(i = 1, \ldots, n)$.

- Bob computes $B_j \leftarrow H^{s_B} X_{i_j} M^{s_B}$ for his choice j ($j = 1, \cdots, k$ and $1 \leq i_j \leq n$). Then Bob sends B_j ($j = 1, \cdots, k$) to Alice.
- Alice computes $K_j \leftarrow H^{s_A} B_j M^{s_A}$. Then Alice sends K_j, c_i to Bob ($j = 1, \cdots, k$ and $i = 1, \ldots, n$).
- Bob selects c_{i_j} ($j = 1, \cdots, k$). Then he computes $A_{i_j} \leftarrow H^{-s_B} K_j M^{-s_B}$ and recovers messages $m_{i_j} \leftarrow c_{i_j} \oplus h(A_{i_j})$ for $j = 1, \cdots, k$.

The correctness of the proposed scheme is shown as follows:

- Alice computes K_j as

$$\begin{aligned} K_j &= H^{s_A} B_j M^{s_A} & (10)\\ &= H^{s_A}(H^{s_B} X_{i_j} M^{s_B}) M^{s_A} & (11)\\ &= H^{s_A+s_B} X_{i_j} M^{s_A+s_B}. & (12)\end{aligned}$$

- Bob recovers A_{i_j} as

$$\begin{aligned} H^{-s_B} K_j M^{-s_B} &= H^{s_A+s_B-s_B} X_{i_j} M^{s_A+s_B-s_B} & (13)\\ &= H^{s_A} X_{i_j} M^{s_A} & (14)\\ &= A_{i_j}. & (15)\end{aligned}$$

The k-out-of-n OT protocol also depends on group factorization problem. Therefore, the security can be analyzed using the same method as presented in Section 3.2. Note that, B_j don't reveal any information about i_j ($j = 1, \cdots, k$). Besides, matrices X_i is not required to be invertible.

4.2 Efficiency analysis

This section will present the efficiency analysis from the aspects of the communication cost and the computational. For simplicity, we denote the proposal reported in Section 3.1 by **Protocol 1**, the proposal in Section 3.3 by **Protocol 2a**, and the k-out-of-n construction of Protocol 2a is denoted by **Protocol 2b**, the proposal in Section 4.1 by **Protocol 3**. The description of Protocol 2b is easy to achieve and thus omitted.

Analysis on Protocol 3.

Communication cost. In the construction of Protocol 3, two passes are needed in the transfer phase: Bob sends one matrix B_j ($j = 1, \ldots, k$) to Alice in Pass 1; Alice sends K_j, c_i ($j = 1, \ldots, k$ and $i = 1, \ldots, n$) to Bob in Pass 2. Here, the communication cost for sending one upper triangular matrix (in the form of V) over group ring $Z_7[S_5]$ is bounded by $4 \times \log(7 \times 120) + 6 \times 120 \times \log 7 = 2200$ bits. Observe that $c_i \in \{0, 1\}^\lambda$ and $K_i \in GL_l(Z_7[S_5])$. Thus, in Pass 1 Bob sends $2200k$ bits to Alice, then Alice sends $2200k + \lambda n$ bits to Bob in Pass 2. Thus, Protocol 3 (the k out of n OT protocol) requires the communication cost of $\lambda n + 4400k$ bits in total.

Computational cost. The computational cost is analyzed for Alice and Bob.

- Alice computes $A_i = H^{s_A} X_i M^{s_A}$ $(i = 1, \ldots, n)$ and needs $2 \log s_A + 2n$ matrix multiplications. After that, she computes $K_j = H^{s_A} B_j M^{s_A}$ $(j = 1, \ldots, k)$, then needs $2k$ matrix multiplications. Thus, Alice costs $2 \log s_A + 2n + 2k$ matrix multiplications and n hash evaluations in total.
- Bob computes $B_j = H^{s_B} X_{i_j} M^{s_B}$ $(j = 1, \ldots, k)$ and needs $2 \log s_B + 2k$ matrix multiplications. After that, he computes $A_{i_j} = H^{-s_B} K_j M^{-s_B}$ $(j = 1, \ldots, k)$, then needs $2k$ matrix multiplications and 2 matrix inverse. Thus, Bob costs $2 \log s_B + 4k$ matrix multiplications, 2 matrix inverse and k hash evaluations in total.

The proposal requires $2 \log s_A + 2 \log s_B + 2n + 6k$ matrix multiplications, 2 matrix inverse and $n + k$ hash evaluations. Furthermore, the number of matrix multiplications is bounded by $4\lambda + 2n + 2k$ for $s_A, s_B \in \{0, 1\}^{\lambda}$.

In Protocol 2a, Alice receives 2200 bits from Bob in Pass 1 while Bob receives $2200n + \lambda n$ bits from Alice in Pass 2. Then the communication cost of $\lambda n + 2200n + 2200$ bits is required. For computational cost, Alice needs $2n \log s_A + 3n$ matrix multiplications and n hash evaluations in total. While Bob needs $2 \log s_B + 3$ matrix multiplications, 2 matrix inverse and 1 hash evaluations in total. The number of matrix multiplications of Protocol 2a is bounded by $2n\lambda + 2\lambda + 3n + 3$. (Protocol 1 can be discussed like the simplified version-Protocol 2a). For the communication cost of Protocol 2b, Bob delivers $2200k$ bits to Alice in Pass 1 and then Alice delivers $(2200n + \lambda n)k$ bits to Bob in Pass 2. Thus, Protocol 2b needs the communication cost of $(\lambda n + 2200n + 2200)k$ bits in total. For Protocol 2b computational cost, Alice needs $2n \log s_A + 2kn$ matrix multiplications and n hash evaluations in total. While Bob requires $2k \log s_B + 3k$ matrix multiplications, $2k$ matrix inverse and k hash evaluations in total. The number of matrix multiplications is bounded by $2n\lambda + 2k\lambda + 2kn + 3k$.

Table 1. Comparison of communication cost

Proposal	Protocol 2a	Protocol 2b	Protocol 3
Communication cost (bit)	$\lambda n + 2200n + 2200$	$(\lambda n + 2200n + 2200)k$	$\lambda n + 4400k$

Table 2. Comparison of computational cost

	Matrix multiplication	Matrix inverse
Protocol 2a	$2n\lambda + 2\lambda + 3n + 3$	2
Protocol 2b	$2n\lambda + 2k\lambda + 2kn + 3k$	$2k$
Protocol 3	$4\lambda + 2n + 2k$	2

5 Conclusions

In this paper, three OT protocols are proposed based on matrices over group rings. The feature of our proposals lies in that its security is based on the defined intractability assumptions of group factorization problem of matrices over the group rings, as well as the non-commutativity in this algebraic structure. Considering the absence of quantum algorithm for the GFP problem, we hope the new protocols have the potential to resist quantum attack algorithms.

References

1. Camenisch J, Neven G, Shelat A.: Simulatable adaptive oblivious transfer. EUROCRYPT, pp. 573-590, 2007.
2. Cao Z.J., Liu L.H.: Improvement of One Adaptive Oblivious Transfer Scheme. IACR Cryptology ePrint Archive 2013. URL: http://eprint.iacr.org/2013/517.
3. Childs A.M. and Ivanyos G.: Quantum computation of discrete logarithms in semigroups. CoRR, Vol. abs/1310.6238, 2013.
4. Green, M. Hohenberger, S.: Practical adaptive oblivious transfer from simple assumptions. In: Yuval, I., (Ed.) TCC 2011. LNCS, 6597, Springer, Heidelberg, pp. 347-363, 2011.
5. Grohmann B.: A new protocol for 1-2 oblivious transfer. CoRR, Vol. abs/0904.2023, 2009. http://eprint. iacr. org/2009/172.
6. Gu L.Z. Wang L.C. Ota K. Dong M.X. Cao Z.F. Yang Y.X.: New public key cryptosystems based on non-abelian factorization problems. Security and Communication Networks, 6(7): pp. 912-922, 2013.
7. Habeeb M. Kahrobaei D. Koupparis C. and Shpilrain V.: Public key exchange using semidirect product of (semi) groups. LNCS, 7954, pp. 475-486, 2013.
8. Huang H.F. Chang C.C.: A new t-out-of-n oblivious transfer with low bandwith. Applied Mathematical Sciences, pp. 311-320, 2007.
9. Kahrobaei D. Koupparis C. Shpilrain V.: Public key exchange using matrices over group rings. Groups-Complexity-Cryptology, 5, pp. 97-115, 2013.
10. Lempken W., Magliveras S.S., van Trung T., Wei W.: A public key cryptosystem based on non-abelian finite groups. Journal of Cryptology: the journal of the International Association for Cryptologic Research, 22(1), pp. 62-74, January 2009.
11. Lipmaa H.: New Communication-Efficient Oblivious Transfer Protocols Based on Pairings. IACR Cryptology ePrint Archieve 2007. Vol. 2007, p. 133, 2007. URL: http://eprint.iacr.org/2007/133.
12. Monico C and Neusel M. D.: Cryptanalysis of a system using matrices over group rings. Groups Complexity Cryptology, 7(2), pp. 175-182, 2015.
13. Myasnikov A.D., Ushakov A.: Quantum algorithm for the discrete logarithm problem for matrices over group rings. Groups Complexity Cryptology, pp. 31-36, 2014.
14. Myasnikov A.G., Shpilrain V., Ushakov A., Mosina N. Non-commutative cryptography and complexity of group-theoretic problems. vol. 177. American Mathematical Society Providence, RI, USA; 2011.
15. Parakh A.: Oblivious transfer using elliptic curves. CRYPTOLOGIA: Cryptologia 31(2), pp. 125-132, 2007.
16. Rabin, M. How to exchange secrets by oblivious transfer. Technical Report TR-81, Aiken Computation Laboratory, Harvard University 1981.
17. Shor P.W.: Polynomial-time algorithms for prime factorization and discrete logarithme on a quantum computer. SIAM Journal on Computing, pp. 1484-1509, 1997.
18. Yang YG., Sun SJ., Pan QX., Xu P.: Reductions between private information retrieval and blivious transfer at quautam level. Optik, 126(21), pp. 3206-3209, 2015.

E-Voting Scheme Using Secret Sharing and K-Anonymity

Quanyu Zhao[1] and Yining Liu[2]

Abstract An e-voting scheme is proposed, in which the voter's ballot is shared among all the candidates in voting phase. All candidates and voting system participate in recovering and tallying the ballot, and voting system publishes the ballot on the bulletin board in the post-voting phase. Moreover, the proposed scheme satisfies the coercion-resistant and unconditional security. In addition, other properties of electronic voting are satisfied, such as non-cheating, universal verifiability, confidentiality, anonymous.

1 Introduction

Voting schemes were introduced by both [1] and [2] independently, the traditional paper voting scheme using the paper and ballot boxes cannot be trusted to guarantee all these properties, such as, (1) the paper ballots spend much money since it cannot be reused and a lot of workers must be hired; (2) the manipulation or destruction of votes during tallying are possible; (3) inconvenient, for instance, if election day is not a public holiday and the voting booth is far away from peoples workplaces, its cumbersome to vote for many people [3-4].

In 1981, Chaum proposed the first e-voting scheme [5] to overcome the shortcomings of the traditional voting, moreover, the participants can't track the voters from the result. Due to the properties of security and fairness, e-voting scheme has attracted an increasing interest in recent years [6-12]. Its main properties include: (1) non-cheating, i.e. the vote counting center and the voter cannot cheat; (2) confidentiality, i.e. malicious listener can not get the content of ballot before it is published; (3) anonymous, i.e. a voter can not be linked with his ballot; (4) verifiability, i.e. every voter can verify whether his ballot is counted or not; (5) coercion-resistant, i.e. the voter can not prove the content of his ballot to others, so the malicious voter cannot sell his vote.

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6_91

In this paper, an unconditional secure e-voting scheme based on secret sharing and k-anonymity is proposed. On the one hand, the aspiration of voters is tallied and published on the bulletin aboard, on the other hand, each voter can verify whether his ballot is counted without knowing any others' information.

The rest of this paper is organized as follows. In the next section, we introduce some preliminaries, and in Section 3 the system model is presented. An electronic voting protocol is proposed in Section 4, and Section 5 analyzes the proposed voting scheme. Conclusion is included in Section 6.

2 Preliminaries

The following cryptographic concepts are necessary for understanding the proposed scheme:

2.1 Shamir's (k,n) secret sharing scheme

There are n shareholders $\{P_1,P_2,\cdots,P_n\}$ and a mutually trusted dealer D. D picks $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_{k-1}x^{k-1} mod p$, where $s = f(0) = a_0$, p is a prime, and computes n shares $y_i = f(x_i) mod p (i = 1,2,\cdots,n)$, then distributes y_i to P_i via a secure channel. No less than k shareholders can recover the secret s by using Lagrange interpolating formula,

$$s = f(0) = \sum_{i=1}^{k} f(x_i) \prod_{v=1,v\neq l}^{k} \frac{-x_v}{x_l - x_v} mod p \tag{1}$$

2.2 Secret sharing homomorphism

Secret sharing homomorphism was introduced by Benaloh [13], and it means that the sum of the sub secret can be obtained in this way, the sum of the shares from different sub-secret are added up and interpolated according to the threshold mentioned. Assuming there are two secrets: K_1, K_2, and they are shared by two polynomials $f(x)$ and $g(x)$, respectively. Then, the share $f(i)+g(i)(1 \leq i \leq n)$ can be regarded as the shares corresponding to the secret K_1+K_2. The secret shares are sent to n shareholders and any k of them can retrieve the result back together.

2.3 k-anonymity

k-anonymity means that any element included in a set appears with the probability no greater than $\frac{1}{k}$. For example, $T(A_1,A_2,\cdots,A_n)$ is a table with n attributes $A_1,A_2,\cdots,A_n$. If each sequence of values in a set of attributes appears with at least k occurrences [14], T is defined k-anonymity. Moreover, the values related to multi-attributes in the k-anonymity are generalized. The expected records cannot be determined since k records are in accordance with some information.

In this voting scheme, any randomly k voters' data are generated and published on the bulletin board by *VS*. Nobody including candidates and voters can efficiently distinguish the personally data [15]. Then this scheme satisfies k-anonymity.

3 The system model

The main participants in the e-voting scheme include an authority center (*AC*), voting system (*VS*), voter(*V*), candidate (*C*) and a bulletin board. The participants functions are described as follows:

AC*:AC* launches the e-voting scheme, authorizes the legal voter to elect candidate only once. Moreover, *AC* is responsible for arbitrating disputes and issuing digital certificates to each participant, including public keys, election dates, and verification information.

***VS*:** *VS* is semi-honest to other participants and owns the computing and communication capacity, in the other words, *VS* executes the scheme, but does not attempt to derive or tamper extra information about other voters' private inputs. In addition, *VS* generates the private credential for *V*, but it leaks nothing about the voters' intention. In the post-voting phase, $k(k \leq n)$ ballots are selected randomly to reconstruct the polynomial, then, *VS* aggregates k voters' ballot randomly.

***V*:** *V* is certified by *AC*, selects the favorite candidate with the help of *VS*, then he gets his credential. Usually, *V* is assumed not honest, but *V* cannot prove the content of his election to others for they cannot obtain and infer any information from his credential. In addition, the contents of voters' ballots are mutually independent and different from any others' ballot due to the random number r_i .

C*:C* collaborates to tally the result when *VS* wants to recover the polynomial.

bulletin board*:VS* tallies k voters' ballot and publishes it on the bulletin board once.

It augments the trust and reliability to the currently e-voting scheme. In the currently e-voting scheme, if the voter casts a candidate, the candidates' data base entry is updated and it can easily be tracked by the malicious adversary. However, it is difficult to track the voter's data in the proposed scheme since the voters' shares are tallied by *VS*. On-line verification of the authenticity of the voter is not taken into consideration.

During the secret recovering, inside adversaries (also called "cheaters") can deceive the honest shareholders by altering the shares. In this way, malicious candidates can change the shares and disrupted the elections. Many research papers [16-18] were proposed to address the problems of cheater detection and identification. In this paper, we do not consider this situation that some shareholders change the shares.

Dividing the large number of voters into different optimal groups can satisfy the voter's different privacy requirements. Zhang's research [19] tells us how to optimize the efficiency of the secure ballots aggregation process in a scenario where the total number of voters is large. In this way, *VS* can divide the voters into groups according to voters privacy requirements before running the protocol within each voter group. With the optimal grouping solution, the efficiency of the entire voters' ballot aggregation process can be optimal. Dividing the voter into different groups isn't taken into consideration, we use k-anonymity in this paper.

4 An electronic voting protocol

In this section, we introduce a novel unconditional secure e-voting scheme based on the secret sharing and k-anonymity. Our scheme including *Pre-voting Phase*, *Voting Phase* and *Post-voting Phase*, is described as follows.

Pre-voting Phase

The voting takes place in a polling station and a polling station may have many voting terminals. The list of all candidates and their symbols are showed on the display of the voting machine. The voters and candidates are allowed to take part in election after they have been verified by the *AC*. Assume that there are n voters $V_i(i = 1,2,\cdots,n,n \leq N)$ and m candidates $C_i(i = 1,2,\cdots,m,m < n)$, the candidates can be or not be the voters.

Voting Phase

In this phase, *VS* generates an interpolation polynomial $f(x)$ of degree m: $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m modp$. Each voter's ballot is set as the secret data and distributed into $m+2$ shares, m shares are assigned to m candidates and one share is allotted to *VS*. Moreover, the private credentials are sent to the voters, which is used to verify whether the ballot is counted or not. Notation $R = SN_{prefix}$ denotes the vote's private credential $\{SN_{prefix}|r_i, (x_i,y_i)\}$, where r_i is a random number and (x_i,y_i) is the ultimate share. In addition, the private credential can't be taken out of the polling station. At the end of the voting, *VS* and all candidates can reconstruct a polynomial of degree m and aggregate k ballots together. Fig.1 shows the operation procedures.

The voting procedure is described as follows.

Step1. After V is certified by *AC*, *VS* distributes a temporary *ID* to V , where *VS* does not know any information about the current voter, and *AC* does not know voters' *ID* distributed by *VS*.

Step2. According to V_i's intention, *VS* generates an interpolation polynomial of degree m , which is $f_i(x) = a_{i,0} + a_{i,1}x + a_{i,2}x^2 + \cdots + a_{i,m}x^m modp$. If V_i chooses the

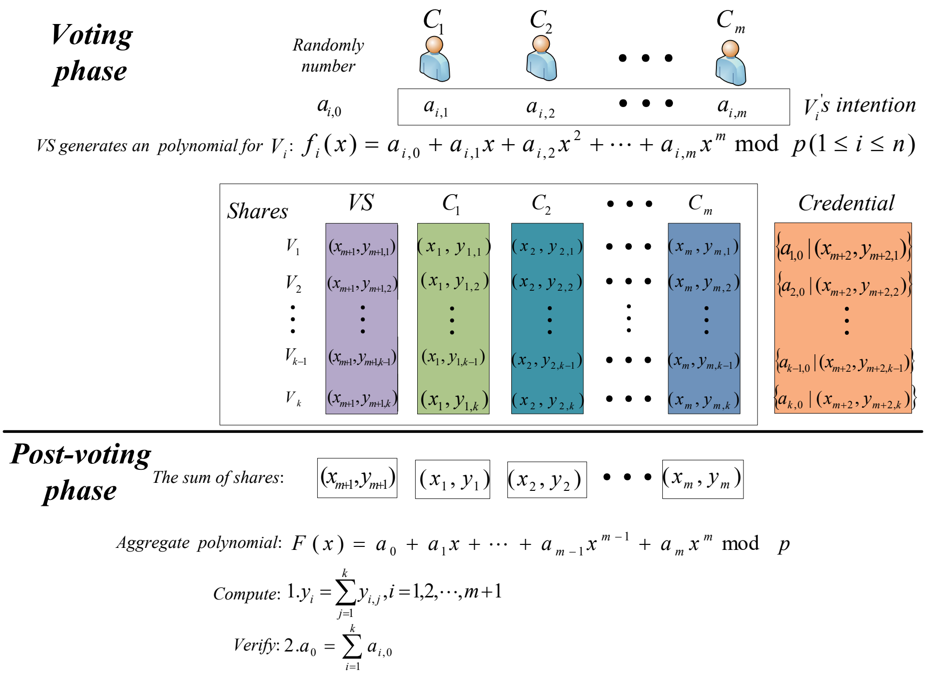

Voting phase
Randomly number
C_1 C_2 C_m
$a_{i,0}$
$a_{i,1}$ $a_{i,2}$ $a_{i,m}$
V_i's intention
VS generates an polynomial for V_i: $f_i(x) = a_{i,0} + a_{i,1}x + a_{i,2}x^2 + \cdots + a_{i,m}x^m \bmod p(1 \le i \le n)$
Shares VS C_1 C_2 C_m Credential
V_1 $(x_{m+1}, y_{m+1,1})$ $(x_1, y_{1,1})$ $(x_2, y_{2,1})$ $(x_m, y_{m,1})$ $\{a_{1,0} \mid (x_{m+2}, y_{m+2,1})\}$
V_2 $(x_{m+1}, y_{m+1,2})$ $(x_1, y_{1,2})$ $(x_2, y_{2,2})$ $(x_m, y_{m,2})$ $\{a_{2,0} \mid (x_{m+2}, y_{m+2,2})\}$
V_{k-1} $(x_{m+1}, y_{m+1,k-1})$ $(x_1, y_{1,k-1})$ $(x_2, y_{2,k-1})$ $(x_m, y_{m,k-1})$ $\{a_{k-1,0} \mid (x_{m+2}, y_{m+2,k-1})\}$
V_k $(x_{m+1}, y_{m+1,k})$ $(x_1, y_{1,k})$ $(x_2, y_{2,k})$ $(x_m, y_{m,k})$ $\{a_{k,0} \mid (x_{m+2}, y_{m+2,k})\}$
Post-voting phase
The sum of shares: (x_{m+1}, y_{m+1}) (x_1, y_1) (x_2, y_2) (x_m, y_m)
Aggregate polynomial: $F(x) = a_0 + a_1x + \cdots + a_{m-1}x^{m-1} + a_mx^m \bmod p$
Compute: $1.y_i = \sum_{j=1}^{k} y_{i,j}, i = 1,2,\cdots,m+1$
Verify: $2.a_0 = \sum_{i=1}^{k} a_{i,0}$

Fig. 1 Operation procedures of the voting phase

C_l, $a_{i,l} = 1(l = 1,2,\cdots,m)$, otherwise $a_{i,l} = 0$. Moreover, $a_{i,0} = 1$ is selected by *VS* randomly with non-zero value, and i is the serial number of ballot;

Step3. *VS* computes $m+2$ shares: $y_{i,r} = f_i(x_r)(r = 1,2,\cdots,m+1,m+2)$, where x_r is the public identifying with the corresponding $C_l(l = 1,2,\cdots,m)$, *VS* and V_i. Then *VS* distributes $(x_r, y_{i,r})(r = 1,2,\cdots,m)$ to all C_l via a secure channel. Finally, V_i gets a credential $R_i = SN_{prefix} = \{a_{i,0}, (x_{m+2}, y_{m+2})\}$ from *VS* in order to verify whether his ballot was counted or not after voting.

Post-voting Phase

Step1. When the results need to be published, *VS* divides the voter randomly into some groups with k voters and publishes the selective V_i'$(i = 1,2,\cdots,k)$ temporary *ID*. Then, *VS* asks the selective k voters to publish the $a_{i,0}(i = 1,2,\cdots,k)$ on the bulletin board.

Step2. Each $C_l(l = 1,2,\cdots,m)$ and *VS* compute the sum of their shares $(x_r, y_{i,r})(r = 1,2,\cdots,m,m+1)$ and publish it. By using the sum of the shares, $C_l(l = 1,2,\cdots,m)$ and *VS* can recover an interpolation polynomial, $F(x) = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m modp$, where $a_j = \sum_{i=1}^{k} a_{i,j}(j = 0,1,2,\cdots,m)$. Thereafter, *VS* publishes the aggregative ballots of the k voters $\{a_0, a_1, \cdots, a_m\}$ on the bulletin board.

Step3. If the sum of the published $a_{i,0}(i = 1,2,\cdots,k)$ on the bulletin board isn't equal to a_0. *VS* and all candidates would be asked to check out their declaratory information, and reconstruct the corresponding ballots again. If V_i's claim is true, the ballots with prefix R_i will be published again, or else they will remain unchanged.

Step4. Finally, according to the published information on the bulletin board, everyone computes the tallying result of C_l,$vote_l = \sum a_l$.

5 Security Analysis

The proposed scheme not only satisfies the correctness, unconditional security, anonymous, confidentiality, and non-cheating, but also achieves additional two security features: universal verifiability and coercion-resistant.

1. ***Correctness*****:** Every voter ensures that the published information is correct by checking $a_0 = \sum_{i=1}^{k} a_{i,0}$. For knowing the content of his vote and randomly number from the credential, voter can restructure the polynomial easily. Thereafter, the voter knows the result is correct by checking whether the polynomial passes through the shares on the credential. Then the scheme satisfies correctness.
2. ***Anonymity*****:** Actually VS and AC do not know any information about the current voter since V can get a temporary ID distributed by VS after he was certified by AC. All the candidates know nothing about the voter. At last, the result is published in this way that some ballots are aggregated and posted. Hence, our scheme satisfies the anonymity.
3. ***Confidentiality*****:** VS sends the shares to m candidates and VS, hereafter, the content of the ballots is confidential, and the ballots are also confidential after publishing due to the aggregate ballots are posted on the bulletin board together. Then our scheme satisfies the confidential.
4. ***Unconditional security*****:**In the proposed scheme, the shares are sent to the candidates and VS. Although the malicious adversary has unconfined computing power, he can't infer some information about the content of the vote. Then, our scheme satisfies the unconditional security.
5. ***Non-cheating*****:** If there are some voters who found that $a_0 = \sum_{i=1}^{k} a_{i,0}$ isn't correctness, the information about these ballots will be checked, reconstructed and published again. Then the proposed scheme meets the non-cheating.
6. ***Universal verifiability*****:**Each voter knows that his ballot is counted by checking that a_0 is equal to the sum of $a_{i,0}(i = 1, 2, \cdots, k)$. Voter uses his intention and randomly number to restructures the polynomial. Thereafter, the voter can verify his intention is counted by judging the polynomial passes through the shares on the credential. Therefore, the scheme is satisfied the universal verifiability.
7. ***Coercion-resistant*****:**The credential included $R_i = SN_{prefix} = \{a_{i,0}, (x_{m+2}, y_{m+2})\}$ contains no information about the content of the vote. The voter can proof the content of his ballot to others for they cannot infer whether the voter casts him or not. Then, the electronic voting scheme satisfies the coercion-resistant.

6 Conclusion

An electronic voting scheme based on secret sharing and k-anonymity was proposed. In the scheme, the voting system generates an interpolation polynomial according to the aspiration of the voters, computes and sends the shares to all candidates and *VS*. They can reconstruct the polynomial and aggregate the ballot together. Moreover, the proposed scheme satisfies the correctness, unconditional security, non-cheating, universal verifiability, confidentiality, anonymous. At last, the voter can proof the content of his credential to others and for they cannot infer any information about the content of ballot.

Acknowledgements This work was partly supported by National Natural Science Foundation of China (Nos. 61363069, 61301166,61662016), Innovation Project of Guangxi Graduate Education (No. YCSZ2015149), and Innovation Project of GUET Graduate Education (No. 2016YJCX44).

7 References

1. FujiwaraT.(2015). Voting technology, political responsiveness, and infant health: evidence from Brazil. *Econometrica*, vol. 83, no. 2, pp. 423-464.
2. Aggarwal,R., Saffi,P.A.C., Sturgess, J.(2015). The role of institutional investors in voting: evidence from the securities lending market. *The Journal of Finance*,70(5), pp. 2309-2346.
3. Liaw,H.T.(2014). A secure electronic voting protocol for general elections. *Computers and Security*, 23(2), pp. 107-119.
4. Chang,C.C., Lee, J.S.(2006). An anonymous voting mechanism based on the key exchange protocol. *Computers and Security*, 25(4), pp. 307-314.
5. Chaum,D.L.(1981). Untraceable electronic mail, return addresses and digital pseudonyms. *Communications of the ACM*, 24(2), pp. 84-88.
6. Cortier,V.,Eigner,F., Kremer,S., et al.(2015). Type-based verification of electronic voting protocols. *Principles of Security and Trust and Springer Berlin Heidelberg*, pp. 303-323.
7. Grewal,G.S., Ryan, M.D., Chen, L., et al.(2015). Du-vote: remote electronic voting with untrusted computers. *2015 IEEE 28th Computer Security Foundations Symposium*, pp.155-169.
8. Ryan,P.Y.A., Schneider,S., Teague, V. (2015). End-to-end verifiability in voting systems, from theory to practice. *IEEE Security and Privacy*, 13(3), pp. 59-62.
9. Cubric,M., Jefferies,A.(2015). The benefits and challenges of large-scale deployment of electronic voting systems: university student views from across different subject groups. *Computers and Education*, 87, pp. 98-111.
10. Chun,T.L., Min,S.H., Chi,Y.L.(2008). An electronic voting protocol with deniable authentication for mobile ad hoc networks. *Computer Communications*, 31, pp. 2534-2540.

11. Fan,C.I., Sun, W.Z.(2008). An efficient multi-receipt mechanism for un-coercible anonymous electronic voting. *Mathematical and Computer Modelling*, 48, pp. 1661-1627.
12. Francesc,S., Josep, M., Miret, J.P., Jordi,P.(2010). Simple and efficient hash-based verifiable mixing for remote electronic voting. *Computer Communication*, 33, pp. 667-675.
13. Benelux,J.C.(1986). Secret sharing homomorphism: keeping shares of a secret secret. *Conference on the Theory and Application of Cryptographic Techniques, Springer Berlin Heidelberg*, pp. 251-260.
14. Ciriani,V., Vimercati, S.D.C.D., Foresti, S., Samarati, P.(2007). K-anonymity. secure data management in decentralized systems, *Springer US*, 33, pp. 323-353.
15. Zhang,Y., Chen, Q., Zhong, S. (2016). Privacy-preserving data aggregation in mobile phone sensing. *IEEE Transactions on Information Forensics and Security*, 11(5), pp. 980-992.
16. Xu,R., Morozov, K., Takagi, T. (2013). On cheater identifiable secret sharing schemes secure against rushing adversary. *International Workshop on Security. Springer Berlin Heidelberg*, pp. 258-271.
17. LinP.Y.2016. Distributed secret sharing approach with cheater prevention based on QR code. *IEEE Transactions on Industrial Informatics*, 12(1), pp. 384-392.
18. Chen,Z., Li, S., Zhu, Y., et al.(2015). A cheater identifiable multi-secret sharing scheme based on the Chinese remainder theorem. *Security and Communication Networks*, 8(18), pp. 3592-3601.
19. Zhang,Y., Chen, Q.J., Zhong, S.(2016). Privacy-preserving data aggregation in mobile phone sensing. *IEEE Transactions on Information Forensics and Security*, 11(5), pp. 980-992.

Author Index

L. Barolli et al. (eds.), *Advances on Broad-Band Wireless Computing, Communication and Applications*, Lecture Notes on Data Engineering and Communications Technologies 2, DOI 10.1007/978-3-319-49106-6