

James J. (Jong Hyuk) Park

Han-Chieh Chao

Hamid Arabnia

Neil Y. Yen

Editors

Advanced Multimedia and Ubiquitous Engineering

Future Information Technology Volume 2



Springer

Lecture Notes in Electrical Engineering

Volume 354

Board of Series editors

Leopoldo Angrisani, Napoli, Italy
Marco Arteaga, Coyoacán, México
Samarjit Chakraborty, München, Germany
Jiming Chen, Hangzhou, P.R. China
Tan Kay Chen, Singapore, Singapore
Rüdiger Dillmann, Karlsruhe, Germany
Haibin Duan, Beijing, China
Gianluigi Ferrari, Parma, Italy
Manuel Ferre, Madrid, Spain
Sandra Hirche, München, Germany
Faryar Jabbari, Irvine, USA
Janusz Kacprzyk, Warsaw, Poland
Alaa Khamis, New Cairo City, Egypt
Torsten Kroeger, Stanford, USA
Tan Cher Ming, Singapore, Singapore
Wolfgang Minker, Ulm, Germany
Pradeep Misra, Dayton, USA
Sebastian Möller, Berlin, Germany
Subhas Mukhopadhyay, Palmerston, New Zealand
Cun-Zheng Ning, Tempe, USA
Toyoaki Nishida, Sakyo-ku, Japan
Bijaya Ketan Panigrahi, New Delhi, India
Federica Pascucci, Roma, Italy
Tariq Samad, Minneapolis, USA
Gan Woon Seng, Nanyang Avenue, Singapore
Germano Veiga, Porto, Portugal
Haitao Wu, Beijing, China
Junjie James Zhang, Charlotte, USA

About this Series

“Lecture Notes in Electrical Engineering (LNEE)” is a book series which reports the latest research and developments in Electrical Engineering, namely:

- Communication, Networks, and Information Theory
- Computer Engineering
- Signal, Image, Speech and Information Processing
- Circuits and Systems
- Bioengineering

LNEE publishes authored monographs and contributed volumes which present cutting edge research information as well as new perspectives on classical fields, while maintaining Springer’s high standards of academic excellence. Also considered for publication are lecture materials, proceedings, and other related materials of exceptionally high quality and interest. The subject matter should be original and timely, reporting the latest research and developments in all areas of electrical engineering.

The audience for the books in LNEE consists of advanced level students, researchers, and industry professionals working at the forefront of their fields. Much like Springer’s other Lecture Notes series, LNEE will be distributed through Springer’s print and electronic publishing channels.

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

James J. (Jong Hyuk) Park
Han-Chieh Chao · Hamid Arabnia
Neil Y. Yen
Editors

Advanced Multimedia and Ubiquitous Engineering

Future Information Technology Volume 2



Springer

Editors

James J. (Jong Hyuk) Park
Seoul National University of Science
and Technology
Seoul
Korea, Republic of (South Korea)

Han-Chieh Chao
National Ilan University
Yilan City
Taiwan

Hamid Arabnia
The University of Georgia
Athens, GA
USA

Neil Y. Yen
The University of Aizu
Aizuwakamatsu
Japan

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-3-662-47894-3 ISBN 978-3-662-47895-0 (eBook)
DOI 10.1007/978-3-662-47895-0

Library of Congress Control Number: 2015940892

Springer Heidelberg New York Dordrecht London
© Springer-Verlag Berlin Heidelberg 2016

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer-Verlag GmbH Berlin Heidelberg is part of Springer Science+Business Media
(www.springer.com)

Message from the FutureTech 2015 General Chairs

FutureTech 2015 is the 10th event of the series of international scientific conference. This conference takes place on May 18–20, 2015, in Hanoi, Vietnam. The aim of the FutureTech 2015 is to provide an international forum for scientific research in the technologies and application of information technology. FutureTech 2015 is the next edition of FutureTech 2014 (Zhangjiajie, China), FutureTech 2013 (Gwangju, Korea), FutureTech 2012 (Vancouver, Canada), FutureTech 2011 (Loutraki, Greece), and FutureTech 2010 (Busan, Korea, May 2010) which was the next event in a series of highly successful the International Symposium on Ubiquitous Applications and Security Services (UASS-09, USA, January 2009), previously held as UASS-08 (Okinawa, Japan, March 2008), UASS-07 (Kuala Lumpur, Malaysia, August 2007), and UASS-06 (Glasgow, Scotland, UK, May 2006).

The conference papers included in the proceedings cover the following topics: Hybrid Information Technology High-Performance Computing, Cloud and Cluster Computing, Ubiquitous Networks and Wireless Communications Digital Convergence, Multimedia Convergence, Intelligent and Pervasive Applications, Security and Trust Computing, IT Management and Service Bioinformatics and Bio-Inspired Computing, Database and Data Mining, Knowledge System and Intelligent Agent, Game and Graphics Human-centric Computing and Social Networks, Advanced Mechanical Engineering, Computer-Aided Machine Design, Control and Automations, and Simulation. Accepted and presented papers highlight new trends and challenges of future information technologies. We hope readers will find these results useful and inspiring for their future research.

We would like to express our sincere thanks to Steering Chair: James J. Park (SeoulTech, Korea) and Hamid R. Arabnia (The University of Georgia, USA). Our special thanks go to the Program Chairs: Joon-Min Gil (Catholic University of Daegu, Korea), Neil Y. Yen (The University of Aizu, Japan), and Muhammad Khurram Khan (King Saud University, Saudi Arabia); all program committee members; and all reviewers for their valuable efforts in the review process that helped us to guarantee the highest quality of the selected papers for the conference.

We cordially thank all the authors for their valuable contributions and the other participants of this conference. The conference would not have been possible without their support. Thanks are also due to the many experts who contributed to making the event a success.

C.S. Raghavendra, University of Southern California, USA
Jason C. Hung, Oversea Chinese University, Taiwan
Doo-soon Park, SoonChunHyang University, Korea
Jianhua Ma, Hosei University, Japan
FutureTech 2015 General Chairs

Message from the FutureTech 2015 Program Chairs

Welcome to the 10th International Conference on Future Information Technology (FutureTech 2015), which will be held in Hanoi, Vietnam, on May 18–20, 2015. FutureTech 2015 will be the most comprehensive conference focused on the various aspects of information technologies. It will provide an opportunity for academic and industry professionals to discuss recent progress in the area of future information technologies. In addition, the conference will publish high-quality papers which are closely related to the various theories and practical applications in multimedia and ubiquitous engineering. Furthermore, we expect that the conference and its publications will be a trigger for further related research and technology improvements in these important subjects.

For FutureTech 2015, we received many paper submissions, and after a rigorous peer review process, we accepted only articles with high quality for the FutureTech 2015 proceedings, published by the Springer. All submitted papers have undergone blind reviews by at least two reviewers from the technical program committee, which consists of leading researchers around the globe. Without their hard work, achieving such a high-quality proceeding would not have been possible. We take this opportunity to thank them for their great support and cooperation. We would like to sincerely thank the following invited speakers who kindly accepted our invitations and, in this way, helped to meet the objectives of the conference: Prof. Han-Chieh Chao, National Ilan University, Taiwan, and Prof. Timothy K. Shih, National Central University, Taiwan. Finally, we would like to thank all of you for your participation in our conference and also thank all the authors, reviewers, and organizing committee members. Thank you and enjoy the conference!

Joon-Min Gil, Catholic University of Daegu, Korea
Neil Y. Yen, The University of Aizu, Japan
Muhammad Khurram Khan, King Saud University, Saudi Arabia
FutureTech 2015 Program Chairs

Organization

Steering Chair	James J. Park, SeoulTech, Korea Hamid R. Arabnia, The University of Georgia, USA
General Chairs	C.S. Raghavendra, University of Southern California, USA Jason C. Hung, Oversea Chinese University, Taiwan Doo-soon Park, SoonChunHyang University, Korea
General Vice-Chairs	Jianhua Ma, Hosei University, Japan Hwa-Young Jeong, Kyung Hee University, Korea Cho-Li Wang, University of Hong Kong, Hong Kong
Program Chairs	Joon-Min Gil, Catholic University of Daegu, Korea Neil Y. Yen, The University of Aizu, Japan Muhammad Khurram Khan, King Saud University, Saudi Arabia
Workshop Chairs	Deok-Gyu Lee, Seowon University, Korea Vincent Huang, National Taichung University of Science and Technology, Taiwan Ka Lok Man, Xi'an Jiaotong-Liverpool University, China
International Advisory Committee	Yi Pan, Georgia State University, USA Qun Jin, Waseda, Japan Fatos Xhafa, Technical University of Catalonia, Spain

	Hsiao-Hwa Chen, National Cheng Kung University, Taiwan
	Ivan Stojmenovic, University of Ottawa, Canada
	Laurence T. Yang, St Francis Xavier University, Canada
	Young-Sik Jeong, Dongguk University, Korea
Publicity Chairs	Eunyoung Lee, Dongduk Women's University, Korea
	Byung-Gyu Kim, Sun Moon University, Korea
	Sung-Ki Kim, Sun Moon University, Korea
Program Committee	Alfredo Cuzzocrea, University of Calabria, Italy
	Amagasa Toshiyuki, University of Tsukuba, Japan
	Byna Suren, Lawrence Berkeley National Laboratory, USA
	Caldelli Roberto, Universita degli Studi di Firenze
	Cerquitelli Tania, Politecnico di Torino, Canada
	Cha Yue-Shan, National Taipei University, Taiwan
	Chen Bing, Memorial University of Newfoundland, Canada
	Chen WeiFeng, California University of Pennsylvania, USA
	Chi-Fu Huang, National Chung Cheng University, Taiwan
	Ching-Hsien Hsu, Chung Hua University, China
	Davidovic Tatjana, Mathematical Institute of the Serbian Academy of Sciences and Arts, Serbia
	Edward Hua, QED Systems, USA
	Gonzalez JoseAntonio, Universidad de Malaga, Spain
	Homenda Wadysaw, Instytut Badan Systemowych Polskiej Akademii Nauk, Poland
	Hu Yu-Chen, Providence University, Taiwan

Ivetic Dragan, University of Novi Sad, Serbia
Jiqiang Lu, Institute for Infocomm Research, Singapore
Kapetanios Epaminondas, University of Westminster, UK
Klyuev Vitaly, University of Aizu, Japan
Kyungbaek Kim, Chonnam National University, Korea
Lu Leng, Southwest Jiaotong University Emei Campus, China
Maumita Bhattacharya, Charles Sturt University, Australia
Pai-Ling Chang, Shih-Hsin University, Taiwan
Picard Willy, Poznan University of Economics, Poland
Qiang He, Swinburne University of Technology, Australia
Raylin Tso, National Chengchi University, China
Ren-Song Ko, National Chung Cheng University, Korea
Rios Ruben, Universidad de Malaga, Spain
Rudolf Tsoy, Far Eastern State Academy, Russia
Salem Abdelbadeeh, Ain Shams University, Egypt
Song Fu, University of North Texas, USA
Wei-Chuen Yau, Multimedia University, Malaysia
Wookey Lee, Inha University, Korea
Wyne Mudasser, National University, USA
Xiao Liu, East China Normal University, China
Yeun Chan, Yeob Khalifa University of Science, Technology and Research, UAE
Yo-Ping Huang, National Taipei University of Technology, Taiwan
Zhang Yunquan, State Key Lab of Computer Science, China
Zhang Zhiqiang, Harbin Engineering University, China

Message from the MUE 2015 General Chairs

MUE 2015 is the 9th event of the series of international scientific conference. This conference takes place on May 18–20, 2015, in Hanoi, Vietnam. The aim of the MUE 2015 is to provide an international forum for scientific research in the technologies and application of multimedia and ubiquitous engineering. Ever since its inception, International Conference on Multimedia and Ubiquitous Engineering has been successfully held as MUE-14 (Zhangjiajie, China, May 2014), MUE-13 (Seoul, Korea, May 2013), MUE-12 (Madrid, Spain, July 2012), MUE-11 (Loutraki, Greece, June 2011), MUE-10 (Cebu, Philippines, August 2010), MUE-09 (Qingdao, China, June 2009), MUE-08 (Busan, Korea, April 2008), and MUE-07 (Seoul, Korea, April 2007).

The conference papers included in the proceedings cover the following topics: Multimedia Modeling and Processing, Ubiquitous and Pervasive Computing, Ubiquitous Networks and Mobile Communications, Intelligent Computing, Multimedia and Ubiquitous Computing Security, Multimedia and Ubiquitous Services, Multimedia Entertainment, and IT and Multimedia Applications. Accepted and presented papers highlight new trends and challenges of multimedia and ubiquitous engineering. We hope readers will find these results useful and inspiring for their future research.

We would like to express our sincere thanks to Steering Chair: James J. (Jong Hyuk) Park (SeoulTech, Korea). Our special thanks go to the Program Chairs: Gangman Yi (Gangneung-Wonju National University, Korea) and Chengcui Zhang (The University of Alabama at Birmingham, USA); all program committee members; and all reviewers for their valuable efforts in the review process that helped us to guarantee the highest quality of the selected papers for the conference.

Shu-Ching Chen, Florida International University, USA
Young-Sik Jeong, Dongguk University, Korea
Han-Chieh, Chao National Ilan University, Taiwan
MUE 2015 General Chairs

Message from the MUE 2015 Program Chairs

Welcome to the 9th International Conference on Multimedia and Ubiquitous Engineering (MUE 2015), which will be held in Hanoi, Vietnam, on May 18–20, 2015. MUE 2015 will be the most comprehensive conference focused on the various aspects of multimedia and ubiquitous engineering. It will provide an opportunity for academic and industry professionals to discuss recent progress in the area of multimedia and ubiquitous environment. In addition, the conference will publish high-quality papers which are closely related to the various theories and practical applications in multimedia and ubiquitous engineering. Furthermore, we expect that the conference and its publications will be a trigger for further related research and technology improvements in these important subjects.

For MUE 2015, we received many paper submissions, and after a rigorous peer review process, we accepted only articles with high quality for the MUE 2015 proceedings, published by the Springer. All submitted papers have undergone blind reviews by at least two reviewers from the technical program committee, which consists of leading researchers around the globe. Without their hard work, achieving such a high-quality proceeding would not have been possible. We take this opportunity to thank them for their great support and cooperation. Finally, we would like to thank all of you for your participation in our conference and also thank all the authors, reviewers, and organizing committee members. Thank you and enjoy the conference!

Gangman Yi, Gangneung-Wonju National University, Korea
Chengcui Zhang, University of Alabama at Birmingham, USA
MUE 2015 Program Chairs

Organization

Steering Chair	James J. Park, SeoulTech, Korea
General Chairs	Shu-Ching Chen, Florida International University, USA
General Vice-Chairs	Young-Sik Jeong, Dongguk University, Korea Han-Chieh, Chao National Ilan University, Taiwan Weijia Jia, City U. of Hong Kong, Hong Kong
Program Chairs	Hwa-Young Jeong, Kyung Hee University, Korea Gangman Yi, Gangneung-Wonju National University, Korea
Workshop Chairs	Chengcui Zhang, University of Alabama at Birmingham, USA Namje Park, Jeju National University, Korea Xu Shuo, ISTIC, China
International Advisory Committee	Borko Furht, Florida Atlantic University, USA Thomas Plagemann, University of Oslo, Norway Roger Zimmermann, National University of Singapore, Singapore Hamid R. Arabnia, The University of Georgia, USA Stephan Olariu, Old Dominion University, USA Albert Zomaya, University of Sydney, Australia Yi Pan, Georgia State University USA Koji Nakano, University of Hiroshima, Japan Cheonshik Kim, Anyang University, Korea (Leading Chair)
Publicity Chairs	Mohamed Gaber, University of Portsmouth, UK Ryan Leong Hou U, Universidade De Macau, China Chengjiu Yin, Kyushu University, Japan Jian-Lian Chen, Aletheia University, Taiwan Junbo Wang, University of Aizu, Japan

Program Committee

Nan-Chen Hsieh, National Taipei University and Health Sciences, Taiwan
Afrand Agah, West Chester University of Pennsylvania, USA
Akihiro Sugimoto, National Institute of Informatics, Japan
Angel D. Sappa, Universitat Autonoma de Barcelona, Spain
Bin Lu, West Chester University, USA
Ch. Z. Patrikakis, Technological Education Institute of Pir, Greece
Chao-Tung Yang, Tunghai University, Taiwan
Dakshina Ranjan Kisku, Asansol Engineering College, India
Dalton Lin, National Taipei University, Taiwan
Debzani Deb, Winston-Salem State University, USA
Dongkyun Kim, KISTI, Korea
Ezendu Ariwa, University of Bedfordshire, UK
Guillermo Camara Chavez, Universidade Federal de Minas Gerais, Brasil
Hai Jin, Huazhong University of Sci & Tech, China
HaRim Jung, Sungkyunkwan University, Korea
HeonChang Yu, Korea University, Korea
Hermann Hellwagner, Klagenfurt University, Austria
Jin Kwak, Ajou University, Korea
Jungong Han, Civolution Technology, Netherlands
Jun-Won Ho, Seoul Women's University, Korea
Kilhung Lee, Seoul National University of Science, Korea
Kwang Sik Chung, Korea National Open University, Korea
Marco Cremonini, University of Milan, Italy
Mario Doeller, University of applied science, Germany
Paisarn Munesawang, Naresuan University, Thailand
Pascal Lorenz, University of Haute Alsace, France
Rainer Unland, University of Duisburg-Essen, Germany
Reinhard Klette, The University of Auckland, New Zealand

Se-Hak Chun, Seoul National University of Science,
Korea
Seunghae Kim, KISTI, Korea
Sokratis Katsikas, University of Piraeus, Greece
Teng Li, Baidu Inc., China
Wesley De Neve, Ghent University, Belgium
Yan Liu, Hong Kong Polytechnic University, China
Young-Gab Kim, Sejong University, Korea
Zheng-Jun Zha, National University of Singapore,
Singapore

Contents

The Study on the Detection of the Damaged File Using the Graph of the Information Entropy for File Trust Management	1
Chae Ho Cho, SungSuk Kim, Seonmi Han and Kwang Sik Chung	
Enhancing Dataset Processing in Hadoop YARN Performance for Big Data Applications	9
Ahmed Abdulhakim Al-Absi, Dae-Ki Kang and Myong-Jong Kim	
A Multimetric Approach for Discriminating Distributed Denial of Service Attacks from Flash Crowds	17
Mourad Elhadef	
A Supporting Tool for Spiral Model of Cryptographic Protocol Design with Reasoning-Based Formal Analysis	25
Kazunori Wagatsuma, Tsubasa Harada, Shogo Anze, Yuichi Goto and Jingde Cheng	
Idea of Personal Digital Memories Using Smart Application	33
Martin Zmitko and Ondrej Krejcar	
Proxy Based Mobility Management Scheme Using Prediction Algorithm	41
Daewon Lee, Daeyong Jung, Doo-Soon Park and HwaMin Lee	
Principles of Usability in Human-Computer Interaction	51
Tomas Hustak and Ondrej Krejcar	
Design and Performance Evaluation of a VANET-Based Adaptive Overtaking Assistance System	59
Ihn-Han Bae and Jae-Kon Lee	

Vulnerability Analysis on Smartphone Fingerprint Templates	71
Young-Hoo Jo, Sung-Yun Jeon, Jong-Hyuk Im and Mun-Kyu Lee	
Study of Measures for Detecting Abnormal Access by Establishing the Context Data-Based Security Policy in the BYOD Environment.	79
Changmin Jo	
Incremental Multilevel Association Rule Mining of a Dynamic Database Under a Change of a Minimum Support Threshold	87
Nophadon Pumjun and Worapoj Kreesuradej	
Compressing Method of NetCDF Files Containing Clustered Sparse Matrix	95
Suntae Hwang, Gyuyeun Choi and Daeyoung Heo	
An Approach to Discovering Weighted Service Pattern	107
Jeong Hee Hwang and Mi Sug Gu	
Cyber Security Modeling for the Operation of Virtualized Trusted Networks	115
Yong-Hee Jeon	
Development of a Quantitative Evaluation Method for Vehicle Control Systems Based on Road Information	123
Jinyong Kim, Changhyun Jeong, Dohyun Jung and Byeongwoo Kim	
Design Challenges and Implementation of a Shipborne Gateway for Safe and Secure Navigational Networks	129
Kwangil Lee and Moonsub Song	
A Design and Implementation of Lightweight ENC for Android App	137
Moonsub Song, Kwangil Lee, Byungtae Jang and Soonghwan Ro	
Study of Censorship in Named Data Networking	145
Xingmin Cui, Lucas C.K. Hui, S.M. Yiu and Yu Hin Tsang	
Research on Design of LTE-Based High-Speed Railway Networks in Korea: Current and Emerging Issues	153
Jin-Kyu Choi, Hanbyeog Cho, Hyun-Seo Oh, Kyong-Ho Kim and Heung-Gyo Ryu	

Contents	xxiii
Alternative Way to Manage the Research Documents	161
Jeong Ah Kim, Jae-Young Choi, Jong-Won Ko, Sun-Tae Kim and Young-Hwa Cho	
R&D Project Management Using Context in Document: Research Descriptor	169
Jong-Won Ko, Jae-Young Choi, Jung-Ah Kim, Sun-Tae Kim and Young-Hwa Cho	
Research Descriptor Based Project Management Approach for R&D Projects	177
Jong-Won Ko, Jae-Young Choi, Jung-Ah Kim, Sun-Tae Kim and Young-Hwa Cho	
Attributes for Characterizing Java Methods	185
Illo Lee, Suntae Kim, Sooyong Park and Younghwa Cho	
A Study on Real Time Circular Motion in Robots Using Kalman Filters	193
Malrey Lee, Suntae Kim and Younghwa Cho	
A Study on Traceability Between Documents of a Software R&D Project	203
Suntae Kim, HyunYoung Kim, Jeong Ah Kim and Younghwa Cho	
Path Planning for Avoiding Obstacles for Unmanned Ground Vehicles	211
Gyoungeun Kim, Deok Gyu Lee and Byeongwoo Kim	
Improved AEB Performance at Intersections with Diverse Road Surface Conditions Based on V2V Communication	219
Sangduck Jeon, Deok Gyu Lee and Byeongwoo Kim	
Integrated Information Retrieval for Distributed Heterogeneous Ontology Systems	225
Sang-Won Hwang, Young-Kwang Nam, Lee-Nam Kwon, Jae-Soo Kim and Byoung-Dai Lee	
hFractal: A Cloud-Assisted Simulator of Virtual Plants for Digital Agriculture	233
Fei Hao, Doo-Soon Park, Young-Sik Jeong and Jong Hyuk Park	

Proposal and Validation of AEB System Algorithm for Various Slope Environments	241
Ming Lin, Jaewoo Yoon and Byeongwoo Kim	
A Study on the V2V-Communication-Based AEB System for High-Speed Driving Under Various Road Surface Conditions	247
Hyeonggeun Mun and Byeongwoo Kim	
Implementation of Power off Recovery Scheme for Block Mapping FTL	253
Ji-Hwan Chung and Tae-Sun Chung	
PLC Monitoring and Protection for SCADA Framework	259
Ahmed AlShemeili, Chan Yeob Yeun and Joonsang Baek	
Modeling of Smart Supply Chain for Sustainability	269
Kyeongrim Ahn, Sangwon Lim and Younggyo Lee	
Human Action Classification Using Multidimensional Functional Data Analysis Method	279
Wanhyun Cho, Sangkyoon Kim and Soonyoung Park	
Light-Weight Authentication Scheme for NFC mCoupon Service in IoT Environments	285
Sung-Wook Park and Im-Yeong Lee	
Model of CPU-Intensive Applications in Cloud Computing	301
Junjie Peng, Yongchuan Dai, Yi Rao and Xiaofei Zhi	
A Study of Effects of UTAUT-Based Factors on Acceptance of Smart Health Care Services	317
Yoo-Jin Moon and Young-Ho Hwang	
Detection and Recognition of Road Markings for Advanced Driver Assistance System	325
JongBae Kim	
Color and Depth Image Correspondence for Kinect v2	333
Changhee Kim, Seokmin Yun, Seung-Won Jung and Chee Sun Won	
Equivalent Test Model of Wireless Optical Communication System for Automotive Environment	341
Sang Yub Lee, Jae Kyu Lee, Duck Keun Park and Jae Jin Ko	

Feasibility Study of Non-linear Apodization for IVUS B-mode Imaging	349
Jin Ho Sung, Seon Mi Ji, Chan Yuk Park, Sung Yun Park, Sung Min Kim, Won Seuk Jang, Byeong Cheol Choi and Jong Seob Jeong	
An Empirical Study of Impacts of User Intention for Smart Wearable Devices and Use Behavior	357
Yoo-Jin Moon, Young-Ho Hwang and Sungkap Cho	
Lightweight Context-Aware Activity Recognition	367
Byung Gill Go, Asad Masood Khattak, Babar Shah and Adil Mehmood Khan	
Efficient Sliding Window Join in Data Stream Processing	375
Hyeon Gyu Kim	
The Effect of the User Interface Design of Smartphone Applications on Users' Individual Experiences of Performance	383
Wonjin Jung	
Data Hiding for H.264/AVC Based on the Motion Vector of 16 Grids	389
Cheng-Hsing Yang, Yih-Kai Lin, Chun-Hao Chang and Jin-Yi Chen	
Idle-Time Processing in Time-Slide Window Join	397
Hyeon Gyu Kim	
A Large-Scale Object-Based Active Storage Platform for Data Analytics in the Internet of Things	405
Quanqing Xu, Khin Mi Mi Aung, Yongqing Zhu and Khai Leong Yong	
Concurrent Regeneration Code with Local Reconstruction in Distributed Storage Systems	415
Quanqing Xu, Weiya Xi, Khai Leong Yong and Chao Jin	
A Technique for Streaming Multiple Video Parts in Parallel Based on Dash.js	423
Konthorn Sangkul, Sucha Smanchat and Jo Yew Tham	
Prioritized Medical Image Forwarding Over DTN in a Volcano Disaster	431
Muhammad Ashar, Morihiko Tamai, Yutaka Arakawa and Keiichi Yasumoto	

A Framework of Personal Data Analytics for Well-Being Oriented Life Support	443
Seiji Kasuya, Xiaokang Zhou, Shoji Nishimura and Qun Jin	
Secure Initial Attach Based on Challenge-Response Method in LTE Conforming to LTE Standards	451
Jungho Kang	
A Design and Development of Secure-Coding Check System Based on CVE and CWE	457
Hyungjoo Kim and Moon-seog Jun	
Lightweight Encryption Technique for Content Security in Open IPTV Environment	465
Jaewoo Kim, Younggu Lee and Moonseog Jun	
Lightweight Mutual Authentication Routing Protocols Design Based on LEACH in Sensor Network Environment	473
Jaeseung Lee, Jae-pyo Park, Eunhwan Kim and Moon-seog Jun	
Method Research for Safe Authentication in Cloud Environment	479
Junho Song, Jaesoo Kim, ManSik Kim and Moon-seog Jun	
A Design of Dual Encryption Based on Data Obfuscation and Mutual Authentication in the Smart-Grid Environment	487
Wonkyu Choi, Jungoh Park, Jaesik Lee and Moon-seog Jun	
A Study on Realtime Detecting Smishing on Cloud Computing Environments	495
Ayoung Lee, Kyounghun Kim, Heeman Lee and Moonseog Jun	
Application of Data Mining for Crime Analysis	503
Aziz Nasridinov, Jeong-Yong Byun, Namkyoung Um and HyunSoon Shin	
Graph-Based Motor Primitive Generation Method of UAVs Based on Demonstration-Based Learning	509
Yunsick Sung, Jeonghoon Kwak and Jonghyuk Park	
Author Index	511

The Study on the Detection of the Damaged File Using the Graph of the Information Entropy for File Trust Management

Chae Ho Cho, Sungsuk Kim, Seonmi Han and Kwang Sik Chung

Abstract Information entropy refers to the complexity of information included in set of data in a mathematical way. Entropy is now usually used for the classification of files or detection and analysis of malicious code. Information entropy graph shows the probability of occurrence of each information included in set of data using information entropy. Each Well Known File has different entropy and each file can be sorted using this. When it comes to binary file, however, different files can have the same entropy values so there is error possibility. Thus, the identification of files for the least errors can be possible when using entropy and graph patters. In the forensic analysis process, detections of hidden and tampered files are handled. With existing forensic method, the extensions of header and footer of tampered files are not automatically detected. When the other functions such as calculation and comparison of graphs are added, accuracy of experiment is increased in the forensic process. In this study, we proved that different files but have the same entropy values are assorted with the information entropy graphs. The information entropy graphs of Well Known Files showed the meaningful patterns for analysis and detection. When it comes to the damaged file header, footer, and even body, they sustained the same graph patterns even though they showed different entropy values.

C.H. Cho

Korea Cyber Forensics Professional Association, Seoul, Republic of Korea

e-mail: greatopen@gmail.com

S. Kim

Department of Computer Science, Seokyeong University, Seoul,
Republic of Korea

e-mail: sskim03@skuniv.ac.kr

S. Han

Department of Nano Convergence Engineering, Seokyeong University, Seoul,
Republic of Korea

e-mail: gkstjsal83@gmail.com

K.S. Chung (✉)

Department of Computer Science, Korea National Open University, Seoul,
Republic of Korea

e-mail: kchung0825@knou.ac.kr

Keywords Forensic · Information entropy · Information entropy graph · Augmented reality · File steganography · Forgery of files

1 Introduction

Entropy refers to the physical concept that indicates the measure of disorder in certain system in thermodynamics. Claude E. Shannon, the mathematician first suggested Information Entropy which means the amount of information or signals related to the events using entropy [1]. Given situation, information entropy is characterized that the more different information mixed with, the higher entropy it has while the more same information mixed with, the lower entropy it has. That is, the characteristics of information can be measured by entropy value and as uncertain and disordered information is increased, entropy is also increased. Forensic method using information entropy is broadly used for the malicious code detection or analysis of compressed file.

Damaged files in Meta area has to be recovered with using self-information of the files not using the information of file system. One of the carving technologies using the characteristics of file itself is the way to use information entropy. The files which have various extensions have different header and footer information for each and then store the information through certain forms of files. According to each kind of file, information entropy value varies and even if header or footer is damaged, information entropy value is rarely changed. Although the file is deleted and Meta information area is damaged, deleted or hidden files can be restored with information entropy value in data area. When the entropy values of different forms of files are similar, however, those files can be recognized as the same file.

In this study, beyond the existing research which only uses the information entropy value, detection method of damaged files using the information entropy graph is suggested. The detection of files utilizing the information entropy graph can be indicated that different files with similar entropy values can have different entropy graphs. Especially, when the different files have the same entropy values because of damaged header or footer information of files, the information entropy graph shows their differences for sure.

In Chapter “[Enhancing Dataset Processing in Hadoop YARN Performance for Big Data Applications](#)”, we will discuss the concept of the information entropy suggested by Claude E. Shannon and header and footer information of Well Known files. Also, existing study for file detection and restoration using information entropy will be dealt with. In Chapter “[A Multimetric Approach for Discriminating Distributed Denial of Service Attacks from Flash Crowds](#)”, the method for damaged file detection with information entropy and entropy graphs will be suggested. It proved that the files are not able to detected with existing forensic technology can be identified as files before damaged with the information entropy and entropy graphs. In Chapter “[A Supporting Tool for Spiral Model of Cryptographic Protocol Design](#)

with Reasoning-Based Formal Analysis”, the further development and study for the forensic technology using the information entropy and graphs will be addressed.

2 Related Research

The information entropy $H(X)$ that Claude E. Shannon suggested can be digitized as following formula [1].

$$\begin{aligned} H(X) &= \sum_{i=1}^n p(x_i)I(x_i) \\ &= \sum_{i=1}^n p(x_i) \log_b \frac{1}{p(x_i)} \\ &= - \sum_{i=1}^n p(x_i) \log_b p(x_i) \end{aligned}$$

The information entropy is the total value of information probability that each information can have in a group. In the information entropy H , the certain information refers to x_i and its probability is $p(x_i)$. Also its amount of information refers to $I(x_i)$. Shannon used \log to get $I(x_i)$. The entropy of the x_i would be gotten through logarithmic calculations with multiplying p_i , all possible probability value by its reciprocal. The base number of \log ‘b’ is generally 2 and Euler’s constant ‘e’ or 10 would be used either. The higher the entropy is, the more the amount of information is and vice versa.

The various Well Know Files have their own header and footer information. In general computer system, even if operating system, possessor or the size of files are changed, its header and footer information never change. In cyber forensic filed, changed or corrupted files are identified with these characteristics of files.

In Study [2], the information entropy of Well Know File was measured and verified that the entropy values of execution file or compressed execution file are higher than general text or image files. After measuring the entropy of each file, even though the different files have the same extensions caused by malicious changes, with the entropy values for each, they can be sorted. In Study [3], the malicious code having the form of compressed execution file showed higher entropy than other general execution files. Using the properties that show different information entropies according to the tools used for compressing the execution files, tools used in the process of malicious code can be detected and analyzed.

In Study [4], the way to use the information entropy for the restoration of broken files was suggested. The information entropy was used for the restoration of broken files in more than 300 hard drives. JPEG files, Microsoft OLE files, and compressed file format ZIP files were selected and then, each entropy value of each defragmented cluster was measured. After measuring the entropies of adjacent cluster, clusters having similar entropies were gathered together then, reconstructed.

Except that, there is another way to compare the patterns of files by using the restoration technology that data area is interpreted with stream units. This method, however, has high probability of error when it has the similar patterns and takes long time for the restoration. The Carving technology using Slack Space is RAM Slack. When the files saved with cluster units don't use the sound cluster, the final sector area would be filled with 0×00 . If each cluster of sector detects the sector filled with 0×00 , header and footer of file can be automatically detected [4]. In the case of text file, there are two ways; calculating the frequency that character string composed of ASCII Code appears and the confirmation of line feed character. Line feed character is generally composed of '\n' which means line feed and '\r' which means carriage return. This method, however, has restriction for use because it is confined to text file and has different values of line feed characters in Windows or UNIX operation system.

The existing forensic tools don't have functions for the automatic detection of files which have been tampered because of the change of the extensions or signatures. In this case, suspicious files should be opened respectively and then be analyzed the structures with manual sorting. With this reason, the camouflaged files having changed names, extensions, headers and footers should be evaluated only by the forensic analyst. Even the way to use the script function offered by forensic tool has the probability of error so new detection method would be inevitable.

3 File Detection Using the Information Entropy Graph

In this study, the information entropy graph is suggested for the detection of damaged file. During the forensic analysis process, the information entropy graph can help the accuracy of damaged or hidden file's detection increase.

This study suggests the information entropy values and its graphs as a method for the detection of damaged files. In this study, the problem has been solved that the existing forensic tools are not able to verify which file is fake or not because header or footer of file has been damaged. It has an advantage that files don't need to be executed or read to verify that the files are modulated or not in forensic process, but only need patterns of graphs to identify the types of files. The information entropy values and its graphs of Well Known File are compared in this. The files used in this experiment have different resolutions and were selected randomly.

3.1 doc File

doc file is the file extension of MS-Word developed by MS-Office, Microsoft. After comparing the information entropy graphs, confirmed that $p(x_0)$ is higher than other $p(x_i)$. The graphs are in Table 1.

Table 1 Comparison of the information entropy in docx file

File	Information entropy	Information entropy graph	Note
1.docx	7.91		Original file
2.docx	7.94		Change in header, footer and extensions

Table 2 Comparison of the information entropy in the damaged jpg file

File	Information entropy	Information entropy graph	Note
1.jpg	7.76		Original file
2.jpg	7.76		Change in header, footer and extensions

3.2 Damaged jpg File

The detection and analysis of the damaged file is an important element in whole forensic process. In this test, the differences between when the header or footer and extensions were changed and when the part of body information was changed were compared. As a result, when the header, footer, and extension were changed, the information entropy value showed the same and there was no visible change in the information entropy graph. Also, when the part of body information was changed, the information entropy value showed meaningful changes but only minimal change of pattern was found in the graph. The entropy graph of damaged files is seen in Table 2.

	A	B	C	D	E	F
	freqList[i]	Original File	20% Damaged File	30% Damaged File	Standard Deviation of 20% Damaged File	Standard Deviation of 30% Damaged File
1						
2	0	0.024384781	0.035997715	0.039489602	0.476236965	0.619436432
3	1	0.004747106	0.006412434	0.006992432	0.350809209	0.472988375
4	2	0.004407331	0.006117025	0.007015156	0.387920452	0.591701449
5	3	0.003643379	0.005845422	0.007020566	0.604395604	0.926937927
6	4	0.003866289	0.006148406	0.00739605	0.590260285	0.912958298
7	5	0.004610763	0.006471949	0.007214259	0.403661112	0.564656184
8	6	0.002985472	0.00530438	0.006549859	0.7767307	1.193910837
9	7	0.003846811	0.005842176	0.006882059	0.518706048	0.789029536
10	8	0.003237598	0.004893187	0.005793482	0.511363636	0.789438503
11	9	0.003853304	0.005690684	0.006741388	0.47683235	0.749508565
12	10	0.004347817	0.006120272	0.00694482	0.407665505	0.597312096
13	11	0.003672596	0.004973262	0.005660386	0.35415439	0.541249263
14	12	0.003852222	0.00530979	0.005943892	0.378370787	0.542977528
15	13	0.0033945	0.00457289	0.005243783	0.347146956	0.544788014
16	14	0.00308286	0.004326175	0.005074978	0.403299403	0.646191646
17	15	0.004555577	0.005570572	0.006042361	0.22280285	0.326365796
18	16	0.00352435	0.004604271	0.00521132	0.306416948	0.478661345
19	17	0.003499462	0.004029684	0.004548002	0.151515152	0.299628942
20	18	0.003871699	0.004715726	0.005175612	0.217998882	0.336780324
21	19	0.003428045	0.004274235	0.004727628	0.246843434	0.379103535

Fig. 1 Dynamic rage of the entropy value for each unit

3.3 Graph Changes in Information Entropy

The information entropy graphs of Well Known Files stayed the same pattern regardless of the degree of damage. In this study, to measure the rate of change of the entropy graphs with pre-damaged and post-damaged files, standard deviation was used. Pre-damaged file $p(x_i) \cdot I(x_i)$ and post-damaged file $p(x_i) \cdot I(x_i)$ were compared each other and their range of fluctuation was examined. Almost no variation when calculating the entropies of neighboring units (Fig. 1).

4 Conclusion

In recent Cyber infringement accidents, deleting or damaging the critical evidences has been increasing. The tampered files can be detected when only using the information of header and footer or extensions of files in existing forensic tools. It has certain limitations to detect and verify files when both extension and header or footer are tampered in forensic process. In this study, the new way to verify the damaged files using the information entropy values and their graphs of ‘well-Known Files’ were suggested. The same kinds of files have the similar patterns of entropy graphs, which can help identify the characteristics of files. Although the entropy values were changed according to the degree of damage, their graphs maintained the same patterns. To sum up, if the entropy graphs is utilized to verify the damaged files, the possibility of errors in forensic analysis process can be decreased.

References

1. Shannon, C.E.: A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423, 623–656 (1948)
2. Lyda Sparta, R., Hamrock, J., McDonald, B.: Using entropy analysis to find encrypted and packed malware. *IEEE Secur. Priv.* 40–45 (2007)
3. Jeong, G., Choo, E., Lee, J., Bat-Erdene, M., Lee, H.: Generic Unpacking using Entropy Analysis, pp. 98–105. IEEE (2010)
4. Garfinkel, S.L.: Carving contiguous and fragmented files with fast object validation. *Digital Invest.* **4S**, S2–S12 (2007)

Enhancing Dataset Processing in Hadoop YARN Performance for Big Data Applications

Ahmed Abdulhakim Al-Absi, Dae-Ki Kang and Myong-Jong Kim

Abstract In Hadoop MapReduce distributed file system, as the input dataset files get loaded and split to every worker, workers start to do the required computation according to user logic. This process is done in parallel using all nodes in the cluster and computes output results. However, the contention of resources between the map and reduce stages cause significant delays in execution time, especially due to the memory IO overheads. This is undesired because the task execution in the Hadoop MapReduce induces an overhead in considering redundant data in case of imprecise applications which increases the execution time. Thus, in this paper we present our approach to optimize local worker memory management mechanism to reduce the presence of null schedule slots. Efficient utilization of slots leads to reduce execution times. The local memory management mechanism adopted enables efficient parallel execution and reduced memory overheads. The approach effectively reduced the MapReduce computation time which minimizes the budget for application execution in the cloud.

Keywords Dataset · Hadoop YARN · MapReduce · Big data · Cloud computing

A.A. Al-Absi (✉) · D.-K. Kang

Division of Computer and Information Engineering, Dongseo University, Busan, Korea
e-mail: absiahmed@gmail.com; ahmed_absi2005@yahoo.com

D.-K. Kang
e-mail: dkkang@dongseo.ac.kr

M.-J. Kim
School of Business, Pusan National University, 63 Beon-gil 2, Busandaehag-ro,
Geumjeong-gu, 609-735 Busan, Korea
e-mail: mjongkim@pusan.ac.kr

1 Introduction

With the increase in population and beneficiary of internet services, data size is getting increased day by day where 100s of quadrillion of data files are there in cloud available in unstructured nature [1]. On the other hand the application of data-insensitive applications does need certain optimum approach to manage these data files and retrieve the data even without mammoth task and complexity. Various applications like IaaS and PaaS do need the applications which can be effective for providing optimum data access in real time operations. Taking into account of these all circumstances, now days, a number of research works being going on, and Hadoop was one of the significant outcomes that is being used extensively for cloud frameworks.

Hadoop [2] is open-source software in the form of a highly scalable and fault tolerant distributed system which plays a very significant role in data storage and its processing. This framework Hadoop encompasses two dominant parts, first Hadoop distributed file system (HDFS) while second refers for MapReduce. HDFS is the mechanism to classify data on nodes or clusters and provide an interface for data management between users, tasks trackers and nodes or machines. The space where there is certain optimization scope is MapReduce.

According to [3] the main issue with MapReduce framework is its batch-processing oriented nature, which means stateless mapper followed by a stateless reducer, that are executed by a batch job scheduler. This paradigm makes repeated querying of datasets difficult and imposes limitations. Moreover, the process of mapping and converting to intermediate combiner is a time consuming and need to be optimized. In cloud context and due to the cloud heterogeneous behavior existing between the central servers and storing disks, there must be something parallel architecture that could enhance the processing speed and data retrieval rate in MapReduce [4].

A number of researches like in [5, 6] have been done for optimization for MapReduce framework, but still there exists a huge scope for further optimization with diverse cloud platform to come up with the optimum cloud computing model. Taking into consideration of these factors, here in this paper, we introduce a parallel execution model based on MapReduce framework. The major contributions of the parallelized model are in the local worker memory management mechanism and the optimization technique adopted to reduce the presence of null schedule slots. The worker nodes i.e. the Map workers and Reduce workers operate based on the slots assigned. Efficient utilization of slots leads to reduced execution times. The local memory management mechanism adopted enables efficient parallel execution and memory overhead reduction. As in Hadoop our approach also considers the data in chunks. The Map Workers in MapReduce framework perform computations on the chunks of data. In our work, these chunks of data are further split to enable parallel execution in the Map Worker nodes.

2 Hadoop Scenario and Proposed Solution

2.1 *Motivation*

In MapReduce programming model, Hadoop runs MapReduce files in the form of (Key, Value). Converting these input text files into form of (Key, value) requires passing the values from the input split to mapper by one more pre-defined interface called Record Reader [7]. The key of the split file is associated with each line by its byte offset in which the Record Reader is invoke repeatedly on the input until the entire InputSplit file completed whereas each invocation of the RecordReader leads to another call to the map() method of the Mapper and store the intermediate result into the combiner [7].

The main issue is that the contention of resources between the map and reduce stages causes significant delays in execution time, especially due to the memory IO overheads. As the Hadoop map stage is initially completed, the reduce task is performed. This kind of a serial execution mode can hinder execution performance. MapReduce Map processes are initiated sequentially as splits the dataset into small chunks. The map and reduce workers perform their tasks in their pre assigned slots. The presence of ideal or unutilized slots affects system performance. The task execution in the Hadoop MapReduce induces an overhead in considering redundant data in case of data-intensive applications which increases the execution time. This is undesired for a number of reasons because it requires more storage and consumes more computation time. Therefore, the next section addresses this issue by a solution to optimize map computation time and improves the storage efficiently.

2.2 *Proposed Solution*

In Hadoop MapReduce, the contention of resources between the map and reduce stages cause significant delays in execution time, especially due to the memory IO overheads. This is undesired because it requires more storage and computation time. Thus, to enhance performance of MapReduce and resource utilization, we present our approach to optimize MapReduce performance by data reduction technique. In our approach, the map and the reduce slots assignments are optimized to minimize the occurrence of null slots or unoccupied slots. The Map Workers receive the chunk data from the cloud storage and store the data in the local cache memory. The received data in MapReduce is further split into parallel data blocks in our approach. Figure 1 illustrates our parallel data model in Hadoop MapReduce.

As is noted, each data block is executed in a parallel fashion. In addition to the parallel execution, redundant and previously computed data is eliminated from the local memory to achieve reduction in execution time and storage overheads. It is observed that the local memory available with the Map Workers reduces in size as the execution time increases. The reduction in the local memory is dependent on the

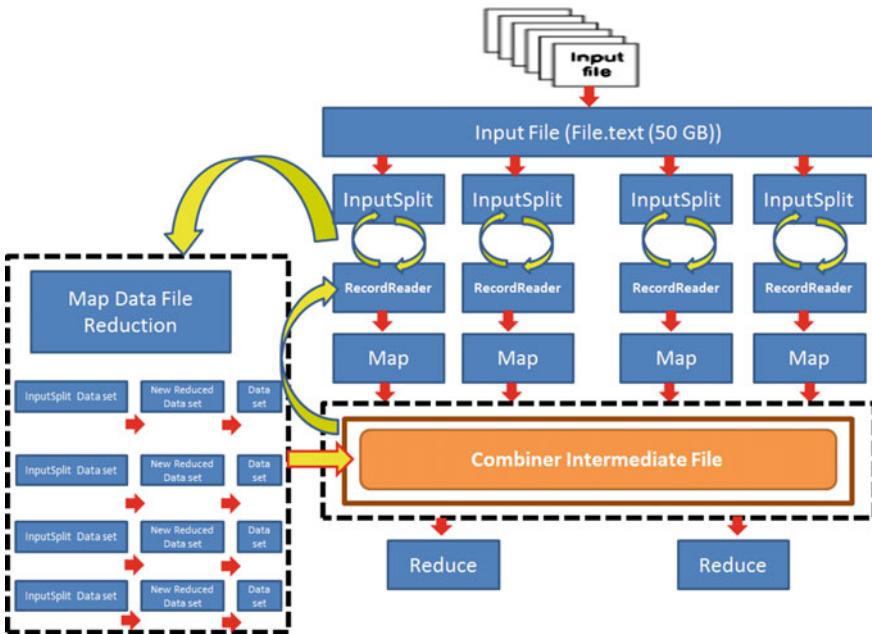


Fig. 1 The proposed data reduction in Hadoop MapReduce

actual computation function. The Map Workers store the intermediate results in the cloud storage. The Reduce workers utilize the intermediate data to perform the reduce tasks. For example, in wordcount, suppose we have a file contains A to Z, if worker-1 has done the computation for “A”, the file in which the next worker will get is the file with “A” already removed. Workers will do their individual work and as soon as one worker finishes its job it goes to the next level in parallel processing. The words are eliminated and stored in central store. Therefore, the loading and computation process will also be expedited in our approach. Figure 2 presents our data reduction in Hadoop MapReduce.

3 Performance Evaluation

3.1 Experiment Setup

In order to implement our system to optimize Hadoop MapReduce performance by data reduction, we considered Microsoft Azure cloud platform for our developed and optimized Hadoop implementation. Here we implement Microsoft Azure HDInsight technology for creation of virtual machines.

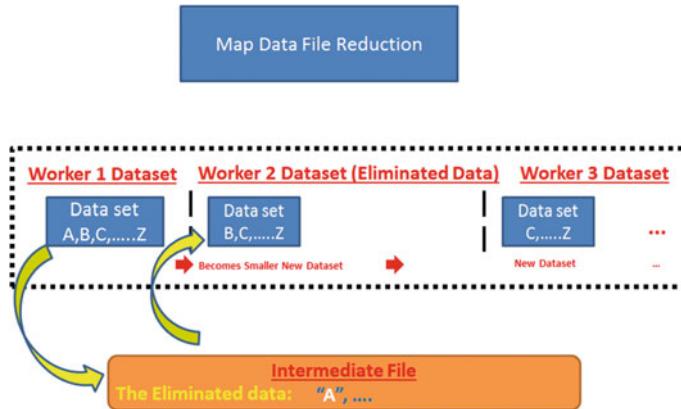


Fig. 2 The proposed data reduction process

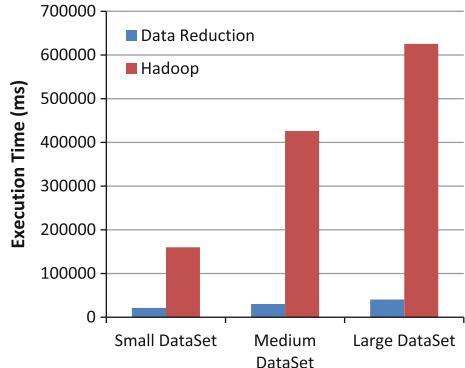
Wikipedia dataset [8] were downloaded as text files for our work. These text files converted into 10 small chunk files, where each file was just over 25 MB. Each Map task processes one of these chunk files. This resulted in a total dataset size of approximately 250, which was copied to the local storage as well as to Azure blob storage service. We used three datasets for Wordcount application benchmark (Small S 100 MB, Medium M 200 MB, Large L 250 MB). The datasets used to test MapReduce loaded input files execution times issue on the wordcount example jobs that come with the Hadoop distribution. We used wordcount application as a benchmark for our experiment and we installed our work on the ongoing development of Apache Hadoop YARN (Yet Another Resource Negotiator) version 2.4.0.2.1.3.0.

3.2 Evaluation Result

This section presents the evaluation results for our work in enhancing MapReduce using the data reduction. The experiment has been executed for the purpose of evaluating the Map phase aiming to observe the Hadoop map phase execution time before and after applying data reduction approach. We sequentially have executed several requests for the word count application and on each request, we have checked the execution time of map's total time spent by all maps in occupied slots in milliseconds (ms). Figure 3 represents the execution time of our work in Microsoft Azure for wordcount with 4 workers when we vary the input dataset size.

From Fig. 3, it is clear that the execution time increases as the size of input data scales. Furthermore, Hadoop acquires poor performance compared to our work with data reduction. It is because MapReduce's performance is affected due to the contention of resources between the map and reduce stages which cause significant

Fig. 3 Workers over all execution time performance comparison of “wordcount” benchmark between Hadoop and data reduction approach in different datasets



delays in execution time, especially due to the memory IO overheads which affect I/O performance and require more storage. Note that our approach has effectively reduced the MapReduce computation time by considering the input data computation in a parallel fashion. The local memory management aids in memory overhead reductions. The optimization strategy to reduce the occurrences of unutilized slots improves the execution efficiency.

Figure 4 reports the wordcount execution with data reduction compared to traditional Hadoop as we scale the number of workers in small dataset. In this graph, the less the execution time the higher the performance.

We can find it clear from Fig. 3 that the reduction of the execution times had a big impact on datasets inputsplit execution performance. The reduction approach was able to successfully complete execution of the Map before the Hadoop with the scalability of workers in different input file size small, medium and large datasets.

From the experiment results, we conclude that the data input loading in MapReduce is an important factor in terms of I/O performance on cloud environment. The contention of resources between the map and reduce stages cause significant delays in execution time, especially due to the memory IO overheads. In

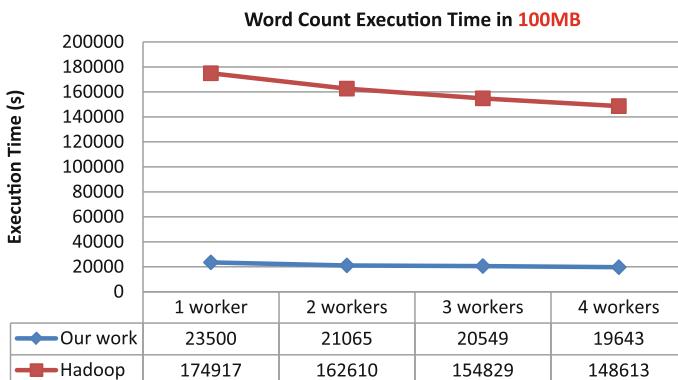


Fig. 4 Scale Wordcount Execution Time for small dataset input file

our approach, The Map Workers receive the chunk data from the cloud storage and store the data in the local cache memory. The received data is further split into parallel data blocks and therefore our work approach outperforms Hadoop by a factor of 7.44 in wordcount application. Our approach has provided faster computation time with less storage, which minimizes the budget for application execution in the cloud.

4 Conclusion and Future Work

In Hadoop based cloud computing systems, the contention of resources between the map and reduce stages cause significant delays in execution time, especially due to the memory IO overheads. This is undesired because the task execution in the Hadoop MapReduce induces an overhead in considering redundant data in case of imprecise applications which increases the execution time. Thus, this paper presented our approach to optimize local worker memory management mechanism to reduce the presence of null schedule slots. The approach effectively reduced the MapReduce computation time which minimizes the budget for application execution in cloud.

For future work, we are planning to analyze and compare the performance of our work in other high performance computing application like gene sequencing, sequence matching, and page ranking.

Acknowledgment This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. NRF-2013R1A1A2013401).

References

1. Gupta, P.K.: Introduction to Analytics and Big Data/Hadoop. Implementing Information Infrastructure Summit (IIIS). Marina Mandarin, Singapore, 30 May 2013. http://issuu.com/fairfaxbm/docs/cws_jul-aug2013/17
2. <http://hadoop.apache.org/>
3. Dean, J., Ghemawat, S.: MapReduce: simplified data processing on large clusters. Commun. ACM **51**(1), 107–113 (2008)
4. Kolb, L., Thor, A., Rahm, E.: Load balancing for mapreduce-based entity resolution. In: 2012 IEEE 28th International Conference on Data Engineering (ICDE), pp. 618–629 (2012)
5. Luo, Y., Guo, Z., Sun, Y., Plale, B., Qiu, J., Li, W.: A hierarchical framework for cross-domain MapReduce execution. In: Proceedings of ECMLS, pp. 15–22 (2011)
6. Zaharia, M., Konwinski, A., Joseph, A. D., Katz, R. H., Stoica, I.: Improving mapreduce performance in heterogeneous environments. In: OSDI. USENIX, pp. 29–42 (2008)
7. <https://developer.yahoo.com/hadoop/tutorial/module4.html>
8. Thottethodi, M., Ahmad, F., Lee, S., Vijaykumar, T.N.: Puma: Purdue mapreduce benchmarks suite. Technical Report, Purdue University (2012)

A Multimetric Approach for Discriminating Distributed Denial of Service Attacks from Flash Crowds

Mourad Elhadef

Abstract Distributed Denial of Service (DDoS) attack, whether at the application or network layer, continues to be a critical threat to the Internet. In a DDoS attack, attackers run a massive number of queries through the victim's search engine or database query to bring the server down. This massive number of queries results in a very high traffic generated within a short period of time. Or in the Internet, researchers have identified a legitimate high traffic, known as a *flash crowd*, where a very large number of users simultaneously access a popular web site, which produces a surge in traffic to the web site and might cause the site to be virtually unreachable. Thus the need to be able to discriminate between DDoS attack traffics and flash crowds. In this project, a hybrid discrimination mechanism is proposed to detect DDoS attacks using various features that characterize the DDoS traffics, and that distinguish it from flash crowds. These features include among others the entropy variation, the information distance, and the correlation coefficient.

Keywords DDoS attacks · Flash crowds · Similarity · Entropy · Information distance · Discrimination

1 Introduction

One of the emerging security threats to the Internet is Distributed Denial-of-Service (DDoS) attacks. In fact, a recent survey [1] of the 70 largest Internet operators in the world has demonstrated that DDoS attacks not only have dramatically increased in

This work is supported by Abu Dhabi University's Faculty Research Incentive Grant.

M. Elhadef (✉)

College of Engineering, Abu Dhabi University, Abu Dhabi, United Arab Emirates
e-mail: mourad.elhadef@adu.ac.ae

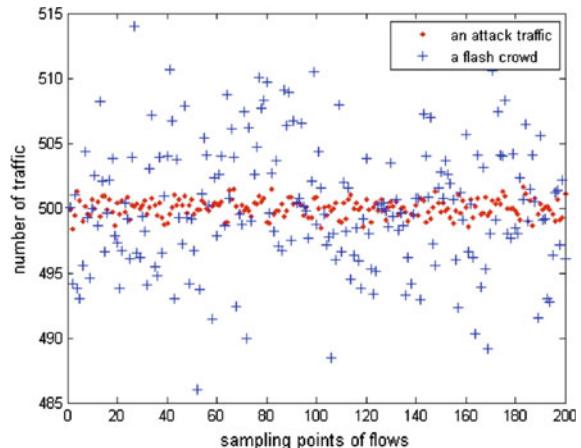
recent years; but, they became stronger and more sophisticated. Hackers are becoming more encouraged to organize botnets to commit these crimes motivated mainly by huge financial rewards, such as renting out their botnets for attacks or collecting sensitive information for malicious purposes [2]. Since it is easy to change the source addresses of IP packets [3], verifying the pattern of attack flows [4], and in addition the memory-less feature of the Internet, it became extremely hard to defend against DDoS attacks.

In this research project, we will investigate one particular challenge in dealing with DDoS attacks: the discrimination of DDoS attacks from fresh cloud traffic, and propose a hybrid solution to deal with this challenge. To date, there is no effective and efficient algorithm available to defend against mimicking traffic patterns of flash crowds, i.e., legitimate dramatic surge of accessing to a service site for special events (such as breaking news), in DDoS attack traffic.

In the last three decades various DDoS detection approaches have been proposed, including, activity profiling [5], sequential change-point detection [6, 7], wavelet analysis [8], chi-square/entropy detector [5, 9], to name a few. Most of these techniques relies on a single feature or fingerprint of specific DDoS attacks. Since hackers can easily mimic these features to fool these detection methods. This includes for example spoofing the source IP addresses using the real Internet IP address distribution to counter the source address distribution based detection [10, 11]; or modifying TTL values according to the real hop distance between bots and the victim in order to foul the hop-count-based detection [12]. In addition, attackers can mimic the behaviors of flash crowds [13], where a sudden increase of legitimate traffic occur. Examples of flash crowds' traffic include for example thousands of fans accessing the official website when an important match is ongoing, or millions of people visiting a popular news channel after a special event had happened.

DDoS attacks and flash crowds share similar behaviors, and we have to differentiate them effectively, otherwise, we may raise false alarms. Figure 1 shows the

Fig. 1 Difference between DDoS traffic and flash crowds traffic



difference between DDoS traffic and flash crowds traffic [14]. We propose to use various properties of traffic such the entropy rate, the information distance, the correlation coefficient, and others to discriminate DDoS attack traffics from flash crowd traffics. We will analyze the efficiency of the new proposed approach in achieving the two goals: raising DDoS alarm as early as possible and discriminating DDoS attacks from flash events, using simulations of DDoS attacks. Discriminating DDoS attacks from flash crowds have been extensively studied in the last decades, [2–13, 15]. All previous research, in particular [13, 16, 17], focused on extracting DDoS features. Using these features they were able to detect and filter DDoS packets. We believe that these methods by themselves cannot detect DDoS attacks. To the best of our knowledge, the most popular defense against flash crowd attacks is the use of graphical puzzles to differentiate between humans and bots [18]. This method involves human responses and can be annoying to users. In [19, 20], Xie and Yu tried to differentiate DDoS attacks from flash crowds at the application layer based on user browsing dynamics. In [21], Oikonomou and Mirkovic tried to differentiate the two by modeling human behavior. These behavior-based discriminating methods work well at the application layer. However, we have not seen any detection method at the network layer, which can extend our defense diameter far from the potential victim.

The objective of this work is to combine some of the existing solutions developed for discriminating DDoS attacks, by combining them into a hybrid approach that will perform the discriminations based on many features. This will help in overcoming the weaknesses of each technique, and hence, form a solution that is more efficient and more effective. The remainder of this paper is organized as follows. First, we present preliminary concepts and some related definitions. Next, we describe, in Sect. 3, the new hybrid approach for discriminating DDoS attacks from flash crowds, followed by an analysis of the proposed discrimination approach in the same section. Finally, we summarize our work and present future directions in Sect. 4.

2 Preliminaries

In this section, we start by presenting preliminary definitions, and then we formalize the features that characterizes Internet traffic. In particular, we define the entropy rate, the information distance, and the correlation coefficient following [9, 14, 22], respectively. Consider the simple community network, shown in Fig. 2, showing two flows of packets, X_i and X_j , from the different edge routers R_1 and R_2 , with the server being the victim which requires protection.

We characterize the network packets sharing the same destination address as one *network flow*. The number of packets is sampled at regular time interval, and hence any network flow X_i will be characterized by a vector comprising N measures of the number of packets at given time interval t : $X_i^t = \{x_i[1], x_i[2], \dots, x_i[N]\}$. The *flow strength* is defined using $E[X_i] = \frac{1}{N} \sum_{n=1}^N x_i[n]$, and the *flow fingerprint* is given by

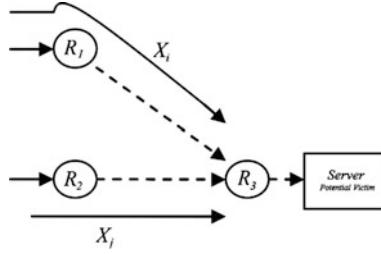


Fig. 2 Two network flows within a simple community network

$X'_i = \left\{ \frac{x_i[1]}{N \cdot E[X_i]}, \frac{x_i[2]}{N \cdot E[X_i]}, \dots, \frac{x_i[N]}{N \cdot E[X_i]} \right\}$. The similarity between two different network flows will be characterized by the correlation between them. Considering that two flows may be correlated with a phase difference, we define the correlation between two network flows X_i and X_j , denoted by $r_{X_i, X_j}[k]$, using: $r_{X_i, X_j}[k] = \frac{1}{N} \sum_{n=1}^N x_i[n] x_j[n+k]$, where $0 \leq k < N$ is the position shift of the X_j flow. The *flow correlation coefficient* measures the similarity between two flows and is defined by:

$$\rho_{X_i, X_j}[k] = \frac{r_{X_i, X_j}[k]}{\frac{1}{N} \sqrt{\left[\sum_{n=1}^{N-1} x_i^2[n] \sum_{n=1}^{N-1} x_j^2[n] \right]}} \quad (1)$$

The combined approach we are proposing uses the information distance as a second metric on which discrimination will be based. According to results, it has been shown that the *Sibson distance* is the best metric for flow distance. We calculate the probability of distribution of the flow X_i using $p(x_i[k]) = x_i[k] \cdot (\sum_{n=1}^N x_i[n])^{-1}$. For any given two network flows, X_i and X_j , with probability distributions $p(X_i)$ and $p(X_j)$, respectively, the *Sibson distance* $D_S(X_i, X_j)$ is defined as follows, where $D(X_i, X_j)$ is the *Kullback-Leibler distance* [23]:

$$D(X_i, X_j) = \sum_{n=1}^N p(x_i[n]) \cdot \log \frac{p(x_i[n])}{q(x_i[n])} \quad (2)$$

$$D_S(X_i, X_j) = \frac{1}{2} \left\{ D\left(X_i, \frac{X_i + X_j}{2}\right) + D\left(X_j, \frac{X_i + X_j}{2}\right) \right\}$$

The third metric on which we will rely to discriminate between DDoS attacks and flash crowds is the *entropy variations*. The entropy variation allows to measure the changes of randomness of flows at a router for a given time interval [24]. For a given time interval ΔT , we define the variation of the number of packets for a given flow as follows: $N_i(t + \Delta T) = X_i^{t+\Delta T} - X_i^t$. Now, based on the large number theorem, the probability of each flow at a local router is calculated using:

$$p(X_i, t + \Delta T) = \frac{N_i(t + \Delta T)}{\sum_{n=1}^{\infty} N_i(t + n \cdot \Delta T)}$$

We define the *entropy* [23] of flows for the local router as follows:

$$H(F) = - \sum_k p(X_k, t + \Delta T) \log p(X_k, t + \Delta T) \quad (3)$$

In this paper, our objective to detect DDoS and discriminate it from flash crowds by measuring at the same time the similarity among the flows using the flow correlation coefficient and the Sibson distance, and the changes in randomness of flows using the entropy variations.

3 Multimetric Discrimination Approach

In this section, we present the design of the multimetric discrimination approach, and present the details about the algorithm implementation. In Fig. 3, we describe the main steps of the discrimination algorithm. The discrimination process is initiated by cooperating routers once there is a surge of network flows that is assign of possible DDoS attack. The cooperating routers, e.g. R_1 and R_2 in Fig. 2, start by sampling the suspicious flows for a given time interval. They will then exchange the metric measurements. Cooperating routers will be able to calculate the similarity degrees and the changes in randomness of the flows using the previously defined metrics. Routers will label a flow as DDoS attack if the distance is smaller than a given threshold α , the correlation coefficient is larger than a second threshold δ , and the entropy variations is less than a given threshold β .

Algorithm *DiscriminateDDoS* **Begin**

1. Let t , N , α , δ , and β denote respectively, the sampling time interval, the number of samples, distance, correlation, and randomness thresholds.
2. Initialize sampling at router R_i and R_j upon detecting a suspicious flow: $X_i^t = \{x_i[1], x_i[2], \dots, x_i[N]\}$ and $X_j^t = \{x_j[1], x_j[2], \dots, x_j[N]\}$.
3. Compute the flow correlation coefficient $\rho_{X_i, X_j}[k]$ using equation (1).
4. Calculate the probability distributions $p(X_i)$ and $p(X_j)$, and the Sibson distance $D_s(X_i, X_j)$ using equation (2).
5. Evaluate changes randomness of flows X_i^t and X_j^t entropy variations following equation (3).
6. Each cooperating router performs Steps 1-5, then it exchanges the computed information with its peers by broadcasting it.
7. **If** $(\rho_{X_i, X_j}[k] > \delta \& D_s(X_i, X_j) < \alpha \& \text{Entropy variation} < \beta)$
Then Discard the suspicious flow
Else Forward packets to final destination.

Fig. 3 The multimetric discrimination algorithm

4 Conclusion

In this paper, we developed a new discrimination algorithm to differentiate DDoS attacks from flash crowds using three combined metrics: the flow correlation coefficient, the Sibson distance, and the entropy variations. Each of these metrics has been used by itself to solve the DDoS discrimination problem. The combined approach is will profit from the advantages of the three metrics, and we believe it will provide a higher accuracy when used combined. In the future, we are planning first to perform an extensive simulation study of the multi-metric detection approach using real datasets. Second, we will investigate other metrics that we believe will improve further the accuracy of the discrimination of DDoS attacks from flash crowds.

References

1. Arbor: IP Flow-Based Technology (2011). <http://www.arbornetworks.com>
2. Stone-Gross, B., Cova, M., Cavallaro, L., Gilbert, B., Szydlowski, M., Kemmerer, R., Kruegel, C., Vigna, G.: Your Botnet is my Botnet: analysis of a Botnet takeover. In: Proceedings of ACM Conference on Computer Communications Security (2009)
3. Peng, T., Leckie, C., Ramamohanarao, K.: Survey of network-based defense mechanisms countering the DoS and DDoS problems. *ACM Comput. Surv.* **39**(1) (2007)
4. Chen, Y., Hwang, K.: Spectral analysis of TCP flows for defense against reduction-of-quality attacks. In: The 2007 IEEE International Conference on Communications (ICC'07), pp. 1203–1210, June 2007
5. Feinstein, L., Schnackenberg, D., Balupari, R., Kindred, D.: Statistical approaches to DDoS attack detection and response. In: Proceedings of DARPA Information Survivability Conference and Exposition, vol. 1, pp. 303–314, 22–24 April 2003. IEEE CS Press (2003)
6. Blazek, R.B., Kim, H., Rozovskii, B., Tartakovsky, A.: A novel approach to detection of ‘Denial-of-Service’ attacks via adaptive sequential and batch-sequential change-point detection methods. In: Proceedings of IEEE Workshop Information Assurance and Security, pp. 220–226, June 2001. IEEE CS Press (2001)
7. Wang, H., Zhang, D., Shin, K.G.: Change-point monitoring for the detection of DoS attacks. *IEEE Trans. Dependable Secure Comput.* **1**(4), 193–208 (2004)
8. Barford, P., Kline, J., Plonka, D., Ron, A.: A signal analysis of network traffic anomalies. In: Proceedings of ACM SIGCOMM internet measurement workshop, pp. 71–82, Nov 2002. ACM Press (2002)
9. Kumar, K., Joshi, R.C., Singh, K.: A distributed approach using entropy to detect DDoS attacks in ISP domain. In: The International Conference on Signal Processing of Communications and Networking (ICSCN'07), pp. 331–337, Feb 2007
10. Duan, Z., Yuan, X., Chandrashekhar, J.: Controlling IP spoofing through interdomain packet filters. *IEEE Trans. Dependable Secure Comput.* **5**(1), 22–36
11. Yi, F., Yu, S., Zhou, W., Hai, J., Bonti, A.: Source-based filtering algorithm against DDOS attacks. *Int. J. Database Theory Appl.* **V1**(1), 9–20 (2008)
12. Wang, H., Jin, C., Shin, K.G.: Defense against spoofed IP traffic using hop-count filtering. *IEEE/ACM Trans. Netw.* **V15**(1), 40–53 (2007)
13. Carl, G., Kesidis, G., Brooks, R.R., Rai, S.: Denial-of-service attack detection techniques. *IEEE Internet Comput.* **10**(1), 82–89 (2006)

14. Yu, S., Zhou, W., Jia, W., Guo, S., Xiang, Y., Tang, F.: Discriminating DDoS attacks from flash crowds using flow correlation coefficient. *IEEE Trans. Parallel Distrib. Syst.* **23**(6) (2012)
15. Zargar, S.T., Joshi, J., Tipper, D.: A survey of defense mechanisms against distributed denial of service (ddos) flooding attacks. *IEEE Commun. Surv. Tutorials* **15**(4), 2046–2069 (2013)
16. Jung, J., Krishnamurthy, B., Rabinovich, M.: Flash crowds and denial of service attacks: characterization and implications for CDNs and websites. In: *Proceedings of 11th International Conference on World Wide Web (WWW)*, pp. 252–262 (2002)
17. Chenand, Y., Hwang, K.: Collaborative detection and filtering of shrew DDoS attacks using spectral analysis. *J. Parallel Distrib. Comput.* **V66**(9), 1137–1151 (2006)
18. Kandula, S., Katabi, D., Jacob, M., Berger, A.: Botz-4-Sale: surviving organized DDoS attacks that mimic flash crowds. In: *Proceedings of Second Symposium on Networked Systems Design and Implementation (NSDI'05)* (2005)
19. Xie, Y., Yu, S.-Z.: A large-scale hidden semi-Markov model for anomaly detection on user browsing behaviors. *IEEE/ACM Trans. Netw.* **V17**(1), 54–65 (2009)
20. Xie, Y., Yu, S.-Z.: Monitoring the application layer DDoS attacks for popular websites. *IEEE/ACM Trans. Netw.* **17**(1), 15–25 (2009)
21. Oikonomou, G., Mirkovic, J.: Modeling human behavior for defense against flash crowd attacks. In: *Proceedings of IEEE International Conference on Communications* (2009)
22. Yu, S., Thapngam, T., Liu, J., Wei, S., Zhou, W.: Discriminating DDoS flows from flash crowds using information distance. In: *Proceedings of Third International Conference on Network and System Security*, pp. 351–356, Washington, DC, USA (2009)
23. Cover, T.M., Thomas, J.A.: *Elements of Information Theory*, 2nd edn. Wiley-Interscience (2006)
24. Shui, Yu., Zhou, W., Doss, R., Jia, W.: Traceback of DDoS attacks using entropy variations. *IEEE Trans. Parallel Distrib. Syst.* **22**(3), 412–425 (2011)

A Supporting Tool for Spiral Model of Cryptographic Protocol Design with Reasoning-Based Formal Analysis

Kazunori Wagatsuma, Tsubasa Harada, Shogo Anze, Yuichi Goto and Jingde Cheng

Abstract Many cryptographic protocols proposed to securely send and receive information with someone in unsecured network for various purposes. To design secure cryptographic protocols, formal analysis for cryptographic protocols should be included as an essential activity in a process of cryptographic protocol design. In other word, the ideal process consists of design, formalization, formal analysis, interpretation, and improvement, and the five activities are done repeatedly as similar as activities in spiral model of software development. This paper presents a supporting tool for the ideal process of cryptographic protocol design. At first, the paper presents the spiral model of cryptographic protocol design, and introduces formal analysis method with reasoning as a suitable formal analysis method for the spiral model. The paper also presents design of the supporting tool and its implementation for key exchange protocols. By the supporting tool, designers can only focus on design and improvement activities in the spiral model.

Keywords Cryptographic protocols · Cryptographic protocol design · Spiral model · Formal analysis with forward reasoning

K. Wagatsuma · T. Harada · S. Anze · Y. Goto · J. Cheng (✉)

Department of Information and Computer Sciences, Saitama University, Saitama 333-8570, Japan

e-mail: cheng@aise.ics.saitama-u.ac.jp

K. Wagatsuma

e-mail: wagatsuma@aise.ics.saitama-u.ac.jp

T. Harada

e-mail: tsubasa@aise.ics.saitama-u.ac.jp

S. Anze

e-mail: anze@aise.ics.saitama-u.ac.jp

Y. Goto

e-mail: gotoh@aise.ics.saitama-u.ac.jp

1 Introduction

Many cryptographic protocols proposed to securely send and receive information with someone in unsecured network for various purposes [2, 6, 12]. For example, key exchange, authentication, digital signature, secret splitting, e-voting, zero-knowledge proof, and so on, are proposed. Especially, many key exchange protocols are proposed and used [2, 6]. However, flaws of proposed protocols were found and attacked after those protocols have been designed and used [2].

To design secure cryptographic protocols, formal analysis for cryptographic protocols should be included as an essential activity in a process of cryptographic protocol design. Formal analysis for cryptographic protocols is used to consider about security before using the protocols [2, 12]. Various tools for analyzing cryptographic protocols based on proving have been proposed. These tools are based on model-checking, for example, Scyther [7], ProVerif [1], or theorem-proving, for example, CafeOBJ [8], Isabelle [11].

In cryptographic protocol design, the ideal process consists of design, formalization, formal analysis, interpretation, and improvement. Furthermore, the five activities are done repeatedly as similar as activities in spiral model of software development. In spiral model of cryptographic protocol design, activities of formalization and interpretation are performed many times. Moreover, it is time-consuming for designers to perform those activities, and designers may make mistakes in those activities.

This paper presents a supporting tool for the ideal process of cryptographic protocol design. At first, the paper presents the spiral model of cryptographic protocol design, and introduces formal analysis method with reasoning as a suitable formal analysis method for the spiral model. The paper also presents design of the supporting tool and its implementation for key exchange protocols.

2 Spiral Model of Cryptographic Protocol Design

2.1 *Spiral Model*

In spiral model of cryptographic protocol design, designers perform five activities as shown in Fig. 1. At first, designers design a cryptographic protocol. Then, designers formalize and do formal analysis of the designed cryptographic protocol. After that, designers interpret what the result of formal analysis means. Finally, designers try to find improvements of the cryptographic protocol. Designers redesign the cryptographic protocol based on the improvements. These activities are repeated many times for improved cryptographic protocol.

By spiral model, designers can design more secure cryptographic protocols progressively by repeating the five activities. When flaws are detected in formal

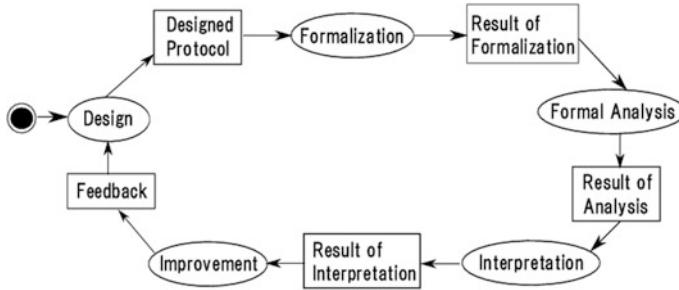


Fig. 1 Spiral model of cryptographic protocol design

analysis, designers improve designed cryptographic protocol without those flaws. As the result, flaws in the designed cryptographic protocol can be reduced.

2.2 Formal Analysis Method with Reasoning

Formal analysis method with reasoning [4] is a method that deduces flaws of cryptographic protocols by forward reasoning from formalized cryptographic protocol as premises. In formal analysis method with reasoning, at first, analysts formalize a specification of cryptographic protocol. After that, many conclusions are deduced by forward reasoning from the result of formalization as premises. Finally, analysts interpret the deduced conclusions whether the analyzed cryptographic protocol has flaws or not. Strong relevant logic [3] is used for a logic system as basis of forward reasoning [4]. Furthermore, FreeEnCal [5] is used for a forward reasoning engine to reduce cost of forward reasoning. FreeEnCal can automatically deduce conclusion from premises represented by logical formulas. FreeEnCal can apply in formal analysis method with reasoning because result of formalization is represented by logical formulas [4].

Traditional formal analysis method based on proving, analysts must enumerate flaws before analysis. In other word, the method can only detect flaws that designers assumed at the design activity. Therefore, traditional formal analysis method based on proving is not suitable for spiral model of cryptographic protocol design.

Formal analysis method with reasoning is suitable for spiral model of cryptographic protocol design. The method is hopeful for detecting flaws that designers did not assume because flaws are deduced by forward reasoning without enumerating those flaws before formal analysis. For improving designed cryptographic protocol, flaws that designers did not assume at activity of design should be detected. Therefore, formal analysis method with reasoning can support to improve designed cryptographic protocol more efficiently.

3 Supporting Tool for Spiral Model of Cryptographic Protocol Design

It is time-consuming for designers to perform formalization and interpretation. If designers perform those activities manually, they may make mistakes in those activities. Furthermore, those activities are repeated many times in the spiral model. Therefore, a supporting tool for those activities is demanded.

The supporting tool consists of sub-tools for formalization, interpretation. Figure 2 explains overview of the supporting tool. In this figure, “FAT” denotes a format of formal analysis tool based on logic, and “LF” denotes logical formulas. The supporting tool includes 4 functions: formalization, transformation into FAT and LF, and interpretation. In the supporting tool, there are 3 types of format that users can input or output: LF, FAT, and format of specification. At first, users input a designed cryptographic protocol that is represented by format of specification, and obtain result of formalization that is represented by LF and FAT. Formal analysis is performed by using the result of formalization. After that, users input the result of formal analysis that is represented by FAT, and obtain result of interpretation that is represented by LF and format of specification. There are various tools for formal analysis of cryptographic protocols based on logic. Input and output FAT tool is different for each tool. Therefore, the result of transformation and interpretation that are represented by logical formulas are necessary for users to understand result of formalization.

By the supporting tool, designers can easily formalize cryptographic protocol, and interpret result of formal analysis. Therefore, designers can only focus on design and improvement activities.

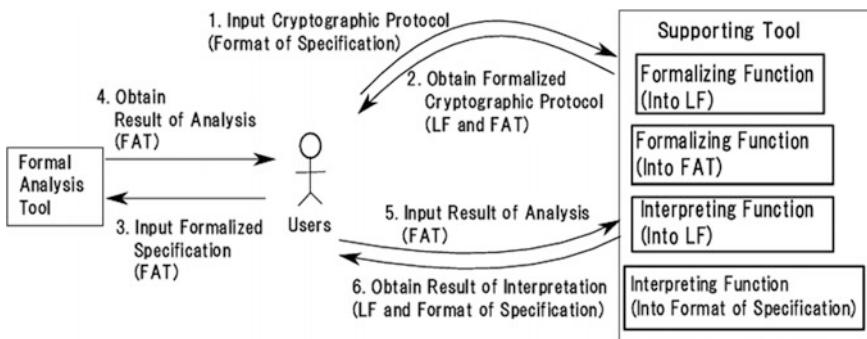


Fig. 2 Overview of the supporting tool

4 Implementation for Key Exchange Protocols

We implemented a supporting tool for formalization and interpretation in spiral model of key exchange protocol design with reasoning-based formal analysis. The supporting tool includes 2 functions, formalizing a specification of key exchange protocol, and interpreting deduced logical formulas. In formalizing function, outputs are logical formulas, and form of FreeEnCal based on proposed formalization method [14]. In interpreting function, outputs are logical formulas and form of specification.

Users can use the supporting tool by following steps. At first, users input a specification of key exchange protocol. A specification of key exchange protocol is represented by following form (1). In the specification, N denotes a step of protocol, X_1 denotes a sender, X_2 denotes a receiver, and Y_1, Y_2, \dots, Y_n denotes sent data. Participants, for example A, B , and S are substituted in X_1 and X_2 . Data types, for example identifier X , nonce N_X , and timestamp T_X are substituted in Y_i . $\{Y_1, Y_2, \dots, Y_n\}_K$ is encrypted data by key K .

$$N: X_1 \rightarrow X_2 : Y_1, Y_2, \dots, Y_n \quad (1)$$

Input specification of key exchange protocol is transformed into logical formulas based on a method of formalization for formal analysis with reasoning [13]. After that, users obtain the logical formulas and form of FreeEnCal. Example of the result of output is shown in Fig. 3. The obtained logical formulas are used as premises, and many conclusions are deduced by FreeEnCal. After that, user input deduced conclusions that are represented by form of FreeEnCal, and obtain result of transformation that are represented by form of specification. Example of the result of output is shown in Fig. 4. Finally, users check output that is represented by form of specification whether the output represents flaws or not.

In order to confirm correctness of the tool, we detected flaws of 4 key exchange protocols that are pointed out by using output from the implemented tool. The target key exchange protocols of formal analysis are Second Protocol Attempt [2], Third

```

 $\forall p1 \forall p2 (\text{Start}(p1, p2) \rightarrow \text{Send}(p1, p2, \text{data}(n, \text{id}(p1), \text{id}(p2),)))$ 
 $\forall p1 \forall p2 \forall x1 (\text{Recv}(p2, \text{data}(x1, \text{id}(p1), \text{id}(p2),)) \rightarrow (\text{Part.i}(p2) \rightarrow \text{Send}(p2, s, \text{data}(x1, \text{id}(p1), \text{id}(p2), \text{enc}(\text{synk}(p1, s), \text{nonce}(p1), x1, \text{id}(p1), \text{id}(p2),)))$ 
 $\forall p1 \forall p2 \forall x1 (\text{Recv}(s, \text{data}(x1, \text{id}(p1), \text{id}(p2), \text{enc}(\text{synk}(p2, s), \text{nonce}(p2), x1, \text{id}(p1), \text{id}(p2), \text{enc}(\text{synk}(p2, s), \text{nonce}(p2), x1, \text{id}(p1), \text{id}(p2)))) \rightarrow (\text{Part.i}(p1) \rightarrow$ 
 $\text{Send}(s, p2, \text{data}(x1, \text{enc}(\text{synk}(p1, s), \text{nonce}(p1), \text{sesk}), \text{enc}(\text{synk}(p2, s), \text{nonce}(p2), \text{sesk}))))$ 
 $\forall p1 \forall p2 \forall x1 (\text{Recv}(p2, \text{data}(x1, \text{enc}(\text{synk}(p1, s), \text{nonce}(p1), k), \text{enc}(\text{synk}(p2, s), \text{nonce}(p2), k))) \rightarrow (\text{Part.i}(p2) \rightarrow$ 
 $\text{Send}(p2, p1, \text{data}(x1, \text{enc}(\text{synk}(p1, s), \text{nonce}(p1), k))))$ 
 $\text{Start}(p1, p2) \rightarrow \text{Get}(i, \text{data}(n, \text{id}(p1), \text{id}(p2),))$ 
 $\text{Start}(p1, p2) \rightarrow \text{Get}(i, \text{data}(n, \text{id}(p1), \text{id}(p2), \text{enc}(\text{synk}(p1, s), \text{old}(\text{nonce}(p1)), m, \text{id}(p1), \text{id}(p2))))$ 
 $\text{Start}(p1, p2) \rightarrow \text{Get}(i, \text{data}(n, \text{enc}(\text{synk}(p1, s), \text{old}(\text{nonce}(p1)), \text{old}(\text{sesk})), \text{enc}(\text{synk}(p2, s), \text{old}(\text{nonce}(p2)), \text{old}(\text{sesk}))))$ 
 $\text{Start}(p1, p2) \rightarrow \text{Get}(i, \text{data}(n, \text{enc}(\text{synk}(p1, s), \text{old}(\text{nonce}(p1)), \text{old}(\text{sesk}))))$ 
 $\forall p1 \forall p2 \forall x1 \forall x2 (\text{Get}(p1, \text{enc}(\text{synk}(p1, p2), x1, x2)) \rightarrow \text{Get}(p1, \text{data}(x1, x2)))$ 
 $\forall p1 \forall p2 \forall x1 \forall x2 (\text{Get}(p1, \text{enc}(\text{synk}(p2, p1), x1, x2)) \rightarrow \text{Get}(p1, \text{data}(x1, x2)))$ 
 $\forall p1 \forall p2 \forall x1 \forall x2 \forall x3 \forall x4 (\text{Get}(p1, \text{enc}(\text{synk}(p1, p2), x1, x2, x3, x4)) \rightarrow \text{Get}(p1, \text{data}(x1, x2, x3, x4)))$ 
 $\forall p1 \forall p2 \forall x1 \forall x2 \forall x3 \forall x4 (\text{Get}(p1, \text{enc}(\text{synk}(p2, p1), x1, x2, x3, x4)) \rightarrow \text{Get}(p1, \text{data}(x1, x2, x3, x4)))$ 
 $\text{Part.i}(a)$ 

```

Fig. 3 Output of transformation into logical formulas

$$\begin{aligned}
 1. A &\rightarrow B:M, A, B, \{N(A), M, A, B\}_{K(AS)} \\
 2'. B &\rightarrow I(S):M, A, B, \{N(A), M, A, B\}_{K(AS)}, \{N(B), M, A, B\}_{K(BS)} \\
 2''. I(B) &\rightarrow S:M, A, I, \{N(A), M, A, B\}_{K(AS)}, \{N(I), M, A, B\}_{K(IS)} \\
 3'. S &\rightarrow I(B):M, \{N(A), K(AB)\}_{K(AS)}, \{N(I), K(AB)\}_{K(IS)} \\
 4. I(B) &\rightarrow A:M, \{N(A), K(AB)\}_{K(AS)} \\
 \vdots
 \end{aligned}$$

Fig. 4 Output of transformation into form of specification

Protocol Attempt [2], Needham-Schroeder Shard Key Protocol [9], and Otway-Rees Protocol [10]. We explain a case in Otway-Rees protocol as follows. Specification of Otway-Rees protocol that we input to the supporting tool is below.

1. $A \rightarrow B:M, A, B, \{N_A, M, A, B\}_{KAS}$
2. $B \rightarrow S:M, A, B, \{N_A, M, A, B\}_{KAS}, \{N_B, M, A, B\}_{KBS}$
3. $S \rightarrow B:M, \{N_A, SK\}_{KAS}, \{N_B, SK\}_{KBS}$
4. $B \rightarrow A:M, \{N_A, SK\}_{KAS}$

We input deduced conclusions from the result of formalization as premises, and confirm that flaws of the protocol are detected. This key exchange protocol is pointed out 2 flaws. At first, attacker falsifies data encrypted by a participant's key into data encrypted by attacker's key. As the result, attacker can obtain session key [2]. First flaw of Otway-Rees protocol is below.

1. $A \rightarrow B:M, A, B, \{N_A, M, A, B\}_{KAS}$
- 2'. $B \rightarrow I(S):M, A, B, \{N_A, M, A, B\}_{KAS}, \{N_B, M, A, B\}_{KBS}$
- 2''. $I(B) \rightarrow S:M, A, I, \{N_A, M, A, B\}_{KAS}, \{N_I, M, A, B\}_{KIS}$
- 3'. $S \rightarrow I(B):M, \{N_A, SK\}_{KAS}, \{N_I, SK\}_{KIS}$
4. $I(B) \rightarrow A:M, \{N_A, SK\}_{KAS}$

Second, if attacker send back encrypted data to participant A , A fails to interpret the received data M, A, B as session key [6]. Second flaw of Otway-Rees protocol is below.

1. $A \rightarrow I(B):M, A, B, \{N_A, M, A, B\}_{KAS}$
- 4'. $I(B) \rightarrow A:M, \{N_A, M, A, B\}_{KAS}$

The result of interpretation is equal to these 2 flaws of Otway-Rees protocol that are pointed out. For example, output of the first flaw is represented in Fig. 4. In other word, these 2 flaws of Otway-Rees protocol are detected by using the supporting tool. In cases of the others key exchange protocol, flaws of these key exchange protocols are detected as well. Therefore, we can say that the result of formalization and interpretation by the supporting tool are correct.

In order to show that the tool can perform formalization and interpretation in a short time compared to perform those activities manually, we performed those activities by manually and using the supporting tool, and measured duration.

Table 1 Duration of formalization and interpretation manually

Name of protocol	Duration manually (formalization) (min)	Duration manually (interpretation) (min)
Second protocol attempt	13	3
Third protocol attempt	13	3
Needham-Schroeder	20	5
Otway-Rees	25	6

As shown in Table 1, it took about 15 min to formalize and about 5 min to interpret each 4 protocols manually. By using the supporting tool, these task can be finished automatically and instantly. Therefore, the tool can reduce duration of formalization and interpretation. If we assumed that such activities are manually performed many times in spiral model of key exchange protocol design, it is not efficient for designers. Therefore, the supporting tool is effective for key exchange protocol design in spiral model.

5 Concluding Remarks

This paper presented a supporting tool for spiral model of cryptographic protocol design. After that, the paper presented design of the supporting tool and its implementation for key exchange protocols, and showed effectiveness of the supporting tool.

We showed that the supporting tool could perform formalization and interpretation correctly in a short time. This effectiveness is same in not only key exchange protocols, but also various cryptographic protocols. It is different for cryptographic protocols only the method of formalization and interpretation. In cryptographic protocol design, designers also perform formalization and interpretation and those activities are time-consuming.

We can extend the supporting tool to be applied to various cryptographic protocols. In the cryptographic protocols, participants send and receive data using cryptography, and participants repeat sending, receiving, and getting data by decryption, are common definition. Therefore, there are common tasks for formalization and interpretation in various cryptographic protocols. These common tasks also can be used in various cryptographic protocols. We only have to change method of the other tasks depending on cryptographic protocols. Currently, target of the supporting tool is limited only to key exchange protocols. In the future, we will extend the supporting tool to be applied to various cryptographic protocols and show its effectiveness.

References

1. Blanchet, B.: An efficient cryptographic protocol verifier based on prolog rules. In: 14th IEEE Computer Security Foundations Workshop, pp. 82–96. IEEE (2001)
2. Boyd, C., Mathuria, A.: Protocols for Authentication and Key Establishment. Springer, Berlin (2003)
3. Cheng, J.: A strong relevant logic model of epistemic processes in scientific discovery. *Inf. Modell. Knowl. Bases* **61**, 136–159 (2000)
4. Cheng, J., Miura, J.: Deontic relevant logic as the logical basis for specifying, verifying, and reasoning about information security and information assurance. In: Proceedings of the 1st International Conference on Availability, Reliability and Security, pp. 601–608. Vienna (2006)
5. Cheng, J., Nara, S., Goto Y.: FreeEnCal: a forward reasoning engine with general-purpose. In: Knowledge-Based Intelligent Information and Engineering Systems. Lecture Notes in Artificial Intelligence, vol. 4693, pp. 444–452. Springer, Berlin (2007)
6. Clark, J., Jacob, J.: A survey of authentication protocol literature: version 1.0. <http://www.cs.york.ac.uk/~jac/> (1997)
7. Cremer, C.: The Scyther Tool: verification, falsification, and analysis of security protocols. In: Proceedings of the 20th International Conference on Computer Aided Verification. Lecture Notes in Computer Science, vol. 5123, pp. 414–418. Springer, Heidelberg (2008)
8. Diaconescu, R., Futatsugi, K.: CafeOBJ Report. AMAST Series in Computing, vol. 6. World Scientific, Singapore (1998)
9. Needham, R.M., Schroeder, M.D.: Using encryption for authentication in large networks of computers. *Commun. ACM* **21**(12), 993–999 (1978)
10. Otway, D., Rees, O.: Efficient and timely mutual authentication. *ACM SIGOPS Operating Syst. Rev.* **21**(1), 8–10 (1987)
11. Paulson, L.C.: Isabelle: a generic theorem prover. In: Lecture Notes in Computer Science, vol. 828. Springer, Berlin (1994)
12. Schneier, B.: Applied Cryptography. In: Schneier, B. (ed.) *Protocols, Algorithms, and Source Code in C*. Wiley, New York (1996)
13. Wagatsuma, K., Anze, S., Goto, Y., Cheng, J.: Formalization for formal analysis of cryptographic protocols with reasoning approach. In: Future Information Technology. Lecture Notes in Electrical Engineering, vol. 309, pp. 211–218. Springer, Heidelberg (2014)
14. Wagatsuma, K., Goto, Y., Cheng, J.: A formal analysis method with reasoning for key exchange protocols. *IPSJ J. (in Japanese)* **56**(3), 903–910 (2015)

Idea of Personal Digital Memories Using Smart Application

Martin Zmitko and Ondrej Krejcar

Abstract The world around us is constantly developing and the life has dramatically changed in the last few decades. Even the requirements themselves grow with the development of technology. The human life is more surrounded by things which enable easier and more convenient life. These could be automatically opening doors, smart regulators for temperature in the house or even mobile application with a specific advantage for users. With this in mind, an application for more convenient work with photographing was developed and is described in this paper. Paper deal with the description of individual parts of the application from the draft, the architecture of the solution, to the implementation.

Keywords Personal · Digital memories · Smart application · Tourist diary

1 Introduction

One of the current trends of personal memories targeted on tourist diary application for mobile devices, which are closets to users where can be easily used as native personal memories source [1–5]. Tourist diary is a mobile application [6–8] which works on the Android platform and helps users who travel and photograph. The aim of this paper was to create an easy and simple application for controlling the creating of image and video to new personalized structure created at the mobile device directory. This means that according to the available sources, the application recognises the position of the device and then creates an address structure for

M. Zmitko · O. Krejcar (✉)

Faculty of Informatics and Management, Center for Basic and Applied Research,
University of Hradec Kralove, Rokitanskeho 62, 500 03 Hradec Kralove, Czech Republic
e-mail: Ondrej@Krejcar.org

M. Zmitko

e-mail: martin.zmitko@uhk.cz

individual photos. As an additional value, the application offers information about the given place which can be useful to the user as a form of a diary for the visited places, including the photos that were taken.

2 Current Solution of Personal Memories—Tourist Diary

As one of the examples for tourist diary applications, We select one [9] which is the tourist diary for travelling (Fig. 1a). The above mentioned tourist diary has also its own mobile application Tourist diary [9]. This application and the printed diary serve the user as a tool through which they can collect visiting cards from the particular places. For the printed form of the diary, it is possible to select the type of transportation, weather or the type of the trip. In contrast to the printed version, the application [9] only allows the viewing of the visiting cards, finding the closest places for visits according to the location of the device or the viewing of the card (Fig. 1a).

However, the aim of this paper is to interconnect photographing which is a common activity during travelling or hiking with different places. In this matter, the newly built application offers excellent possibilities for the user which the above mentioned application does not cover. This includes mainly the automatic localisation from the place where the photo was taken and automatically completing the data about the place instead of having the user do it (contrary to the manual completing in the tourist diary).

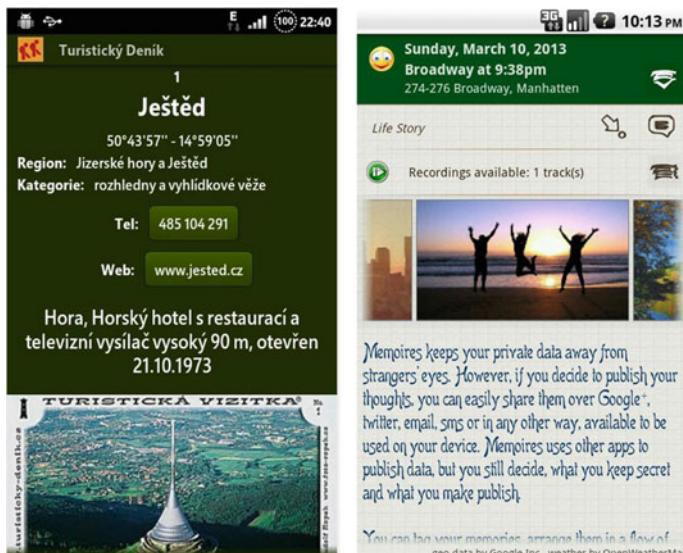


Fig. 1 Example of the tourist diary application (a-left) and personal memories application (b-right)

2.1 Personal Memories: The Diary

From the view of the photographing itself and the principle, the drafted application is most closely similar to the international project Memories: the Diary [10]. This application works on the principle of creating records in the diary (Fig. 1b).

The paper positively judges the way in which the application assigns notes to individual photos. Other very useful function of Memories is the synchronisation with Google account. This function was not implemented in the suggested application of the paper, but it is a possible option for the future to extend the application. Memories application required for each photo the input of a note of place which is not the ideal solution for ordinary photographing. Moreover, the application starts the camera from the Android system which increases the time of preparation for photo taking. From the above mention description, it is obvious that the application is built on individual notes to which the photographs, sound recordings or other metadata are assigned. In this matter the application is different to the suggested solution from this paper where the main subject of interest is the place together with photos and the other elements are viewed as additional data. From this point of view, the suggested application is more convenient for the users who wish to take photos from travels, as well as obtain certain additional information that the application automatically provides.

Common elements which both applications share:

- Obtaining information about the place according to location
- Adding notes to photographs (place)
- Working with camera of the mobile device

In comparison to Memories, the advantage of the suggested application is the focus on the specific place from which the photos are taken. The application concentrates on individual places for which it searches for additional information about weather and temperature and consequently, allows the user to add notes. Another advantage, compared to Memories and common camera, is the fact that the photos are ordered in particular directories on the SD card of the device.

3 Architecture of the New Solution

Both mentioned examples of personal memories solutions are taken into account to use the best ideas from the both to develop a new solution as a smart mobile application. The application was created as a two-layered application (frontend and backend). Frontend is an environment which is perceived by the user and with which the user is in contact. Backend creates its own background of the application which enables the connection with the database, working with the camera and other matters.

Frontend:	own photographing and location according to user's position (Wi-Fi, GPS, GSM [11–13]), entering the weather, type of road, temperature, description or transportation
Backend:	working with SQLite db, shared preferences, obtaining data and their processing
Technologies:	Android Framework version 2.3, Android developers tools
Minimum required SDK:	API: 10 Android 2.3.3 (Gingerbread)
Target SDK:	API: 17 Android 4.0.3 (Jelly Bean)

3.1 Main Functionality

- Obtaining position on the background of the application
- Downloading of useful information (temperature, weather)
- Generating of the tourist diary according to data
- Possibility to add information about transportation, type of the trip, notes by users

3.2 Description of the Current Functionality of the Application

The whole application is designed using few activities among which the user can switch. The main activity is the camera. This activity starts when the application is turned on. From this activity it is possible to switch to the activity settings when pressing the key “menu”, where the user can adjust the basic parameters of the application. Another important activity, which is also accessible from the main activity, is the diary activity. This activity contains the tourist diary which shows particular visited places in a manner which is well-structured according to the year of the picture taking. After the selection of the place, the detail of the activity is shown (Fig. 2). From the detail of the application and after the selection of the specific photo, it is possible to switch to the activity of the enlarged photography.

3.2.1 Main Activity—Camera

Main activity which is the actual camera is immediately ready to be used (Fig. 4). Apart from the basic functions of the camera there are also other buttons designed for this activity. These include most importantly the “My diary” button for the

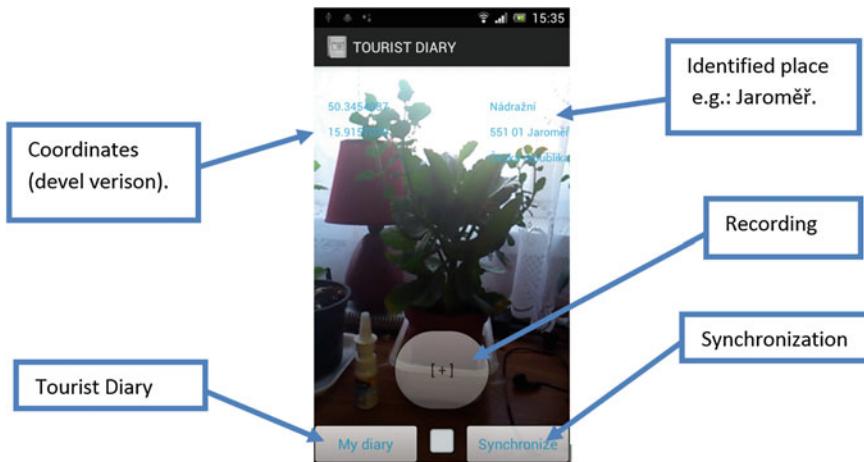


Fig. 2 Main activity—camera

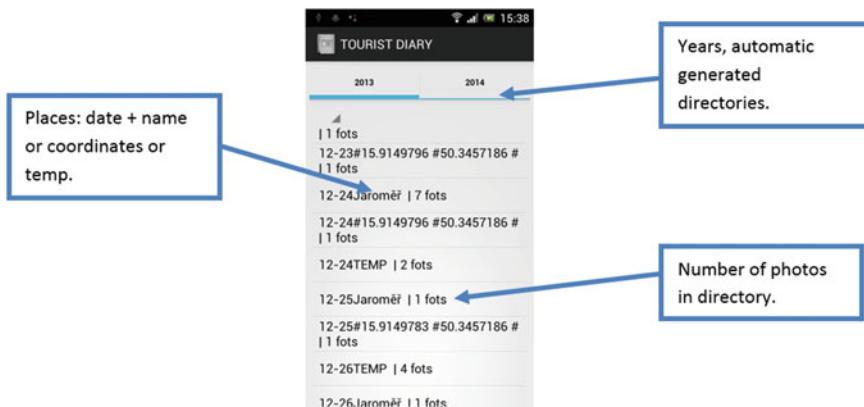


Fig. 3 Overview of places

access to the generated tourist diary. Moreover, there is also the synchronisation button (which will be used in the future) that will allow the user to synchronise data with photos and assign metadata to created albums during an active internet connection.

When starting the main activity, the application itself recognises whether any of the modules (GSM, GPS or Wi-Fi) are running and notifies the user about this fact through Android—toast. In the case of both of the modules for obtaining position being turned off, the user is informed about this matter through dialogue window.

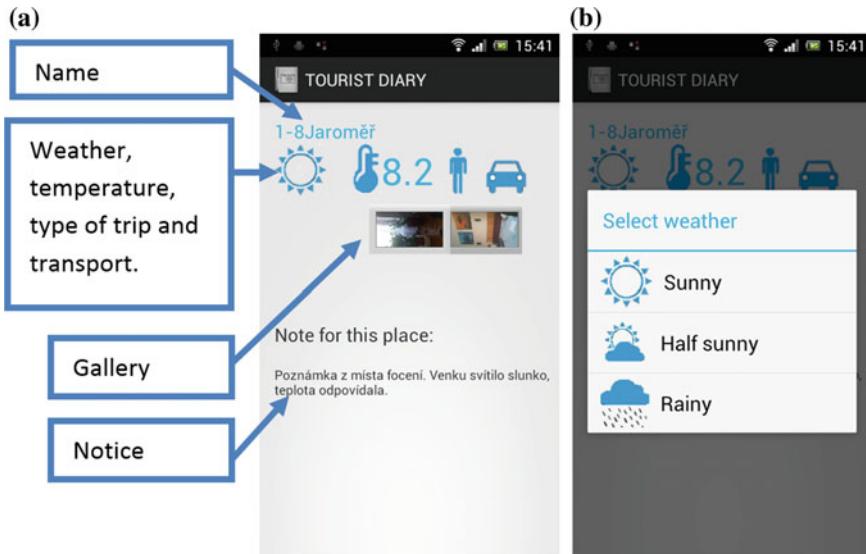


Fig. 4 Detail of a place (a)—left figure; Selecting weather (b)—right figure

3.2.2 Tourist Diary Activity

Another significant advantage of this application for the user is seen in the regeneration of the tourist diary. The tourist diary activity displays a well-structured list of visited places which are grouped into individual years. The activity shows the user month-day of the visit of the place, name of the place (if it exists) and the number of the current photos assigned to this place (Fig. 3). After selecting the chosen place, the following activity is started—tourist diary—detail.

This activity individually displays visited place according to the obtained information and it provides itself the relevant icon showing the weather and the temperature. Furthermore, it shows the details of photographs which belong to this locality together with the possibility to choose the type of transportation and road, as shown on the (Fig. 4a, b). The activity also enables users to add notes about the place and the change of the above described items (temperature, weather, type of the trip, type of the transportation).

4 Conclusions

In the future, the application could have its own connection with Facebook or Google+ (for example as a larger module of the application purchased for a smaller amount) or could generate tourist diary in a web-form or offline form (PDF). Moreover, it is necessary to finish programming the synchronisation. In conclusion

it is important to mention that the application is available in beta version which may contain small mistakes. However, the main function and framework of the application is already fully functional by the result of testing according to parameters [3, 13, 14].

Acknowledgment This work and the contribution were supported by project “SP-2103-2015—Smart Solutions for Ubiquitous Computing Environments” Faculty of Informatics and Management, University of Hradec Kralove, Czech Republic.

References

1. Machacek, Z., Slaby, R., Hercik, R., Koziorek, J.: Advanced system for consumption meters with recognition of video camera signal. *Elektronika Ir Elektrotehnika*. **18**(10), 57–60 (2012)
2. Moulin, C., Lai, C.: Harmonization between personal and shared memories. *Int. J. Software Eng. Knowl. Eng.* **20**(4), 521–531 (2010)
3. Lindley, S.E.: Before I forget: from personal memory to family history. *Hum. Comput. Interact.* **27**(1–2), 13–36 (2012). doi:[10.1080/07370024.2012.656065](https://doi.org/10.1080/07370024.2012.656065)
4. Ozana, S., Pies, M., Hajovsky, R., Haska, J., Koziorek, J., Horacek, O.: Application of PIL approach for automated transportation center. In: 13th IFIP TC8 International Conference on Computer Information Systems and Industrial Management (CISIM 2014), LNCS, Vol. 8838, pp. 501–513 (2014)
5. Cerny, M., Penhaker, M.: Wireless body sensor network in health maintenance systems. *J. Electron. Electr. Eng.* **115**(9), 113–116 (2011)
6. Krejcar, O.: Threading possibilities of smart devices platforms for future user adaptive systems. In: Proceedings of the 4th Asian Conference on Intelligent Information and Database Systems, ACIIDS 2012, Kaohsiung, Taiwan. LNCS Vol. 7197, pp. 458–467, 19–21 Mar 2012
7. Behan, M., Krejcar, O.: Adaptive graphical user interface solution for modern user devices. In: 4th Asian Conference on Intelligent Information and Database Systems, ACIIDS 2012, Kaohsiung, Taiwan. LNCS Vol. 6592, pp. 411–420, 19–21 Mar 2012
8. Behan, M., Krejcar, O.: Modern smart device-based concept of sensoric networks. *EURASIP J. Wirel. Commun. Netw.* **155**(1) (2013)
9. Wander Book: Tourist diary (online). Online at: <http://turisticky-denik.cz/> (2009). Cited in 8 Jan 2015
10. Google Play: Tourist diary. Google Play (online). Online at: <https://play.google.com/store/apps/details?id=org.pavlicek.tdenik&hl=cs> 2013. Cited in 8 Jan 2015
11. Benikovsky, J., Brida, P., Machaj, J.: Proposal of user adaptive modular localization system for ubiquitous positioning. *Lect. Notes Artif. Intell.* **7197**, 391–400 (2012)
12. Mlynka, M., Brida, P., Machaj, J.: Modular positioning system for intelligent transport. In: 5th International Conference on Computational Collective Intelligence Technologies and Applications, Craiova, Romania, Studies in Computational Intelligence, vol. 513, pp. 115–124 (2014)
13. Pies, M., Hajovsky, R., Latocha, M., Ozana, S.: Radio telemetry unit for online monitoring system at mining dumps. *Appl. Mech. Mater.* **548–549**, 736–743 (2014)
14. Balik, L., Horalek, J., Sobeslav, V., Hornig, O.: Remote laboratory for computer networks. In: Proceedings of the 5th International Conference on Data Communication Networking DCNET 2014, Part of ICETE 2014—11th International Joint Conference on e-Business and Telecommunications, pp. 28–34 (2014)

Proxy Based Mobility Management Scheme Using Prediction Algorithm

Daewon Lee, Daeyong Jung, Doo-Soon Park and HwaMin Lee

Abstract Nowadays, mobile users are rapidly increased and many mobile environments are proposed by the improvement of wireless devices and network. There are many mobility management protocols are proposed to provide seamless mobile network environment. In this paper, we focused on transportation environment that is one of bus, train, airplane or ship that takes passengers using mobile devices. Transportation environment has a major problem that Mobile Node (MN) moves fast through multiple cells. Each MN in transportation keeps updating its current point of attachment to its Home Agent (HA) and corresponding node that causes triangle routing problem between HA of MN and HA of Mobile Router (MR). To remove triangle routing problem and minimize cost, we proposed mobility management scheme adapting proxy scheme then we propose prediction algorithm. Based on two features, we design proxy based architecture and extend router advertisement message to alert join in proxy domain. Then, we are classified the mobility into intra proxy mobility and global mobility. And we proposed proxy based mobility management scheme using prediction algorithm. We added heuristic that takes into account real world transportation's movement. By numerical analysis, we shows proposed scheme reduces signaling overheads and increase packet transfer rate than NEMO and proxy MIP.

Keywords NEMO · Proxy MIP · Router advertisement · Binding update

D. Lee

Division of General Education, Seokyeong University, Seoul, Korea
e-mail: daelee@skuniv.ac.kr

D. Jung

Department of Computer Science Education, Korea University, Seoul, Korea
e-mail: karat@korea.ac.kr

D.-S. Park · H. Lee (✉)

Department of Computer Software Engineering, Soonchunhyang University, Asan, Korea
e-mail: hwamin77@gmail.com; leehm@sch.ac.kr

D.-S. Park

e-mail: parkds@sch.ac.kr

1 Introduction

By the improvement of mobile devices and Internet techniques, new types of mobile environments are come out. The transportation environment is one of bus, train, airplane or ship that takes passengers using mobile devices. According to change the mobile environment, many mobility management protocols are proposed to provide seamless network on mobile environment. Transportation, such as bus, train, airplane or ship is one of mobile environment that takes passengers using mobile devices. In this environment, Mobile IP has two disadvantages. In this environment, Mobile IP has two disadvantages. The one is that Home Agent (HA) cannot tunneling packets to Mobile Router (MR) in mobile network. The other is that HA and MR cannot exchange routing information so that they cannot make routing table including entries for prefixes which are used in mobile network [1]. To solve those two disadvantages, NEtwork MObility (NEMO) is proposed that divided problem situation into three models for MR to remain connection to network continuously in case of MR's point of attachment is changed [1]. As people who use smart device such as smart phone keep increasing, signaling cost for Mobile Node (MN) to register also increases. Especially, fast moving mobile nodes only connect little time through its point of attachment or even not connect to network make superfluous traffic to themselves and also to network. Willkie et al. [2] is proposed to overcome these problems by domain based approach.

In transportation environment, Proxy MIP (PMIP) is suitable structurally and easily apply. Transportation environment has a major problem that mobile node moves fast through multiple cells. Each MH in transportation keeps updating its current point of attachment to its HA and Corresponding Node (CN). To overcome this problem we consider feature of transportation. There are two major features. The one is mobile device have fast mobility by transportation. The other one is mobile device moves fixed route. It causes mobile device has intra mobility between several subnets of one agency when it moves into a subway or highway. Based on two features, we design proxy based architecture and extend router advertisement message to alert join in proxy domain. Then, we are classified the mobility into intra proxy mobility and global mobility. And we proposed proxy based mobility management scheme using prediction algorithm.

This paper is organized as follows: Sect. 2 explains system design and architecture. Then proposed mobility management scheme for transportation environment is shown at Sects. 3 and 4 shows numerical analysis between proposed scheme PMIP and NEMO. Finally, Sect. 5 concludes this paper.

2 System Design and Architecture

In transmission environment that we focused on, the mobility of mobile network that consist multiple mobile routers and mobile nodes which connected to the mobile routers couldn't be supported. Transportation that has static characteristic is

similar to NEMO model 2 scenario at [1]. Looking at the solving procedure occurred at MIPv6, if MR gets new CoA from Stationery Router (SR) and send binding update, MN which consider MR as Access Router (AR) never consider whether it moves or not. So they store their prefix to their HA of mobile router that whenever CN sends packet to MN, packet will be sent to HA_{MN} first and then tunnelled to HA_{MR} . Next, it is tunnelled to MR and finally arrived at MN. However, whenever MN changes its point of attachment at NEMO environment, it should register its HA and CN that causes additional cost which called triangle problem among HA_{MN} , HA_{MR} and MN [1].

To overcome this problem, we consider two major characteristics of transportation environment that is focused in this paper. The one is mobile device have fast mobility by transportation. The other one is mobile device moves fixed route. It causes mobile device has intra mobility between several subnets of one agency when it moves into a subway or highway. These characteristics are structurally well suited to the PMIP architecture.

In case of subway, it has fixed route and moves on schedule, prediction of current location is at certain time can be available [3]. Also, there are accurate data such as average distance among stations, length of single train car, increasing rate of velocity, maximum number per single train. So, efficient network infrastructure design can be available. Besides, as the train drives fixed route, binding update (BU) is sent to only one time when the MR/MN enters in a proxy area even the MR crosses multiple cells within a little time. Transportation which has static characteristic in this paper, needs multiple mobile routers for AP of mobile node and SR for AP of MR. Figure 1 shows proposed architecture based on PMIP.

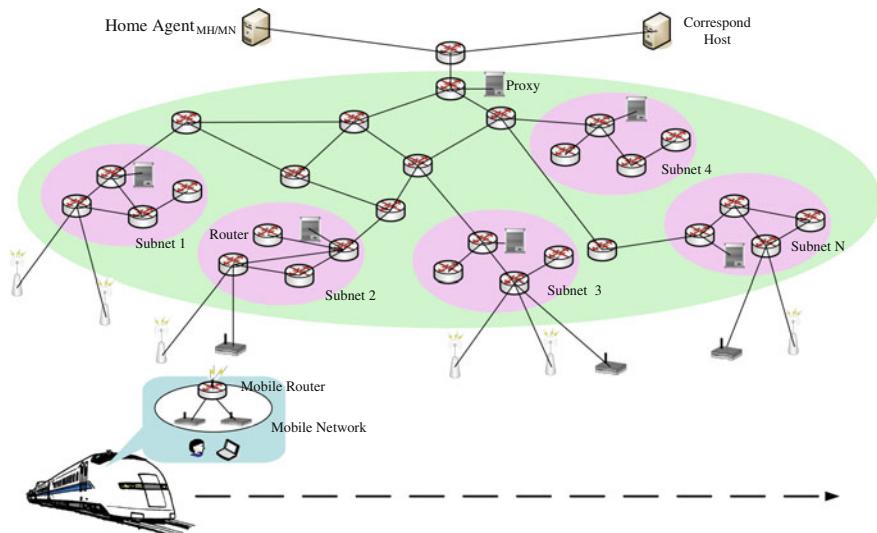


Fig. 1 Network architecture based on PMIP

3 Proposed Mobility Management Scheme for Transportation Environment

Transportation takes multiple MNs and MRs register in several SRs as train pass through cell. Registration procedure is explained as follows. MR firstly enters an SR's coverage and gets Router Advertisement message (RA) from SR. For our scheme we modified RA's reserved field for proxy. To be more we added P flag to check whether this MR supports our scheme. If SR supports our scheme, SR set flag P and contain proxy's unique ID, its own address. The modified RA format is showed at Fig. 2.

Proposed procedure for MR's first registration is described at Fig. 3. Packet delivery path is also described at Fig. 3. CN sends packet to HA_{MR} and after finding MR's information at HA_{MR} then tunnel the packet to proxy according to former binding update. The proxy forwards packet to MR at last.

MN gets on the transportation and takes MR's RA and register in MR. First time to register MR, MN receives RA and checks it is entered in proxy area. In this case, MN doesn't enter proxy area yet, so it sends BU taking its own CoA and MR's CoA to proxy. This procedure store MR to MR mapping information like a chain and can be used to predict location of MR. Proxy receives BU from MN and then stores mapping information, setups timer to calculate MN location. MN adds its own CoA into received RA and relay it. It sends BU to HA_{MN} containing proxy ID so HA_{MN} store proxy ID as MN's CoA at its binding cache. Packet delivery path to MN is nearly same as packet delivery path to MR described at Fig. 3.

4 Performance Evaluation

4.1 Numerical Analysis

This section compares performance of PMIP, NEMO, and proposed scheme from the perspective of signaling cost for registration when a single MR passes through multiple cells and packet delivery cost to MR using provided notations and

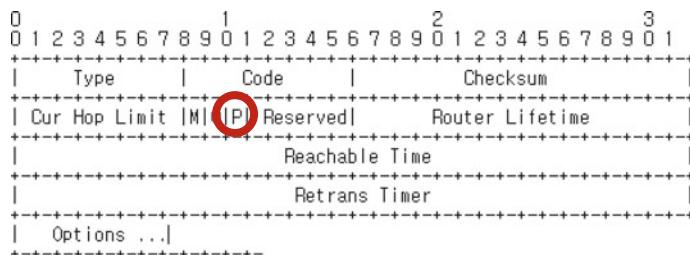


Fig. 2 Revised router advertisement format

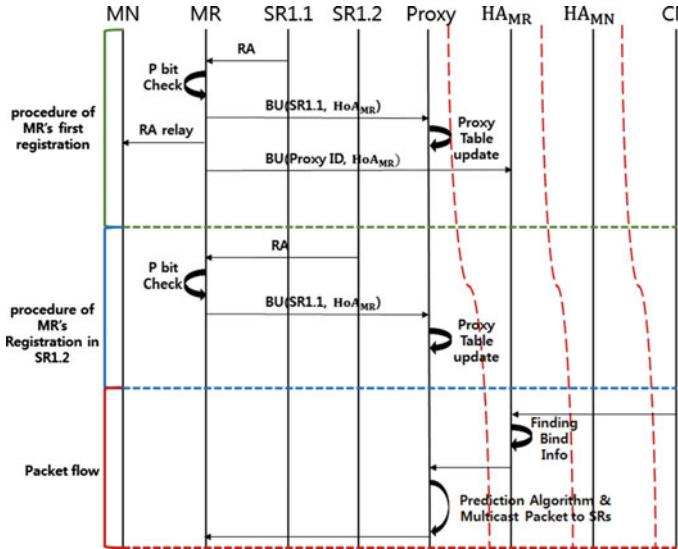


Fig. 3 Call flow for first registration of MR

numerical formulas. We also compare performance of PMIP and proposed scheme from the perspective of packet arrival rate after applying MR location prediction algorithm. We divide MN's mode into idle and active for calculating cost. α means node's idle rate [2] and other necessary variables are shown at below with numerical formulas (Table 1).

$$S_s \times N_{MN} \times N_{cc}$$

$$= ((1 - \alpha) \times S_a + \alpha \times S_i) \times N_{MN} \times N_{cc} \quad (1)$$

$$= [\alpha \{a \times C_{in} + b \times C_{out}\} + (1 - \alpha) \{c \times C_{in} + d \times C_{out}\}] \times N_{MN} \times N \quad (2)$$

The cost is calculated with follow assumptions. A single MR carrying number of MNs and MR pass through a single proxy area composed of m cells. Then extend formula that MR pass through multiple paging areas. Total cost including registration, signaling for location update, and packet delivery is divided into two groups that first one is internal hop signal: signal from MN to proxy or proxy to MN and second one is external hop signal: signal from proxy to HA_{MN} , HA_{MN} , CN or HA_{MN} , HA_{MN} , CN to proxy that need to pass through backbone [4].

Pure packet delivery cost is largely related with hops from departure to destination. So the cost for external signal which pass through backbone is much larger than cost for internal signal. Performance of NEMO, PMIPv6 NEMO, and proposed scheme is compared from the perspective of number of packet delivery, MR's crossing cell number, and number of cells composing a single paging area. Binding update procedure differentiate packet delivery path so those three schemes

Table 1 Parameter for formula (1, 2)

S_s	Signaling cost for a single node	N_{cc}	Number of crossing cell
S_a	Signaling cost for active node	α	Node's idle rate
S_i	Signaling cost for idle node	C_{in}	Internal signaling cost
N_{MN}	Number of mobile node per single MR	C_{out}	Outerinal signaling cost

have different performance on delivering packet. Crossing cell make additional registration cost to PMIP.

Use Timer in proxy table to predict location of MR

divide timer into 120 and according to remainder, separate value into three groups

first group's remainder 0–55: velocity increasing at same rate section

second group's remainder 55–110: velocity decreasing at same rate section

third group's remainder 110–120: down time section

Quotient number means number of station.

So we predict MR is far from $1.2 \text{ km} * N_s + \text{additional distance}$

Divide result by cell coverage and number of cell organizing proxy area. Then the quotient will be number of certain proxy and remainder will be predicted cell including MR in that proxy

Finally, Multicast packet to predicted SR and several SR nearby predicted SR

In numerical analysis, to reflect real world instead of vague design, the real world elements are used such as increase rate of velocity, velocity limit, distance between different stations, etc. MR is assumed a single train. As we mentioned above, paging request cost is main factor of evaluating performance between PMIP and proposed scheme. Proposed scheme reduces packet forwarding cost by predicting MR's location and multicast packet to correspond SR and a few nearby SR. We used timer to calculate how many stations, cells MR pass through. Specific algorithm for calculating cost and formulas which is related to multicasting packet cost is described as follows.

$$p(t) = \lceil d(t)/2R \rceil \quad (3)$$

$$= \left\lceil \left(\left(\int_0^{t_1} V_1 dt + \int_{t_1}^{t_2} V_2 dt \right) + N_S \times 1.2 \right) / 2R \right\rceil \quad (4)$$

$$= \lceil (t_1^2(1/2m + 1/2o) + t_2^2(1/2o) + n t_2 - o^* t_1 t_2) + N_S \times 1.2 / 2R \rceil \quad (5)$$

In this part, we explain specific algorithm for predicting MR's location. As mentioned above, we use real world transportation's movement. The train starts

Table 2 Parameter for formula (3)–(5)

$p(t)$	Predicted location of MR (number of cells)	T_{inc}	(T-downtime)/2
$d(t)$	MR's location at time t	N_S	Timer/120(s)
V_1	$mt+n(0-t_1)$	m	Increasing rate of velocity
V_2	$ot+p(t_1-t_2)$	o	Decreasing rate of velocity

from stationary state and increase its speed at same rate. When the speed is up to limit speed, it decreases its speed at same rate and takes down time for passengers to get on or get off. This is one cycle between two different stations. Train's state is increase speed or decrease speed or down time so we use those three states to calculate location accurately. Formulas and values needed to calculate are described at Table 2.

We use general wireless protocol to calculate cell's coverage and the number of cell and paging area where the MR is at last. Our proposed scheme use timer at page table since MR's first registration to calculate passed time and use passed time, trains movement, and cell's coverage to predict MR's location. In case of error proposed scheme send paging request not to only one SR but to predicted SR and some nearby SRs.

4.2 Analysis Results

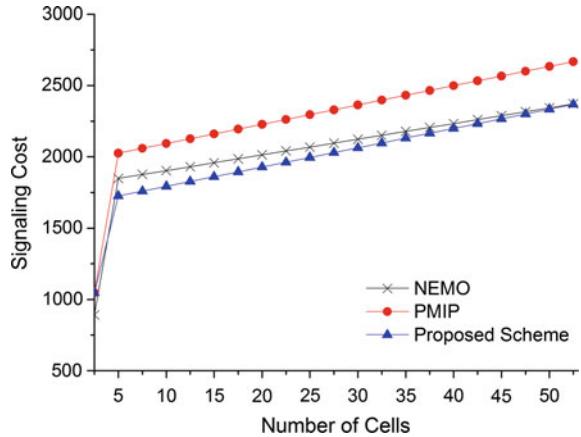
We compare performance of three method based on registration and packet delivery cost. Table 3 shows the parameters used in our numerical analysis, which is discussed in [2, 5, 6]. Main variables are N_{cc} , number of packet to MN per SR, and number of cell organizing a single paging area. Graphs below are results of mathematical analysis. Specific details about graph are explained as follows.

In Fig. 4, we fixed number of packet to MN per SR and controlled N_{cc} and number of cell organizing a single paging area. As there is no triangle problem in proposed scheme if number of packet increase than schemes except proposed scheme cost will drastically increase. There is small number of packet delivery per SR since our environmental assumption makes transportation to cross cell rapidly. The graph shows that cost increasing rates of PMIPv6 NEMO and proposed scheme are bigger than original NEMO. It is because of additional registration cost such as

Table 3 Parameter for numerical analysis

N_{MN}	80	T	120 s
M	3 km/h/s	<i>Downtime</i>	10 s
o_1	-3 km/h/s	<i>Avg. dist. between two stations</i>	1.2 km
<i>Velocity limit</i>	40 km/h	<i>Total distance</i>	10.8 km

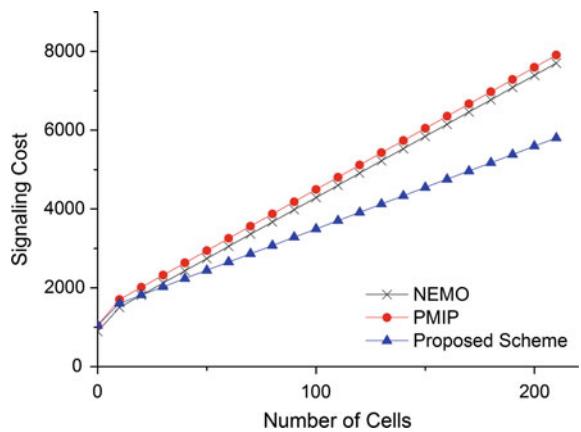
Fig. 4 Total signalling cost on number of cells



PTU. Considering cost of constructing paging network, Paging method on NEMO decrease little cost for registration.

In Fig. 5, we controlled number of packet to MN per SR and fixed Ncc and number of cells per single paging area. We can recognize that packet delivery route correction by solving triangle problem is main strong point of proposed scheme. Furthermore cost for packet delivery to MN is weighted value compared to other variables. Although increasing cost of registration as number of cell per paging area increases is relatively small compared to advantages due to solving triangle problem. As a result packet receive per SR has powerful effect to signaling cost calculation.

Fig. 5 Total signalling cost on number of transmit packets



5 Conclusion

Nowadays, mobile users are rapidly increased and many mobile environments are proposed by the improvement of wireless devices and network. There are many mobility management protocols proposed to provide seamless mobile network environment. In this paper, we focused on transportation environment. To remove triangle routing problem and minimize cost, we proposed mobility management scheme using proxy and prediction algorithm. First of all, we used different binding update procedure to remove triangle problem. We modified router advertisement message and binding update message by adding flag P meaning whether mobile router supports our proposed scheme, Secondly, to reduce registration cost, we used prediction algorithm using proxy table's mapping information and timer value. We also added heuristic that takes into account real world transportation's movement. By numerical analysis, we shows proposed scheme reduces signaling overheads than NEMO protocol. In future work we are concerned about method minimizing registration cost by reducing external packet.

Acknowledgments This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education ((No.2014R1A1A2057878)).

References

1. Okajima, I., Umeda, N., Yamao, Y.: Architecture and mobile IPv6 extensions supporting mobile networks in mobile communications. In: Vehicular Technology Conference, 2001. VTC 2001 Fall. IEEE VTS 54th, vol. 4, pp. 2533–2537 (2001)
2. Willkie, J.J., Lioy, M., Armstrong, S.M.: U.S. Patent No. 6,230,012. U.S. Patent and Trademark Office, Washington, DC (2001)
3. <http://www.seoulmetro.co.kr/>
4. Kabir, H., Mukhtaruzzaman, M., Atiquzzaman, M.: Efficient route optimization scheme for nested-NEMO. J. Netw. Comput. Appl. **36**(3), 1039–1049 (2013)
5. Kato, T., Takechi, R., Ono, H.: A study on mobile IPv6 based mobility management architecture. Fujitsu Sci. Tech. J. **37**(1), 65–71 (2001)
6. <http://www.seoul.go.kr/seoul/summary/sense/transl.html>

Principles of Usability in Human-Computer Interaction

Tomas Hustak and Ondrej Krejcar

Abstract This paper try to address problems that come out when designing a web page and how these problems affect people who use it. We will try to find the way to make our web pages easy to use and to give our user as little obstacles in his efforts as possible. He has to feel that every time he comes back to our web site, everything is on places where it should be and that everything acts as it should. One could say he has to see that everything looks normal. In this paper our aim will be to address web users, their needs and ways to improve their overall efficiency with achieving their tasks. We go through all main parts of usability issues according HCI as well as the current state of the art based on the most cited or specialised literature.

Keywords Usability • Websites evaluation • Human-computer interaction

1 Introduction

Many web designers develop their projects without any feedback from the target audience and mainly focus to meet the contract requirements [1]. They probably think users do not have adequate technical knowledge to understand such a complex field which web designing undoubtedly is. This is mostly true, on the other hand users know exactly what they need and that is something web designers cannot ignore. For even the most experienced web designers it is impossible to design a highly usable web site on their first attempt. This is because nothing like an

T. Hustak · O. Krejcar (✉)

Faculty of Informatics and Management, Center for Basic and Applied Research, University of Hradec Kralove, Rokitanskeho 62, Hradec Kralove 500 03, Czech Republic
e-mail: ondrej@krejcar.org; ondrej.krejcar@uhk.cz

T. Hustak
e-mail: Tomas.Hustak@uhk.cz

average user [2] exists, every single user is different thus it is very hard to say what is good for all of them (or at least most of them).

So if we want to create a good web site we need to listen to our users and communicate with them. The best thing we can do is to discuss with our potential users from the very beginning when we start prototyping. [3, 1] That way we can avoid unnecessary problems and design flaws that could cost us more money in the late phase of the development. Having said that, in this paper we will actually focus mainly on the late phase, it is when our web site is established and running. Sure, it is still possible to invite users and ask them questions, but we would like to have something more flexible and hit as many users as possible, which indicates we will need a tool that would give us an idea about what users do on our web pages.

Before we move on, one important note has to be made. Certain words throughout the whole document mean the same thing, this involves mainly words like users, visitors or customers, these mean people who use a web page. However, in some context a word user can mean a person who accesses a web application in order to analyse obtained data, i.e., an administrator. Words like logging, gathering, listening or sending in the client-server context all mean data transfer from the client to the server and their storage.

In the first part of this paper, we will answer the question “What is usability” and what impact usability has on users. The second part discusses HCI standards and their suitability for practical use. And lastly we will talk about interfaces between machines and humans in terms of usability.

2 What Is Usability

Every web page creator or any program developer wants his web as usable as possible so naturally making UI intuitive, easy-to-use and consistent is the top priority. However, there are certain problems with making such UI, mainly because different web sites offer different things, thus it is very hard to make an ideal pattern for all sites. Another thing is to be original, to look not like a competition. That is why we need some methods that could tell us if we present our data to users in the right way.

Usability is a web design (or an application design in general) approach, by which we decide how difficult for a user is learning and accessing an application. Developers should take into consideration that users are very often already familiar with some GUI patterns. Having said that, using complicated, strange looking or even suspiciously appearing elements on a page might confuse or discourage potential customers. Let us mention a simple example, since the early days of the Internet expansion, users got used to hypertext links are underlined. This convention helps users to distinguish links from general text. It has been proven [4] that users actively seek underlined text when they want to get somewhere else (either another page of a website or another website) and they clearly know it is the area where they are supposed to click.

Usability is defined by International Organization for Standardization (ISO) as “The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use” [5]. There are five main components which define usability [6, 3]:

- Learnability—How easy is to perform basic tasks for new users?
- Efficiency—What time does it take for users to find what they came for?
- Memorability—How hard is it for users to repeatedly perform their tasks?
- Error rate—Errors made by users.
- Satisfaction—The comfort users feel when using the design.

What is quite disturbing and surprising at the same time, the users act differently and do things that a web developer would find hard to believe. To find where users act differently than expected it is needed to conduct testing sessions with a sample of representative potential users. These users are placed in front of a computer and asked several types of questions, depending on what we want to test. Then they are asked to find answers on questions we give them, while browsing a page. Users are preferably people who could be considered as ordinary users, that means no experts on a given field. What happens next is that the person is asked to work with the system naturally while developers are watching this person. They very often realized that the person is working with the UI in very strange ways. Their subject is clicking where he is not supposed to and all those important links and pictures remain unnoticed. With this gained knowledge designers usually change the design a bit and conduct the same experiment again [3, 1–2].

2.1 Scanning

Many think users read on the web. According to [2] the reality is that users seldom actually read all the text on a web page. Users tend to scan the document until they find what catches their eyes. Usually it depends on what the user’s goal on a page is. For web designers it is very important to be aware of this and to adapt the UI with taking into account such behaviour. Visitors just glance at headings, pictures and try to find anything familiar or interesting to click. So it is crucial to follow conventions visitors are used to. It is a great deal that visitors can use something easily, because it resembles something they have seen already before. Lots of web designers ignore this fact and try to reinvent a wheel all the time until the point the page has a very original look, which in theory could be a good marketing advantage, but eventually the page turns out to be very hard to use. As stated previously, when a visitor feels uncomfortable, it is only a matter of time when he runs out of his reservoir of goodwill [2] and leaves. Everybody once already learned how to use many things in their lives, for example how to read a newspaper, how to operate a printer, how to drive a car, how to withdraw money from an ATM and we could go on and on. For people it is impossible to disregard something they already learned, because they cannot go back to the state when they knew nothing [3].

2.2 *Hard to Use*

Unfortunately there is no way to safely design a usable website on first attempt, there are simply so many aspects that make it impossible to create a website based only on your best guess. Mostly, such website is developed by a group of people which brings even more obstacles to the deciding process. Scripters and programmers focus too much on the system, therefore do not really see what a user desires. They probably assume that the user adapts to their application just fine, since most of programmers find it natural and entertaining going through problems and solving them, because it is their world. After all they are hired and paid for their ability to solve technical problems. Another reason why it is so hard to develop a usable application is that audiences change and expand. Not so long time ago it was quite unusual for wide public to use computers and related technologies. This used to be a domain of enthusiasts, experts, scientists and academic employees. No wonder they did not complain, for their tasks it was still significantly better and more amusing than do all the stuff manually. If it was difficult to use, good, for early internet users it was a challenge to muddle through. Nowadays however, the audience is more casual and unforgiving, for them technology is not another hobby, it is a tool they want to successfully use [3, 1].

3 Human-Computer Interaction Standards from Usability View

Standards mean consistency and as already mentioned consistency is good for users, because they can transfer skill from one system to another which leads to decreasing the time needed to learn, thus enhancing their productivity and lowering costs for training them [3]. Human-Computer Interaction (HCI) Standards have been developed over the last 25 years most of them containing general principles for building applicable user interfaces [5]. There are several types of standards according to their origin and scope of use, these are mainly industry, national and international standards. Industry standards are probably the most popular thanks to operating system vendors who promote their solutions and indirectly influence other developers by encouraging them to use their standards. However, various industry standards are platform dependent therefore quite diverse. Those who switch from one standard to another should always learn differences when developing for another platform [3]. National and international standards authorities deal with these problems by creating their own generally accepted standards. The most important American standard-setting body is American National Standards Institute (ANSI) which mostly creates standards to be used within the country. When we need to use HCI standards that are accepted all around the world, we ought to obtain standards that have been developed under the patronage of the International Standards Organization (ISO) and the International Electrotechnical Commission

(IEC). Experts nominated by national standards committees work in Working Groups which is where the technical work occur [7]. It takes several years before a standard gets accepted which implies that industry standards will have the most importance for everyday use thanks to dynamic nature of the computer field. For example ISO/IEC standards for interface components like icons or cursor control have not been widely adopted because of large influence of industry standards [5].

Many assume that each standard has its precise specification, however, HCI standards tend to become out of date as technology evolves. For this reason most of the standards that ISO has created are not about precise specification, but rather about principles that we should employ in order to produce a usable interface. The most comprehensive is ISO 9241 standard family called Ergonomics of human-system interaction which contains a wide range of standards covering software ergonomics, dialogue principles, guidance on software accessibility and many more. We will take a look at some ISO 9241 standards related to usability. ISO 9241-11 is called the Guidance on Usability and the goal of this standard is to let users achieve their goals and measures their satisfaction by the product acceptance. It also measures user performance as the resources such as effort or time needed before the user's task is achieved. ISO 9241-10 called the Dialogue principles covers general principles that apply to designing dialogues between humans and information systems, tasks suitability, learning suitability, individualisation stability, and conformity with user expectations, error tolerance, self-descriptiveness and controllability. The Presentation of information is ISO 9241-12 standard that deals with representing information using screen layout, windows, alphanumeric and graphical codes. ISO 9241-13 (User guidance) issues recommendations for the evaluation and design guidance attributes of user interfaces such as on-line help, feedback or error management. ISO 9241-14 (Menu dialogues) issues recommendations for menu designing, menu structure, navigation, option selection and presentation. Finally ISO 9241-17 recommends how to design forms, their dialogues, structure, navigation and both output/input considerations. So these are the main international standards that specify how to deal with human-computer interaction issues.

Unfortunately there are risks and disadvantages involved when following standards. Having a product solely based on a standard can prevent further enhancements with regards to reduced flexibility and lesser motivation among developers, because they might feel that they do not share ownership of the user interface. Developers are often under the impression that following the standard automatically means they are designing a good user interface and might overlook serious design issues [3].

4 User Interface

When speaking about human-computer interaction, a user interface provides ways for communication between a human and a computer. Usually information is stored on various types of media, but mostly on hard disk drives of either personal computers or servers. Data are presented to users via user interface that describes what kind of data a user will see on a screen. In fact the user interface is where all communication between a human and a machine takes place [8, 3].

4.1 *GUI Design Versus Web Design*

In connection with web applications we mostly talk about a graphical user interface (GUI) or a web user interface (WUI). In a typical GUI application a developer has traditionally full control over the application. He has the means to prevent user entering forbidden parts as well as letting a user see what the developer wants. There are also no restrictions in GUI creation, depending on an operating system you have the option to use predefined system components or you can create and use your own [8]. It is hard to tell a user what he can or what he cannot do on our web page. He has much bigger freedom than when working with a desktop application. By using search engines such as Google he can directly jump on a page that has been never meant to be seen as first. We need to adapt and bring our users support for user-driven navigation. It is easier if the developer anticipates such behaviour.

4.2 *Guidelines and Recommendations for Building a Usable User Interface*

Everything begins when a decision about creating a new web site is made, at that point, people involved such as managers, designers, programmers etc. should take a great care and think about the target audience as the key, inseparable part of the system. Cooperating with potential users from the beginning saves us a lot of work, because problems we solve immediately will not propagate into later versions hence we will be able to focus on fixing other issues. In case a project has a very limited budget and there is no way to afford testing users, it is still important to keep in mind that a web site should serve people who use it. Designing for users is still the top priority and now we will introduce methods that will help us to craft user-friendly interfaces without involving any actual users. However, these methods work best if combined with user testing [3, 1, 2].

5 Conclusions

The purpose of this work was to introduce guidelines, practices and methods to increase users' satisfaction and making users more effective while interacting with our web pages. In this paper we learned about the need of building a usable web site by employing user-centered design principles. The authors believe that this paper clarified the evolution of internet from simple static websites to complex web applications in case of HCI and user interaction. Future research works will be targeted on the development of a software platform that will send user clicks to the server and allows us to go through these data in the web application for administrators [1, 4, 9]. Design and development of this framework along with concrete computation of websites evolution index is planned to be subject of further studies [2, 5, 7].

Acknowledgments This work and the contribution were supported by project "SP-2103-2015—Smart Solutions for Ubiquitous Computing Environments" Faculty of Informatics and Management, University of Hradec Kralove, Czech Republic.

References

1. Benikovsky, J., Brida, P., Machaj, J.: Proposal of user adaptive modular localization system for ubiquitous positioning. *Lect. Notes Comput. Sci.* **7197**, 391–400 (2012)
2. Penhaker, M., Darebnikova, M., Cerny, M.: Sensor network for measurement and analysis on medical devices quality control. *Commun. Comput. Inf. Sci.* **171**, 182–196 (2011)
3. Behan, M., Krejcar, O.: Modern smart device-based concept of sensoric networks. *EURASIP J. Wirel. Commun. Networking.* **2013**(1), 155 (2013). doi:[10.1186/1687-1499-2013-155](https://doi.org/10.1186/1687-1499-2013-155). ISSN 1687-1499
4. Hajovsky, R., Pies, M.: Complex measuring system for longtime monitoring and visualization of temperature and toxic gases concentration. *Elektron. ir Elektrotehnika* **122**(6), 129–132 (2012)
5. Jancikova, Z., Kostial, P., Bakosova, D., Ruziak, I., Frydrysek, K., Valicek, J., Farakasova, M., Puchky, R.: The study of electrical transport in rubber blends filled by single wall carbon nanotubes. *J. Nano Res.* No. **16**(21), 1–6 (2013)
6. Bartuskova, A., Krejcar, O., Kuca, K.: Evolutionary approach of general system theory applied on web applications analysis. ICOCOE 2014, Malacca, Malaysia, 20–21 May 2014. In: Advanced Computer and Communication Engineering Technology. Lecture Notes in Electrical Engineering, vol. 315, part 2, pp. 411–422 (2015)
7. Machacek, Z., Slaby, R., Hercik, R., Koziorek, J.: Advanced system for consumption meters with recognition of video camera signal. *Elektron. Ir Elektrotehnika* **18**(10), 57–60 (2012)
8. Behan, M., Krejcar, O.: Adaptive graphical user interface solution for modern user devices. *Lect. Notes Comput. Sci. (LNCS)* **6592**, 411–420 (2012)
9. Cimler, R., Matyska, J., Balik, L., Horalek, J., Sobeslav, V.: Security issues of mobile application using cloud computing. *Adv. Intell. Syst. Comput.* **334**, 347–357 (2015)

Design and Performance Evaluation of a VANET-Based Adaptive Overtaking Assistance System

Ihn-Han Bae and Jae-Kon Lee

Abstract One of the most risky maneuvers in roads with both directions in the same carriageway and only one lane for each direction is overtaking another vehicle. To overtake becomes particularly dangerous at night, with bad weather conditions and in curves, due to the diminution of visibility. This paper suggests a VANET-based Adaptive Overtaking Assistance System (VAOAS). The VAOAS not only considers both the driving skill of drivers and the length of overtaking/preceding vehicles, but also supports all strategies of overtaking using VANETs. The performance of VAOAS system is evaluated as the number of V2V messages, overtaking and passing sight distances, and overtaking success rate, depending on traffic densities of both directions through a simulation.

Keywords Driver assistance system · Lane change · Overtaking maneuver · Passing sight distance · Time to collision · VANET

1 Introduction

Cooperative vehicular systems are expected to improve traffic safety and efficiency through the real-time exchange of information between vehicles and infrastructure nodes [1]. To avoid potential road dangers, cooperative active safety applications for these systems are being designed. Typical these applications include overtaking assistance, lane change assistance, forward collision warning applications, etc. We

I.-H. Bae (✉)

School of IT Engineering, Catholic University of Daegu, Gyeongsan, Korea
e-mail: ihbae@cu.ac.kr

J.-K. Lee

School of Mechanical and Automotive Engineering, Catholic University of Daegu,
Gyeongsan, Korea
e-mail: leejk@cu.ac.kr

focus on overtaking assistance systems on based on vehicular ad hoc networks (VANETs).

Overtaking a vehicle is cutting in a road by overtaking a vehicle which is driving in the front. Many of road traffic accidents occur when a vehicle intends to overtake another under risky conditions. The fact that vehicles at very different velocities share the same lanes makes it harder to estimate the real velocity of a car which comes in opposite direction. Accident rates become even worse during night trips or in situations with hard rain or fog due to the diminution of visibility [2].

In this paper, we suggest a VAOAS, VANET-based adaptive overtaking assistance system. To support safely and effectively overtaking, the VAOAS not only considers both the driving skill of drivers and the length of overtaking/preceding vehicles, but also supports all strategies of overtaking using VANETs.

The remaining of this paper is organized as follows. Section 2 gives a brief description of related works for vehicular overtaking systems. Section 3 proposes a VANET-based adaptive overtaking system. Section 4 evaluates the performance of VAOAS through a simulation, and shows some simulation results. Finally, the paper is concluded and prospected in Sect. 5.

2 Related Works

A new approach that has recently emerged allows cooperation among vehicles by means of telecommunications. Communicating vehicles can detect, address, and thereby help to prevent hazardous traffic situations. Such a communication platform is provided by VANETs. VANETs provide communications among nearby vehicles and between vehicles and nearby roadside units (RSUs). To this end, special radios and sensors would be embedded within the car. The vehicle to vehicle (V2V) communication infrastructure assumes the presence of high bandwidth with low latency [3].

The research performed by [4, 5] studied and identified variation in overtaking behavior on bi-directional roads. Overtaking maneuver is classified in four categories as follows:

- Accelerative: The host vehicle follows a preceding vehicle and waits for a safe sufficient gap to perform an overtaking maneuver.
- Flying: The host vehicle speed is not adjusted to the preceding vehicle speed whereas host vehicle continues at its current speed during the overtaking maneuver.
- Piggy backing: The host vehicle follows another vehicle that overtakes a slower vehicle.
- Two plus (2+): The host vehicle performs the overtaking maneuver from two or more vehicles.

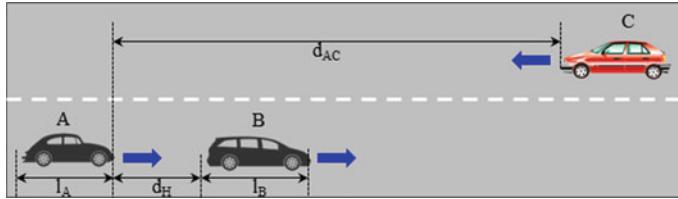


Fig. 1 Overtaking maneuver, initial situation

Figure 1 presents the initial situation preceding overtaking maneuver. The overtaking maneuver on bidirectional roads has been described by the following characteristics [4]:

- The duration of the total overtaking maneuver.
- How this duration correlates with Node *B*'s behavior and how the maneuver is performed.
- The distances between Node *A* and Node *B* prior to and after the maneuver is completed.
- The time left before Node *A* encounters the first oncoming vehicle.

Recently, there are several overtaking assistance systems designed to reduce road accidents and improve both overtaking maneuver and road safety. However, most of them do not consider the entire overtaking and maneuvering process. These systems mainly focus on vehicular lane changing process, which is regarded as one of the multiple phases in an overtaking maneuver. Besides, not all of these related works do overtaking in VANET environments. Vieira et al. [6] presented some important concepts for an application development to assist the driver in overtaking maneuver with prior knowledge of oncoming traffic, even when the road curvature has blind spots. Wang and Cartmell [7] presented a mathematical model that enables the determination of the safe passing sight distance and calculates the desired trajectory for overtaking on two-lane highways. Perez et al. [8] had been implemented a fuzzy decision system based on fuzzy logic able to execute an autonomous overtaking in two-way roads.

3 VANET-Based Adaptive Overtaking Assistance System

In this paper, we design a VAOAS for overtaking in case of bad weather or insufficient line of sight. In the design of VAOAS, we assume the following details:

- The on-board global positioning system (GPS) is equipped in vehicles to calculate the distance between vehicles and vehicle speed.

- Vehicles use a contention-free RPB-MACn mechanism [9] which combines the dedicated communication channel pair and the dedicated directional antenna associated to the relative position of vehicles.

The VAOAS provides drivers with assistance and warning for dangerous situations when a vehicle tries to overtake a slowing vehicle on a rural, 2-lane road. First, VAOAS exchanges information with the host vehicle, preceding vehicle and oncoming vehicle through V2V communication. Then, the VAOAS calculates a safe overtaking and a passing sight distances based on the location and speed of other vehicles, and provides a message, either warning about and approving overtaking.

The overtaking maneuver consists of the three phases as depicted in Fig. 2. In phase 1, the host vehicle A changes lanes, and the host vehicle A accelerates its speed to passing preceding vehicle B in phase 2. Finally, in phase 3, the host vehicle returns to its ongoing lane. In Fig. 2, symbols S_1 , S_2 denote the distance travelled in a forward moving direction by the passing vehicle A during the steps 1 and 2, respectively. X_1 are distance travelled by the preceding vehicle B .

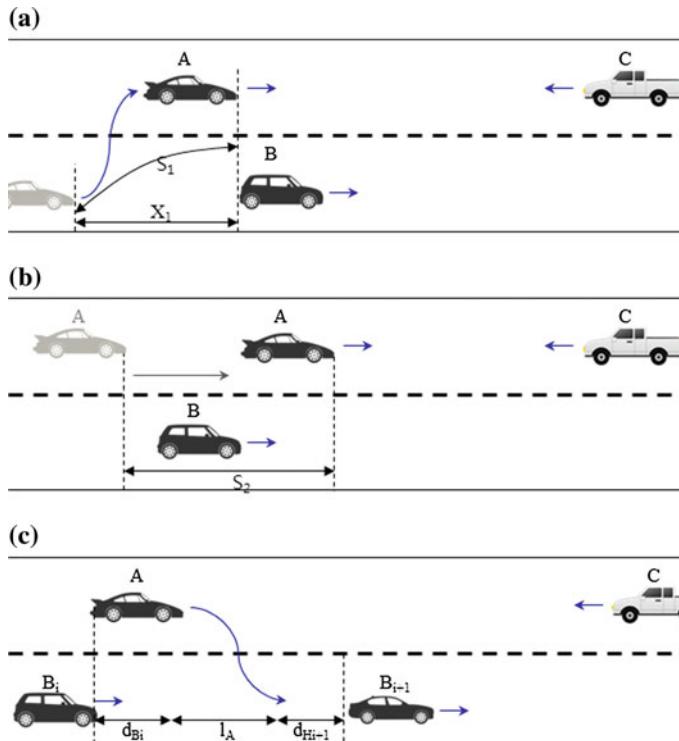


Fig. 2 The phases of overtaking maneuver. **a** Overtaking maneuver, phase 1, **b** overtaking maneuver, phase 2, **c** overtaking maneuver, phase 3_i

In both phases 1 and 3, the host vehicle A has to change lanes in order to overtake the preceding vehicle B and return to its ongoing lane. From Fig. 3, an approximation for the path length is given by (1) [7], where d_{lane} denotes the width of lane.

$$S_i = \frac{7X_i}{15} + \left(\left(\frac{8X_i}{15} \right)^2 + (d_{lane})^2 \right)^{\frac{1}{2}}, \quad \text{for } i = 1 \text{ or } i = 3. \quad (1)$$

The host vehicle A accelerates with a_A^{\max} from its initial speed v_A^{init} until its maximum speed v_A^{\max} is reached. The required distance S_A^{\max} and the consumed time T_A^{\max} that is travelled by the host A to reach its maximum speed are computed by (2) and (3), respectively.

$$S_A^{\max} = \frac{V_A^{\max 2} - V_A^{init 2}}{2a_A^{\max}}. \quad (2)$$

$$T_A^{\max} = \frac{V_A^{\max} - V_A^{init}}{a_A^{\max}}. \quad (3)$$

This paper designs our VAOAS on the assumption that S_A^{\max} is reached within overtaking phase 1. S_1 and X_1 are the distances travelled by the host vehicle A and the preceding vehicle in overtaking phase 1, respectively. These are computed by (1), (4), (5) and (6), where the headway clearance d_H is given by the distance travelled by preceding vehicle B over the safety period T_s .

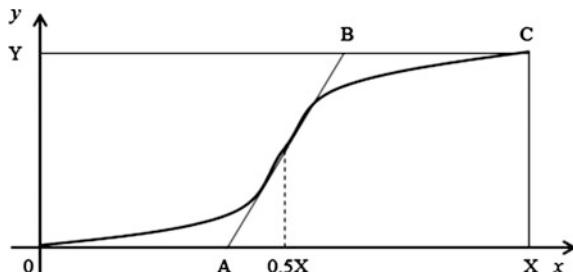
$$S_1 = v_A^{\max} T_1. \quad (4)$$

$$X_1 = v_{B_1} T_1 + d_{H_1}. \quad (5)$$

$$d_{H_1} = T_s v_{B_1}. \quad (6)$$

In overtaking phase 2 of the VAOAS, because the host vehicle A has already reached v_A^{\max} , both vehicle A and vehicle B travel with the constant speeds v_A^{\max} and v_B , respectively. Accordingly, S_2 and X_3 can be computed by (7) and (8).

Fig. 3 Illustration of approximation of S_i in case of transition between lanes



$$X_2 = v_{B_1} T_2. \quad (7)$$

$$T_2 = \frac{l_A + l_{B_1}}{v_A^{\max} - v_{B_1}}. \quad (8)$$

The phase 3 of the VAOAS also consists of three steps: 3_i, 3_p and 3_o. The step 3_i of the VAOAS computes the returning gap that the host vehicle A can return safely its ongoing lane. As shown in Fig. 2c, in the computation of the returning gap between preceding vehicle B_i and the front vehicle B_{i+1} of the preceding vehicle ($RG_{(i, i+1)}$), both the length of the host vehicle A and the driving skill of the driver of vehicle A are considered. And the backward clearance d_B is given by the distance that has to leave a headway distance towards the overtaken vehicle over the safety period T_s . In Eq. (9), s is the variable that represents the level of driving skill which is chosen by the driver himself.

$$RG_{(i, i+1)} = s \cdot (d_{H_{i+1}} + l_A + d_{B_i}), \quad (9)$$

$$s = \begin{cases} 0.85, & \text{experienced driver} \\ 1.0, & \text{normal driver} \\ 1.15, & \text{unexperienced driver} \end{cases}.$$

$$d_{B_i} = T_s' v_{B_i}. \quad (10)$$

If the real distance gap between vehicle B_i and vehicle B_{i+1} , $DGB_{(i, i+1)}$ is longer or equal to $RG_{(i, i+1)}$, step 3_o is performed for completion of overtaking maneuver. Otherwise, step 3_p is performed repeatedly until safe returning gap is found or the maximum overtaking distance is reached.

In step 3_p, the host vehicle A also passes through the preceding vehicle B_{i+1} because $DGB_{(i, i+1)}$ is shorter than $RG_{(i, i+1)}$. The distance and the time travelled by the vehicle B_{i+1} are computed by (11) and (12), respectively.

$$X_{3_i} = v_{B_{i+1}} T_{3_i}. \quad (11)$$

$$T_{3_i} = \frac{DGB_{(i, i+1)}}{v_A^{\max} - v_{B_{i+1}}}. \quad (12)$$

The distances travelled by the vehicle B_i in the step 3_o that the host vehicle A returns to its ongoing lane and finally completes overtaking are computed by (13), (13) and (14).

$$S_{3_i} = v_A^{\max} T_{3_i}. \quad (13)$$

$$X_{3_i} = v_{B_i} T_{3_i} + d_{B_i}. \quad (14)$$

In case that all linear equations of overtaking phases are solved, the passing sight distance PSD can be determined by using (15), where d_S is the safety distance between the overtaking vehicle A and the oncoming vehicle C on the opposite lane as a function on the time to collision (TTC).

$$PSD_{AC} = X_1 + X_2 + \sum_{i=1}^n X_{3_i} + d_S + v_C \left(T_1 + T_2 + \sum_{i=1}^n T_{3_i} \right), \quad (15)$$

$$d_S = TTC(v_A^{\max} + v_C).$$

Therefore, if the real distance between the host vehicle A and the oncoming vehicle C , d_{AC} is longer or equal to PSD_{AC} , VAOAS sends the overtaking possibility message that is you can overtake the number of n vehicles to the host vehicle

```

Algorithm VAOAS(s, lh, maxDov, maxDps)
% s is the driving skill level of a driver.
% lh is the length of overtaking vehicle.
% maxDov is the maximum distance of available overtaking.
% maxDps is the maximum distance for available passing sight.
Begin
  Compute X1, the traveling distance by preceding vehicle B1 in overtaking
  phase 1;
  Compute X2, the traveling distance by preceding vehicle B1 in overtaking
  phase 2;
  i=1; j=0; TDps=0;
  Ov_Success=false;
  Dov = X1 + X2;
  RG(1,2) = ds * (dH2 + lA + dB1);
  While ((DGB(i,i+1) < RG(i,i+1)) && (Dov ≤ maxDov))
    Compute X3p, the traveling distance by preceding vehicle B1 in overtaking
    step 3p;
    Dov = Dov + X3p;
    i=i+1;
    RG(i,i+1) = s * (dHi+1 + lA + dBi);
  End
  Compute X3o, the traveling distance by preceding vehicle B1 in overtaking step
  3o;
  Dov = Dov + X3o;
  While ((Dov ≤ maxDov) && (TDps ≤ maxDps))
    Compute PSDACj+1, passing sight distance;
    If (PSDACj+1 ≤ dACj+1)
      Ov_Success = true;
      Break;
    Else
      TDps = TDps + DGC(j,j+1);
      j=j+1;
    End
  End
  If (Ov_Success == true)
    Send overtaking possibility message (You can overtake the number of i ve-
    hicles) to vehicle A;
  Else
    Send alert message (No overtaking) to vehicle A;
  End
End

```

Fig. 4 Algorithm of the VAOAS

A. Otherwise, VAOAS sends the alert message that is no overtaking to the host vehicle A .

The structure of proposed VAOAS is shown in Fig. 4, where $DGC_{(j, j+1)}$ represents the real distance gap between vehicle C_i and vehicle C_{i+1} .

4 Performance Evaluation

We evaluate the performance of VAOAS in the MATLAB 7.0 [10]. The evaluated metrics for the performance of VAOAS are as follows: average returning gaps depending on driving skills and vehicle types, average distances of overtaking and passing sight depending on forwarding traffic densities, and the average number of required V2V messages and overtaking success rate depending on forwarding and opposing traffic densities. The parameters and values of the performance evaluation for VAOAS are shown in Table 1, where the datasets of traffic density and vehicle speed follow Gaussian distribution, and the evaluation results are derived from the overtaking attempts of 20 times a case except overtaking success rate.

Figure 5 shows average returning gaps by driving skills and vehicle types in case both the forwarding and the opposing traffic densities are 30 vehicles/km. The bigger vehicles and the more unskilled drivers, the longer returning gap is needed to overtake preceding vehicles. We verify that the returning gap of inexperienced drivers is needed about 1.34 times longer than experienced drivers.

Figure 6 shows average distances of overtaking and passing sight by forwarding traffic densities in case the opposing traffic density is 30 vehicles/km, the vehicle type is small, and the driving skill is normal. The overtaking and the passing sight distances are little linearly increased as forwarding traffic densities are increased. We verify that the overtaking distance is depended on forwarding traffic densities.

Figure 7 shows average number of required V2V messages for overtaking by forwarding and opposing traffic densities in case the vehicle type is medium and the

Table 1 Overtaking simulation parameters

Parameter	Value
Transmission range	500 m
v_A^{max}	90 km/h
T_s	1 s
$T_{s'}$	1.5 s
TTC	2 s
d_{lane}	3.6 m
s	0.85, 1.0, 1.15
$l_V, V = A, B$	4.5 m, 7 m, 12 m
maxDov	1 km
maxDps	2 km
Maximum driving speed	70 km/h
Maximum traffic density	70 vehicles/km

Fig. 5 Average returning gaps by driving skills and vehicle types

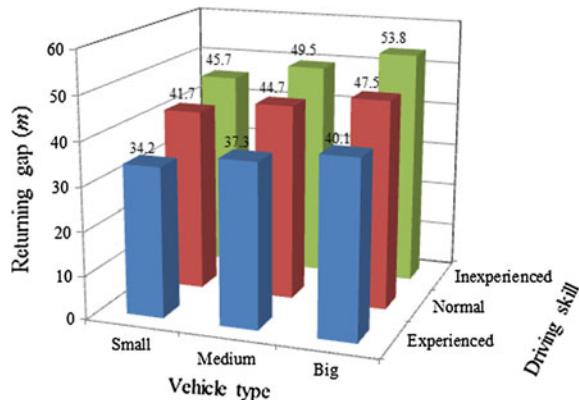


Fig. 6 Average distances of overtaking and passing sight by forwarding traffic densities

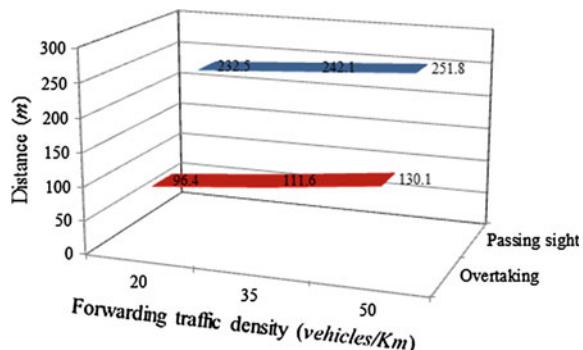
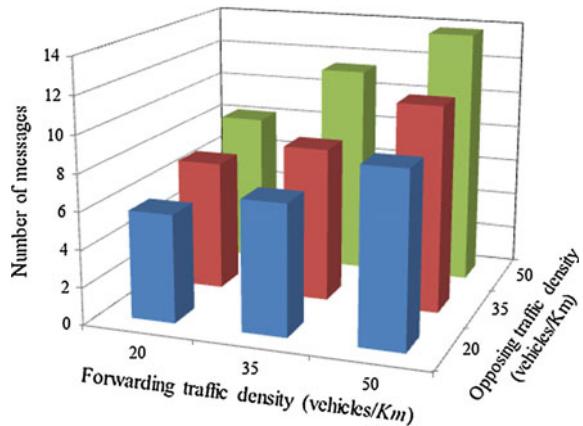


Fig. 7 Average number of V2V messages by forwarding and opposing traffic densities



driving skill is normal. The number of V2V messages is also increased as forwarding and opposing traffic densities are increased. We know that forwarding and opposing traffic densities almost equally affect in the increasing of numbers of V2V messages.

Fig. 8 Overtaking success rate by forwarding and opposing traffic densities

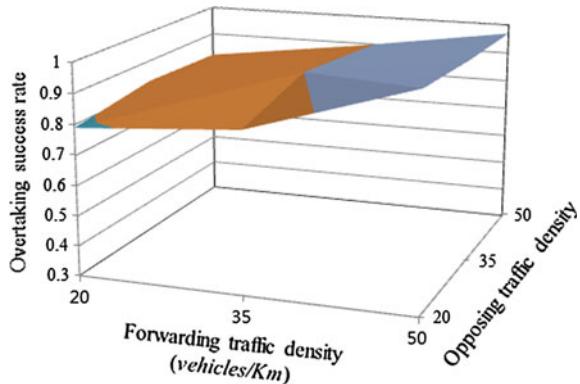


Figure 8 shows overtaking success rate by forwarding and opposing traffic densities in case the vehicle type is small and the driving skill is inexperienced. The result is derived from the overtaking attempts of 100 times. The overtaking success rate is high as forwarding traffic density is bit high and opposing traffic density is low. Also, we verify that the success of overtaking attempts is depended more on forwarding traffic density than opposing traffic density.

5 Conclusion

In this paper, we presented the VANET-based adaptive overtaking assistance system. The proposed VAOAS was designed by considering the driving skill of drivers and the length of vehicle types. Also, the performance of VAOAS was evaluated through a simulation study. From the results of the simulation, we knew that the success of overtaking attempts depended more on forwarding traffic density than opposing traffic density. For future work, applying the VAOAS on lane change collision avoidance systems and blind spot detection systems will be study.

References

1. Sepulcre, M., Gonzalvez, J.: Experimental evaluation of cooperative active safety applications based on V2V communication. In: Proceedings of the Ninth ACM International Workshop on Vehicular Inter-networking, Systems, and Applications, pp. 13–20. ACM Press, New York (2012)
2. Toledo-Moreo, R., Santa, J., Zamora-Izquierdo, M.: A cooperative overtaking assistance system. In: 3rd Workshop: Planning, Perception and Navigation for Intelligent Vehicles, pp. 50–56. IEEE Press, Piscataway (2009)
3. Chandrasekaran, G.: VANETs: the networking platform for future vehicular applications. Department of Computer Science, Rutgers University, New Jersey (2008)

4. Hegeman, G., Brookhuis, K.A., Hoogendoorn, S.P.: Observing overtaking maneuvers to design an overtaking assistance system. In: Proceedings of the 12th World Congress on Intelligent Transport Systems, p. 12. TRB, Washington (2006)
5. Kooten, V.V.: Advanced co-operative overtaking system using vehicular Ad-Hoc networks. Thesis for a Master of Science degree in Telematics, University of Twente (2011)
6. de Sousa Vieira, A.S., et al.: Driver assistance system towards overtaking in vehicular Ad Hoc networks. In: The Ninth Advanced International Conference on Telecommunications, pp. 100–107. IARIA, New York (2013)
7. Wang, Y., Cartmell, M.P.: New model for passing sight distance on two-lane highways. *J. Trans. Eng.* **124**, 536–545 (1998)
8. Perez, J., et al.: Longitudinal fuzzy control for autonomous overtaking. In: International Conference Mechatronics, pp. 188–193. IEEE Press, Piscataway (2011)
9. Chigan, C., Oberoi, V., Li, J.: RPB-MACn: a relative position based collision-free MAC nucleus for vehicular Ad Hoc networks. In: IEEE Global Telecommunications Conference, pp. 1–6. IEEE Press, Piscataway (2006)
10. Kay, M.G.: Basic concepts in matlab. http://www.ise.ncsu.edu/kay/Basic_Concepts_in_Matlab.pdf

Vulnerability Analysis on Smartphone Fingerprint Templates

Young-Hoo Jo, Sung-Yun Jeon, Jong-Hyuk Im and Mun-Kyu Lee

Abstract Currently, many smartphones are adopting fingerprint verification as a method to authenticate their users. Because fingerprint verification is not only used to unlock these smartphones but also used in financial applications such as online payment, it is crucial to secure the fingerprint verification mechanism for reliable services. In this paper, we identify a few vulnerabilities in one of the currently deployed smartphones equipped with fingerprint verification service by analyzing the service application. We demonstrate actual attacks via a proof-of-concept code that exploits these vulnerabilities. By these attacks, an attacker can extract fingerprint features by decoding a file containing them in encrypted form. We also suggest a few possible countermeasures against these attacks.

Keywords Biometrics · Fingerprint authentication · Fingerprint template · Smartphone · Vulnerability

This research was supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (grant number: 2014R1A1A2058514) and in part by the MSIP, Korea, under the ITRC support program (IITP-2015-H8501-15-1008) supervised by the IITP.

Y.-H. Jo · S.-Y. Jeon · J.-H. Im · M.-K. Lee (✉)

Department of Computer and Information Engineering, Incheon 402-751, Korea
e-mail: mklee@inha.ac.kr

Y.-H. Jo

e-mail: yallk@naver.com

S.-Y. Jeon

e-mail: roland.korea@gmail.com

J.-H. Im

e-mail: imjhyuk@gmail.com

1 Introduction

Recent advances in smartphone technologies enabled users to do various tasks using their smartphones. These tasks include critical ones, in particular, those dealing with private information and financial data. In this case, a reliable mechanism is required to verify the identity of a person who tries to use the device. To achieve this goal, fingerprint recognition is used for many smartphones, e.g., iPhone 5s, Galaxy S5, and VEGA Secret Note. It is used both for unlocking a smartphone and for activating other tasks, e.g., to approve transactions in financial applications [1].

Therefore, it is crucial to secure the fingerprint recognition service from possible threats. Unfortunately, however, some of the currently deployed solutions do not seem sufficiently safe. In this paper, we disclose the vulnerabilities in the fingerprint recognition service of VEGA Secret Note by analyzing the service application and demonstrate possible attacks against this service.¹ VEGA Secret Note is an Android-based smartphone with a Qualcomm Snapdragon CPU (Krait 400), 3 GB RAM, and a 5.9-in. IPS touch display. It is equipped with an FPC fingerprint sensor on its back.

Our attack is to extract a stored template from the nonvolatile memory and restore fingerprint minutiae points by decoding the template. By identifying and analyzing a fingerprint service application on the target device, we identified the location of the stored template. In addition, we discovered that the template was encrypted, but the same key and initial vector (IV) are hard-coded and are the same for all devices. This design results in a vulnerability that a malicious user may be successfully authenticated by overwriting a template by another template copied from his/her own device. Also, by analyzing the structure of the decrypted template file, we were able to restore all minutiae points constituting the fingerprint template. This implies that a carefully forged template according to the file structure also may pass the authentication test.

2 Biometric Verification Using Fingerprint Minutiae Format

Many devices that deal with fingerprints, including our target device, use the fingerprint minutia formats based on ISO/IEC 19794-2 [3] and ANSI INCITS 378 [4], where a minutia means a fingerprint feature point which constitutes the characteristic of that specific fingerprint. According to these standards, four main characteristics of minutiae are considered. These four characteristics are the x and

¹The VEGA series is one of the earliest smartphones with fingerprint recognition service, which is prior to recent popular ones such as iPhone 5s and Galaxy S5 [2]. The vulnerability was found on the device with Android 4.2.2 as of April, 2014. We reported this to the vendor. The vulnerability was independently addressed by the vendor through a patch.

y coordinates of the minutia on the original fingerprint image, the angle θ of the ridge corresponding to this minutia point, and ridge types (ridge bifurcation or ridge ending).

For biometric verification, a matcher compares the features extracted from the current sensor image with the stored template which is composed of multiple minutia points. The comparison is done by comparing the (x, y, θ) of each fingerprint minutia point in the stored template with those from the sensor. A matching score is increased whenever each point matches. If the score is larger than a pre-defined threshold, the user is permitted to access the target device.

The fingerprint recognition service application on a VEGA Secret Note supports three main functionalities; registration, verification, and deletion. First, to register a fingerprint, a user is asked to swipe a fingerprint over the fingerprint sensor. For high reliability, the user should swipe his/her fingerprint multiple times. A verification operation is usually used to unlock the smartphone. In this case, the user's task is just to scan his/her finger over the fingerprint sensor on the locked smartphone. The device recognizes the scanned fingerprint, and decides whether to permit this user's access based on the matching result. In addition, other applications may request the fingerprint recognition application to activate the verification functionality to verify if the person who attempts to use the application is the legitimate owner of this smartphone. It is also possible to reset the registered fingerprint by conducting a deletion operation. After unlocking the smartphone by scanning the correct fingerprint, a user may delete the stored fingerprint by scanning his/her fingerprint once again. If this fingerprint matches the registered one, it is deleted from the database.

3 Analysis of Vulnerabilities

3.1 Analysis of Fingerprint Service Mechanism

First of all, it is important to know where the binary code of the fingerprint recognition service application is located in flash memory. To find this location, we examined the list of running applications when the fingerprint service application is running. As a result, we identified application `com.pantech.app.fingerscan`. The next step was to extract the android package file of this application for analysis. After acquiring root user permission through rooting, we analyzed the package file. We remark that the root user permission is only required for the analysis of the application package file, but not all actual attacks require this permission.

Next, by analyzing the package file using a few tools such as *dex2jar 0.0.9.15* and *jd-gui 0.3.7*, we found out that this application uses JNI (Java Native Interface) to use the low-level functions implemented in a C++ library for fingerprint management, and we identified the path of this library loaded by the application. As a

result, we successfully extracted an Android framework file, `framework.odex` (and its corresponding `framework.jar`), and a shared library file, `libfpc1080_jni.so`. We used `framework.odex` to understand the interaction between `class.dex` and `libfpc1080_jni.so`. For the analysis of `framework.odex`, we used a disassembler, *baksmali 2.0.3*.

According to our analysis, the library file, `libfpc1080_jni.so`, which is an ARM-based dynamic linking library, contains the core routines for fingerprint authentication, in particular, fingerprint image processing. Therefore, in order to find attack vectors against fingerprint authentication service, we traced a source code decompiled from `libfpc1080_jni.so` line by line.

3.2 Extraction of Fingerprint Minutiae from an Encrypted Fingerprint Template

To enable the fingerprint verification services explained in Sect. 2 to work, a registered fingerprint should be stored in somewhere in nonvolatile memory storage for later use in fingerprint verification. If the fingerprint template is stored as a readable data file, we may try to analyze its structure and get minutiae points. Therefore, our first goal was to find the location of the stored template file. As explained in Sect. 3.1, we analyzed the fingerprint application, and identified a few essential functions in the library file `libfpc1080_jni.so` dedicated for fingerprint matching. In addition, by analyzing the code segment where the file that stores the fingerprint data is accessed, we found out that the name and path of this file was hard-coded irrespective of a specific device. The file name was `csfp.tpl`.

According to our analysis, the file `csfp.tpl` starts with a 48-byte header and the remainder is a main body containing fingerprint data. To be precise, the header contains a 4-byte identifier (signature) which stands for CSFP, a 12-byte field for file version, a 4-byte field for file size, a 16-byte MD5 checksum, and some additional data. This structure was found by tracing the header updating function in the library file.

The main body starting at 49th byte is not in a readable form, but it is encrypted. By analyzing the point in `libfpc1080_jni.so` when a user's fingerprint template data is stored into `csfp.tpl`, we could decode the encryption logic. According to our analysis, the encryption of a template is performed using the CBC mode of AES [5, 6] with a 256-bit key and a 128-bit IV. We also found out that most routines used for this encryption procedure resemble those of OpenSSL [7], which was helpful for our analysis. The key and IV are generated using very simple `for` statements without involving any randomness. As a result, the key and IV are fixed as `0x00010203...1F` and `0x00010203...0F`, respectively, and all devices use the same values.

We remark that the fact that all devices use the same key may be a critical issue. If the `csfp.tpl` in device A is copied and overwritten to another template file with the same name in another device B, then B will accept the fingerprint of the owner of A. This implies that virtually an attacker may bypass the fingerprint authentication. We will call this attack a *template replacement attack*. Below is a practical scenario where this attack may be a potential threat. If an attacker has a temporary access to B when B is temporarily unattended (e.g., the owner of B may go to a restroom leaving his/her smartphone on the desk), the attacker may inject his/her own `csfp.tpl` into B after rooting B. The attacker then may execute a financial transaction which is approved with fingerprint verification, e.g., a payment on PayPal [1]. If time is sufficient, the attacker may recover the original template file and unroot B, which prevents the owner of B from recognizing what happened with his/her device.

Because the encryption mechanism and its key information has been analyzed, our next step was to decrypt the encrypted template file and obtain the information about the original template, which we call a *template restoration attack*. We wrote a C code to perform AES decryption using the procedures provided by OpenSSL [7]. For decryption, we used the fixed IV and key revealed in the above analysis. This C code reads the `csfp.tpl` as input and outputs the decrypted data. An example result is shown in Fig. 1. We then could identify x and y coordinates by analyzing the code segments in `libfpc1080_jni.so` which refer to the coordinates. In addition, we could identify θ by analyzing the code segment for rotation. According to our analysis, θ was represented in a metric system where a full rotation is defined as an integer, 48. In Fig. 1, the highlighted region, which is 51 bytes long, stands for a single minutia point. The meaning of each field constituting this minutia is summarized in Table 1.

We were able to restore each and every minutia point in the original fingerprint template from the encrypted file, `csfp.tpl`, by using our Proof-of-Concept (PoC) code. This result is shown in Fig. 2. In the left-hand part of this figure, we enumerated restored points. These points can be graphically expressed using the convention in the literature, which is presented in the right-hand part of Fig. 2.

The restored template may be used to reconstruct the original fingerprint image [8, 9]. The reconstructed image can be used to forge a fake fingerprint to fool an image sensor. Moreover, this fake fingerprint may be used for other environments

```

00000000  11 10 10 10 10 10 10 10 11 10 10 10 10 10 0C 10 10 10  .....
00000010  0F 00 00 00 61 00 18 00 01 00 00 00 0E 00 CF F2  ....a.....iö
00000020  C3 FF C3 FC F3 FF CF FF C3 CF 03 FF F3 FF C3 FF  ÁyÁuóýÍyÁÍ.yóýÁy
00000030  33 0F 0F 00 C3 FF 03 FF CC F1 CF 74 03 FF 00 13  3...Áy.yinit.y...
00000040  00 00 00 00 00 00 00 15 00 00 00 FF FF FF FF FF  .......yyyyy
00000050  FF FF FF 01 00 00 00 14 00 00 00 FF FF FF FF FF  yyyy.....yyyyy
00000060  FF FF FF 02 00 00 00 13 00 00 00 FF FF FF FF FF  yyyy.....yyyyy

```

Fig. 1 Part of decrypted content of `csfp.tpl`

Table 1 Size and meaning of each field in a minutia in the decrypted `csfp.tpl`

Index	Size (B)	Meaning
0x00	4	Node (minutia) id
0x04	2	X coordinate
0x06	2	Y coordinate
0x08	4	Duplicate count (weight)
0x0c	2	θ ($0 \leq \theta < 48$)
0x0e	32	Additional information of node
0x2e	1	Minutia type
0x2f	4	(Distance to next minutia in bytes)/16

```

x: 95, y: 23, angle: 105, type: 0
x: 82, y: 35, angle: 277, type: 0
x: 57, y: 41, angle: 98, type: 0
x: 85, y: 71, angle: 277, type: 0
x: 85, y: 73, angle: 285, type: 1
x: 24, y: 74, angle: 247, type: 1
x: 69, y: 82, angle: 98, type: 0
x: 117, y: 90, angle: 112, type: 0
x: 76, y: 95, angle: 270, type: 1
x: 129, y: 97, angle: 300, type: 0
x: 64, y: 130, angle: 82, type: 0
x: 73, y: 135, angle: 270, type: 0
x: 97, y: 148, angle: 292, type: 0
x: 91, y: 149, angle: 105, type: 0
x: 92, y: 149, angle: 285, type: 1
x: 87, y: 158, angle: 105, type: 1
x: 21, y: 183, angle: 225, type: 1
x: 128, y: 187, angle: 315, type: 0
x: 82, y: 200, angle: 270, type: 0
x: 53, y: 202, angle: 225, type: 1
x: 120, y: 215, angle: 315, type: 0
x: 94, y: 241, angle: 277, type: 0
x: 92, y: 243, angle: 240, type: 1
x: 163, y: 247, angle: 157, type: 1
x: 80, y: 256, angle: 217, type: 1
x: 86, y: 257, angle: 232, type: 1

```



Fig. 2 Restored minutiae points and their graphical representation using x , y and θ

where a fingerprint is used for authentication. For example, an attacker may unlock the victim's doorlock with a fake fingerprint obtained through the above attack against the victim's smartphone.

4 Discussion and Conclusion

By analyzing a fingerprint recognition service application, we have identified a few attack vectors in the fingerprint recognition service of VEGA Secret Note and demonstrated actual attacks against this service.

Regarding the template replacement attack, it is not recommended using the same key and IV on all devices because an attacker can thwart the authentication test by overwriting the template file in the target device with that extracted from another device. Therefore, we suggest that a distinct key and a distinct IV should be

used for each device. However, this patch does not completely solve the problem, given that the key generation and encryption procedures are easily recognizable by reverse-engineering the library file. Therefore, it would be desirable to design a fingerprint recognition procedure so that an extracted template should be useless for other devices even after properly decrypted. To achieve this goal, cancelable fingerprints [10, 11] may be adopted. If this technique is adopted, even when an attacker extracts the transformed template, the information about the original minutiae points is not obtained. This approach also prevents a template restoration attack, though it cannot prevent a fake template synthesized according to the rules reverse-engineered from the target device. In addition, the logic to generate a key and an IV should be made hard to be analyzed by using obfuscation techniques.

We may consider a more essential solution including hardware-based isolation technologies such as ARM TrustZone [12]. These techniques might be adopted for secure storage of fingerprint data and isolated execution of fingerprint recognition service.

We finally remark that it would be a good research issue to verify whether other smartphones such as Galaxy series and iPhones equipped with fingerprint recognition service are vulnerable or not to the attacks described in this paper.

References

1. Paypal. <https://www.paypal-pages.com/samsunggalaxys5/us/index.html>
2. Pantech. <http://www.pantech.co.kr/en/board/reportBoardView.do?seq=5870&bbsID=report&ulcd=KO>
3. ISO/IEC International Standard 19794-2. Information Technology—biometric data interchange formats—part 2: finger minutiae data (2011)
4. ANSI INCITS 378-2009: American National Standard for Information Technology—finger minutiae format for data interchange (2009)
5. NIST special publication 800-38A, recommendation for block cipher modes of operation (2001)
6. NIST federal information processing standards publication 197. Advanced Encryption Standard (AES) (2001)
7. OpenSSL. <http://www.openssl.org/>
8. Cappelli, R., Maio, D., Lumini, A., Maltoni, D.: Fingerprint image reconstruction from standard templates. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(9), 1489–1503 (2007)
9. Feng, J., Jain, A.K.: Fingerprint reconstruction: from minutiae to phase. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(2), 209–223 (2011)
10. Ratha, N.K., Chikkerur, S., Connell, J.H., Bolle, R.M.: Generating cancelable fingerprint templates. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(4), 561–572 (2007)
11. Moon, D., Yoo, J.-H., Lee, M.-K.: Improved cancelable fingerprint templates using minutiae-based functional transform. *Secur. Commun. Networks* **7**(10), 1543–1551 (2014)
12. ARM. <http://www.arm.com/products/processors/technologies/trustzone/index.php>

Study of Measures for Detecting Abnormal Access by Establishing the Context Data-Based Security Policy in the BYOD Environment

Changmin Jo

Abstract With the trend of BYOD (Bring Your Own Device), i.e., using personal mobile devices for work, proliferating, one can easily find people doing their work anytime, anywhere. BYOD has the benefits of reducing business expense and increasing work productivity and efficiency. Nonetheless, the uncontrolled access of internal networks by the personal devices, for which enterprises have limitations in controlling, exposes the companies to security threats such as leak of confidential data and access by unauthorized users, yet there are inadequate countermeasures. Therefore, there is a need for a means of collecting the personalized context data of the user accessing the internal network and detecting and controlling abnormal user access by establishing a context-based policy. This paper presents measures for detecting abnormal access by collecting the context data according to the various devices and access environment and establishing the context data-based policy under the BYOD environment.

Keywords BYOD · Security policy · Context · Access control

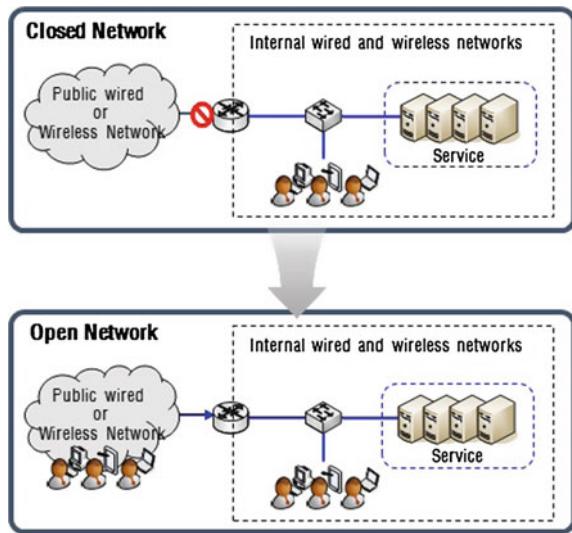
1 Introduction

The recent advancement of wireless communication technology and smart devices enabled users to utilize the Internet through smart devices anytime, anywhere. Similarly, the development of wired/wireless integrated environment has triggered the evolution of the business environment from a closed environment to an open one, with the concept of BYOD becoming the new business operating environment.

BYOD lets employees use their personal mobile devices such as lab top, tablet, and smartphone to access the company internal network to carry out business. It has

C. Jo (✉)
Korea Internet and Security Agency, Songpa-gu, Seoul, Korea
e-mail: jocm1309@kisa.or.kr

Fig. 1 Comparison of existing business environment and BYOD



the benefits of increased business efficiency, flexibility, and productivity as well as decreased business expenses (Fig. 1).

On the other hand, uncontrolled access of internal networks by the personal devices, for which enterprises have limitations in controlling, exposes the companies to security threats such as leak of confidential data and infection by malware using the personal device as the medium. Although personal devices with poor security measures can reportedly become easy targets of an attack through loss and burglary, and attacks through them occur often, there have been no root counter-measures to such.

Therefore, there is a need to identify the various access contexts such as the device accessing the company internal network, accessed network, and accessing user under the BYOD environment, and the policy must be established to detect and control abnormal access.

This paper proposes measures for collecting data concerning the status of the company internal network being accessed by various devices and access environments and establishing the corresponding policy based on the collected data to detect abnormal access.

2 Related Work and Research

Although measures such as NAC (Network Access Control) for controlling network access and MDM (Mobile Device Management) for controlling mobile devices are proposed to ensure security under the BYOD environment, there are

certain limitations. NAC, which performed control based on IP addresses, is not the basic solution even though it has added technologies like diversification of authentication of network access time point, management of harmful traffic after authentication, role-based control, and integrity check with increasingly popular cloud services [1–10].

Since the main purpose of NAC is to authenticate the users and control access, it has limitations in analyzing various contexts based on “who, when, where, what and how”. Moreover, since most NAC solutions provide the agent software to check the device security, users utilizing personal devices containing privacy information may object to it. The same is true for the MDM system, which monitors and controls a personal device with the company security program installed in it.

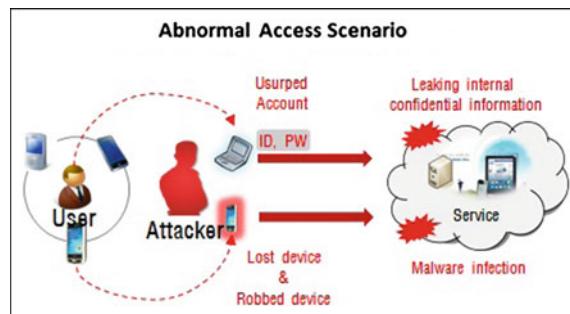
Therefore, an agentless solution is needed as one of the countermeasures to BYOD security threats. The agent software should be installed to control users only when needed, and users accessing the company internal network should be controlled through various context analyses.

3 Abnormal Access Scenario Under a BYOD Environment

3.1 *Information Leak and Malware Distribution Attempts Using Stolen Accounts*

Leaks of personal information using accounts stolen through server access often occur. According to the “Data Breach Report” issued by Verizon, there is 90 % probability of an account being stolen with 10 phishing attacks, and the malware installed by them is not detected by the virus vaccine and network monitoring tool. As such, the attacker can launch attacks such as leaking internal confidential information or distributing malware using the stolen accounts (Fig. 2).

Fig. 2 Abnormal access scenario



3.2 Access to Company Internal Network Through Lost and Stolen Personal Smart Devices

Most users neither lock their personal smart devices nor set them for automatic login to the company internal network. Note, however, that companies have difficulty enforcing functions such as locking and canceling the automatic login of personal devices given their limited control.

Under the situation, any lost or stolen personal smart device of a user will enable unauthorized access to the user's company network and expose it to threats such as leak of confidential information or infection by malware.

4 Detection of Abnormal Access by Establishing the Context Data-Based Policy

This paper proposes the method of detecting abnormal access to company internal networks by collecting the personalized context data of users who access the company internal network using agentless means—such as browser fingerprinting and network traffic mirroring—and subsequently detecting abnormal company internal network by establishing the policy based on the collected context data.

The proposed method builds the in-line based control system to control existing network access and develops the captive portal for user authentication and browser fingerprinting to collect context data related to access to company internal network by users. Based on the collected data, the policy is established, and the data of future user access to the company internal network are analyzed to judge abnormal access (Fig. 3).

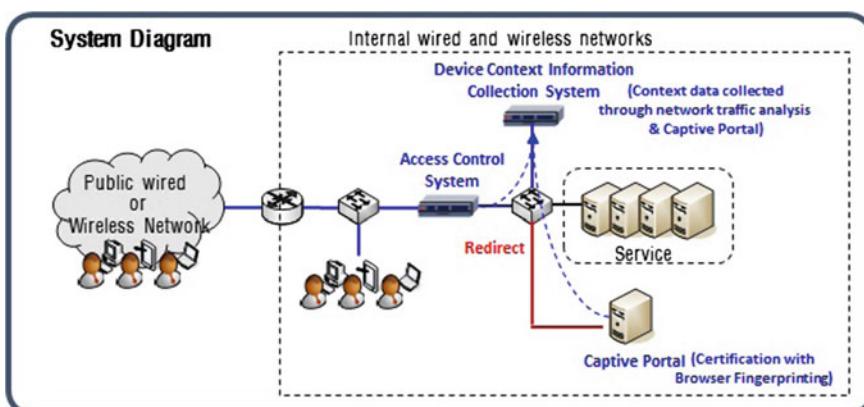


Fig. 3 Schematic diagram of the proposed method

4.1 BYOD Context Data Collection System

Under the BYOD environment, there is a need for network-based context data collection technology independent of the device type of users accessing the company internal network.

The “context data” are the data used to recognize the context by specifying all user behaviors related to accessing and using the company internal network into a data set.

This paper proposes a method of collecting and defining the context data—such as user data, device data, and network data—to identify the personal smart device and access environment in the BYOD environment.

User data include the user ID, privilege, and business travel, which can be collected with user input data during the captive portal authentication.

Device data include the device name, device type, OS, accessing browser, and screen size, which can be collected with browser fingerprinting.

Network data include the access period of the internal business network, user IP, accessing location, and accessed network, which can be collected through the analysis of HTTP traffic generated when the company internal network is accessed.

Based on the defined context data, contextual data such as “when, where, who, what, and how” can be collected for use in establishing the policy.

The proposed system identifies the traffic based on the IP and Session ID and redirects the user to the captive portal or company internal network based on the authorization. Disallowed traffic is redirected to the captive portal, and context data such as user data, device data, and network data are collected during the user authentication.

4.2 Measures to Detect Abnormal Traffic by Establishing the Context Data-Based Policy

The proposed method collects the access context data with the scenario of accessing the company service provided through the web and establishes the policy based on the collected data to detect abnormal access.

Under the BYOD environment, a user can use various smart devices for business; the main service access time, accessing device, and location can vary according to the nature of work of the user.

Therefore, the policy should be established through multidimensional combination based on the personalized context-aware data of “who, when, where, what and how” for each user to detect and control abnormal access (Fig. 4).

Context data are individually managed, and abnormal access to the company internal network can be detected with the policy established by analyzing the contextual data.

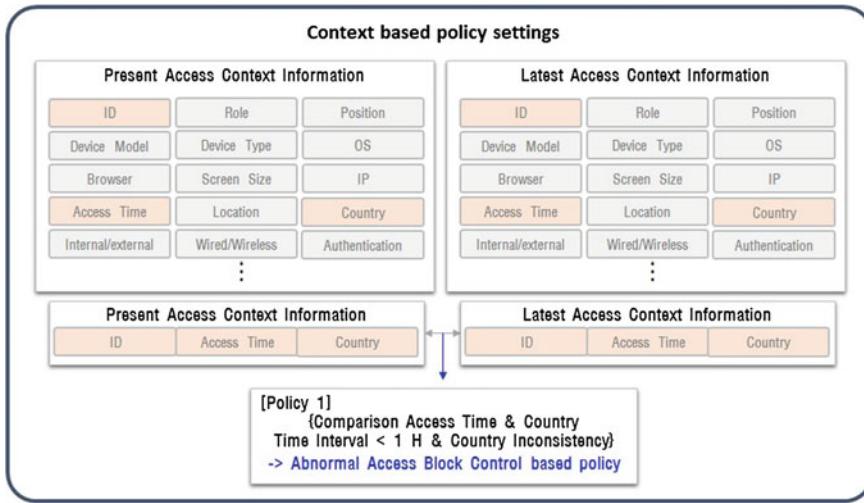


Fig. 4 Context-based policy setting

When a user accesses the company internal network, the stolen account can be identified by comparing the recent accessing location, accessing time, etc. The abnormality can be judged through the policy setting even with concurrent accesses by the same account (Fig. 5).

Moreover, the policy can control the process to force additional authentication when device data such as OS, device name, browser, and device type registered in the company internal network differ from the current device data. If two different accounts are attempting access with the same IP, the policy can block them since they are likely to have connected through a public network with vulnerable security.

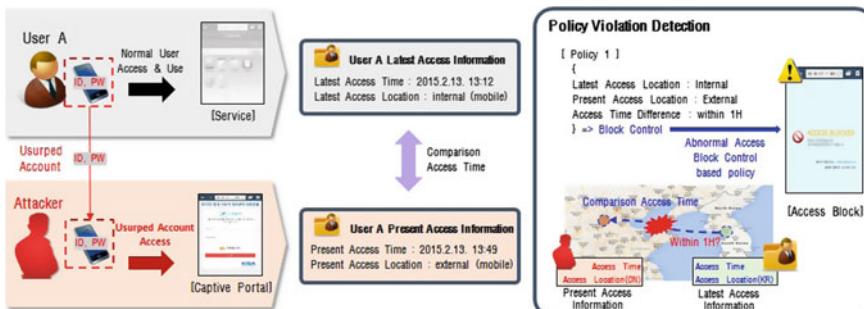


Fig. 5 Detection of abnormal access using stolen accounts

5 Conclusion

With the advancement of communication technologies and smart devices, BYOD has become an essential factor in the business environment. Under the BYOD environment, a user can conduct business using a personal smart device anytime, anywhere. It has the expected benefits of increased business efficiency, flexibility, and productivity as well as decreased business expenses. On the other hand, uncontrolled access of internal networks by the personal devices, for which enterprises have limitations in controlling, exposes the companies to security threats such as leak of confidential data and distribution of malware by malicious users. Therefore, policies must be established based on the context data of users accessing the company internal network to control access under the BYOD environment.

This paper proposes measures for detecting abnormal user access by establishing various policies based on the context-aware data of users accessing the company internal network. The policies can be established by the contextual combinations and correlation analysis of context data, and such will enable the detection of abnormal users to block abnormal access using stolen accounts or lost and stolen devices.

Planned future studies include the strengthening of policy-based detection of users abnormally accessing the company internal network through the contextual combination and correlation analysis of various context data. Other planned studies include measures for establishing the abnormal use detection policy using the context data analysis of service usage after logging into the company internal network.

Acknowledgments This work was supported by the IT R&D program of MSIP/KEIT (Ministry of Science, ICT and Future Planning/Korea Evaluation Institute Of Industrial Technology). (10045109, The Development of Context-Awareness based Dynamic Access Control Technology for BYOD, Smart work Environment.)

References

1. Miller, K.W.: BYOD: security and privacy considerations. *IT Prof.* **14**(5), 53–55 (2012)
2. Kang, D.: Context based smart access control on BYOD environments. *Inf. Secur. Appl. Lect. Notes Comput. Sci.* **8909**, 165–1762 (2015)
3. Singh, M., Patterh, M.S., Kim, T.-H.: A formal policy oriented access control model for secure enterprise network environment. *Int. J. Secur. Appl.* **3**(2), 1–14 (2009)
4. Eckersley, P.: How unique is your web browser?. In: 10th International Symposium, PETS 2010, Berlin, Germany, 21–23 July 2010
5. IDG Deep Dive: Guide to BYOD strategy. IDG Korea (2012)
6. Johnson, K.: Mobility/BYOD security survey. SANS Institute, Bethesda (2012)
7. Henderson, T.: How mobile device management works. *IT WORLD* (2011)
8. Kohno, T., Broido, A., Claffy, K.: Remote physical device fingerprinting. *IEEE Trans. Dependable Secure Comput.* **2**(2), 93–108 (2005)

9. Koh, E.: A study on security threats and dynamic access control technology for BYOD, smart-work environment. IMECS (2014)
10. Kim, T.: A study on context information collection for personal mobile device identification in BYOD and smart work environment. Mobility IoT (2014)

Incremental Multilevel Association Rule Mining of a Dynamic Database Under a Change of a Minimum Support Threshold

Nophadon Pumjun and Worapoj Kreesuradej

Abstract The proposed algorithm can efficiently deal with multilevel association rules mining of a dynamic database and a current support threshold can be different from the previous task. The experimental results show that the proposed algorithm has better performance than ML-T2 algorithm.

Keywords Multilevel association rules · Data mining · Association rules discovery · Incremental association rules discovery

1 Introduction

Recently, several association rule discovery algorithms have been proposed to deal with hierarchical data, set of data items that are related to each other by hierarchical relationships [1–3]. However, these works only deal with static databases while most databases in real-world applications are dynamic and are updated frequently.

On the other hand, MLUp algorithm [4] was proposed to maintain multilevel association rules without re-processing a whole database when new transactions are added and minimum support threshold is the same as a previous mining.

Typically, tasks of association rules mining in dynamic database are required several iterations with various minimum support thresholds until the desired result is obtained. However, the previous algorithms cannot deal with this situation. Therefore, this paper proposes a new algorithm to deal with a dynamic database mining under a change of a minimum support threshold.

N. Pumjun (✉) · W. Kreesuradej (✉)

Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang,
Bangkok, Thailand

e-mail: it3660422@gmail.com; it@gmail.com

W. Kreesuradej

e-mail: worapoj@it.kmitl.ac.th

2 Related Work

MLT-2 [1] is an extension of Apriori algorithm to deal with the multiple level association rules mining. However, when new transactions are inserted into a database, Apriori and ML-T2 cannot deal with this case efficiently because they have to re-process the entire database.

FUp [5] uses frequent itemsets obtained from previous mining in an original database compare with frequent itemsets discovered in an increment database. For each iteration, frequent itemsets in the incremental database which are not a frequent itemsets in the original database will be rescanned in the original database and updated theirs support count. MLUp [4] is based on FUp except that it is designed for a multilevel. However, both FUp and MLUp cannot handle the case of minimum support threshold is changed.

In this paper, we propose a new algorithm to deal with a dynamic database mining under a change of a minimum support threshold. The detail of our algorithm is proposed in Sect. 3.

3 Incremental Multilevel Association Rule Mining of a Dynamic Database Under a Change of a Minimum Support Threshold (IML-ARMCS)

A hierarchical data is a set of data items which are related to each other by hierarchical relationships [1]. The example of a hierarchical data is shown in Fig. 1.

Recently, the association rules mining have been applied to a hierarchical data. To mine a hierarchical data, the association rule algorithms use a hierarchy information encoded transaction database instead of an original database. The example of a hierarchy information encoded transaction data is shown in Fig. 2.

The proposed algorithm called IML-ARMCS can maintain a multiple-level association rule mining of a dynamic database under a change of a minimum support threshold. Like MLT-2 [1], IML-ARMCS starts at level 1, i.e. top level, to

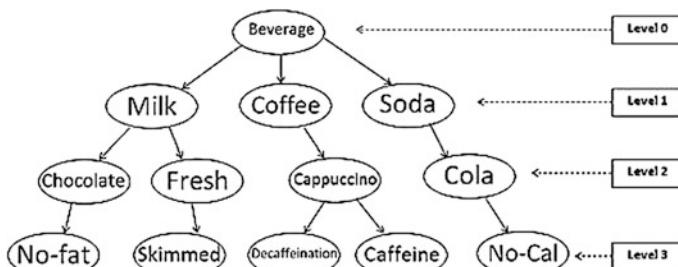


Fig. 1 The example of a hierarchical data

Transaction Table		Encoded Transaction Table (T[1])	
TID	Items	TID	Items
T ₁	{Skimmed Fresh Milk},{No-Cal Cola Soda}	T ₁	{122},{345}
T ₂	{Skimmed Fresh Milk},{Caffeine Cappuccino Coffee}	T ₂	{122},{234}
T ₃	{No-Fat Chocolate Milk},{Decaffeination Cappuccino Coffee},{Caffeine Cappuccino Coffee}	T ₃	{111},{233},{234}
T ₄	{No-Fat Chocolate Milk},{Decaffeination Cappuccino Coffee},{No-Cal Cola Soda}	T ₄	{111},{233},{345}
T ₅	{No-Fat Chocolate Milk},{Decaffeination Cappuccino Coffee}	T ₅	{111},{233}
T ₆	{No-Fat Chocolate Milk},{Caffeine Cappuccino Coffee},{No-Cal Cola Soda}	T ₆	{111},{234},{345}
T ₇	{Decaffeination Cappuccino Coffee},{No-Cal Cola Soda}	T ₇	{233},{345}

Fig. 2 A transaction data and an encoded transaction table

find the frequent 1-itemsets. Then, some transactions of the hierarchy information encoded transaction (T[1]) are pruned if they consist of only infrequent 1-itemsets. The pruned encoded transaction database is called T[2]. After that, T[2] is used as a hierarchy information encoded transaction database for the rest of IML-ARMCS algorithm. For each level of a hierarchical data, IML-ARMCS algorithm continues to search for the frequent k-itemsets in the same manner as Apriori algorithm.

In addition, IML-ARMCS algorithm adopts the principle of Candidate Pruning Threshold (CPT_m) [6] to deal with a change of a minimum support threshold. Candidate Pruning Threshold (CPT_m) is defined as Eq. 1.

$$CPT_m = (S'_m \times d) + D(S'_m - S_m) + 1 \quad (1)$$

where the notations of the equation are shown in Table 1.

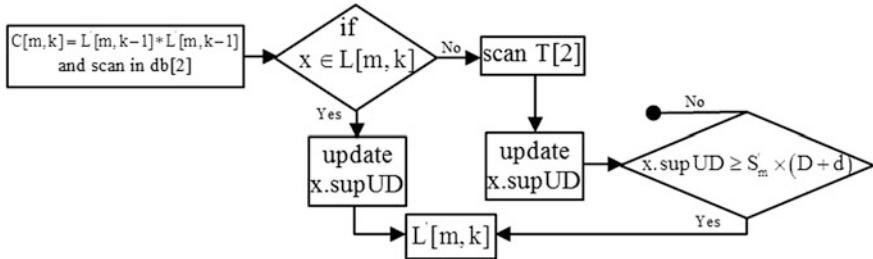
According to [6], a candidate itemset is a possible Winner itemset in an updated database (UD), i.e. itemsets that are not frequent in the original database but are frequent in the updated database, if and only if its support count in an incremental database (db) exceeds CPT_m threshold. Thus, the proposed algorithm uses CPT_m to classify the candidate itemsets that can be the Winner itemsets.

From Eq. 1, the value of CPT_m can be used to indicate whether a minimum support count of an updated database (min sup_{UD}_m) increased or decreased comparing to the minimum support count of an original database (min sup_{DB}_m). CPT_m can be classified into 3 cases: CPT_m ≤ 0, CPT_m > d and 0 < CPT_m ≤ d. Each case of CPT_m is required different techniques to maintain the incremental association rules. Thus, IML-ARMCS algorithm proceeds based on the 3 scenarios of CPT_m.

For CPT_m ≤ 0, the proposed algorithm is shown in Fig. 3. In this case, every candidate item x of an incremental database satisfies x.sup_{db} ≥ CPT_m condition. Then, the frequent itemsets in an original database are also frequent in an updated database. However, the original database is scanned to determine and update their support counts for the itemsets that are not previously frequent in the original database. In this case, i.e., CPT_m ≤ 0, The proposed algorithm consists of 2 main steps:

Table 1 Shows the meaning of notations used

Notation	Meaning
UD	Updated database
DB	Original database
db	Incremental database
T[1]	Original encoded transaction table
T[2]	Filtered T[1] by L'[1, 1]
db[2]	Incremental database after filtered
S'_m	Updated minimum support threshold at level m
S_m	Original minimum support threshold at level m
D	A number of transactions in original database
d	A number of transactions in incremental database
x.supUD	Support count of item x in updated database
x.supDB	Support count of item x in original database
x.supdb	Support count of item x in incremental database
min supUD _m	Minimum support count of updated database at level m
min supDB _m	Minimum support count of original database at level m
min supdb _m	Minimum support count of incremental database at level m
L[m, k]	Original frequent k-itemset at level m
L'[m, k]	Updated frequent k-itemset at level m
C[m,k]	Candidate k-itemset at level m
CPT _m	Candidate pruning threshold at level m

**Fig. 3** IML-ARMCS procedure in case $CPT_m \leq 0$

1. Scanning db[2] to find C[m,1] support count (x.supdb). Winners at level m have to be a descendant of L'[m-1, 1]. For $x \in L[m, 1]$, they are also L'[m, 1]. For $x \notin L[m, 1]$, the algorithm scans each item of C[m,1] in T[2] to update its support count (x.supUD) and any item which its support count (x.supUD) is greater than or equal to $\min supUD_m$ is considered L'[m, 1]. Therefore, all L'[m, 1] are updated. For searching L'[1, 1], db is used for mining then T[2] and db[2] can be derived by filtered out T[1] and db using L'[1, 1] as a filter.

2. At level $m \geq 1$ and $k > 1$, $C[m, k]$ from $L'[m, k - 1]$ are generated, as done by Apriori-gen [7]. Next, $db[2]$ is scanned to obtain $(x.sup db)$ of $C[m, k]$. Searching $L'[m, k]$ with the same as mention in above steps and repeat a similar process with all levels until there is no deeper level in a query or $L'[m, k]$ is empty.

For $CPT_m > d$, the proposed algorithm is shown in Fig. 4. Scanning only the incremental database (db) to update $x.supUD$ is enough to compute $L'[m, k]$ and no need to rescan the original database. In this case, i.e., $CPT_m > d$, the proposed algorithm consists of 2 main steps:

1. Since there is no $x \notin L[m, k]$ in $C[m, k]$ that could possibly be $L'[m, k]$, $db[2]$ is scanned to update $x.supdb$. Winners at level m have to be a descendant of $L'[m - 1, 1]$. Then, update $x.supUD$ and then pruned it by $min supUD_m$. Items that satisfy a condition are considered $L'[m, 1]$. $L'[1, 1]$ is derived from mining in $T[1]$ and db, then $L'[1, 1]$ is used as a filter for $T[2]$ and $db[2]$.
2. $C[m, k]$ from $L'[m, k - 1]$ are generated, as done by Apriori-gen. Next, scan $C[m, k]$ in $db[2]$ to count $x.supdb$. Then, update each $x.supUD$ and then pruned it by $minsup UD_m$ for $L'[m, k]$. Finally, repeat a similar process with all levels until there is no deeper level in a query or $L'[m, k]$ is empty.

For $0 < CPT_m \leq d$, according to Fig. 5, candidate itemsets are pruned depending on whether or not their $x.supdb$ exceed CPT_m . In this case, i.e., $0 < CPT_m \leq d$, the proposed algorithm consists of 2 main steps:

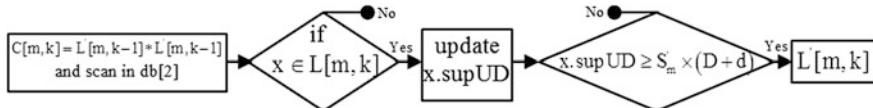


Fig. 4 IML-ARMCS procedure in case $CPT_m > d$

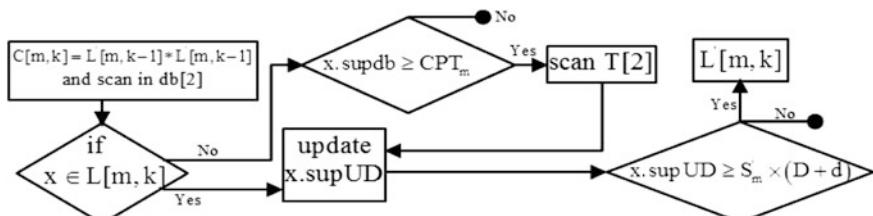


Fig. 5 IML-ARMCS procedure in case $0 < CPT_m \leq d$

1. We scan db[2] to update $x.supdb$. Winners at level m have to be a descendant of $L'[m - 1, 1]$. For each $x \in L[m, 1]$, update $x.supUD$ and then update $L'[m, 1]$. For $x \notin L[m, 1]$, if its $x.supdb$ is lower than CPT_m , it will be the loser. Then, scan the items of $C[m, 1]$ for $x \notin L[m, 1]$ that passed CPT_m in $T[2]$ to update $x.supUD$ and pruned it with $minsupUD_m$ for $L'[m, 1]$. $L'[1, 1]$ is derived from mining in $T[1]$ and db then $L'[1, 1]$ is used as a filter for $T[2]$ and db[2].
2. $C[m, k]$ from $L'[m, k - 1]$ are generated, as done by Apriori-gen. Next, scan $C[m, k]$ in db[2] for $x.supdb$. For $x \in L[m, k]$, update $x.supUD$ and then pruned it by $minsupUD_m$ for $L'[m, k]$. For $x \notin L[m, k]$, item x is a loser if its support count ($x.supdb$) is lower than CPT_m . On the other hand, any $x \notin L[m, k]$ that passed CPT_m will be scanned in $T[2]$ to update $x.supUD$. Items whose $x.supUD \geq minsupUD_m$ will be $L'[m, k]$. Finally, repeat a similar process with all levels until there is no deeper level in a query or $L'[m, k]$ is empty.

4 Experimental Results

The experiments are conducted on a synthetic dataset generated using the same technique in [7]. We used I4T10D90kd10k and I2T5D90kd10k synthetic transaction databases. The number of the original transactions is 90,000 transactions and the number of the incremental transactions is 10,000 transactions. Then, we converted each transaction to an encoded transaction table [1].

In the experiment, the frequent itemsets ($L[m, k]$) of an original database (DB) are obtained with the original minimum support thresholds (S_m) for I4T10 dataset: 0.65, 0.234, 0.082, 0.029 and for I2T5 dataset: 0.44, 0.148, 0.051 and 0.0182 in each level respectively. Then, we varied the different minimum support threshold (S'_m) into 12 cases which 6 cases are increased from S_m and the remaining 6 cases are decreased. For performance comparison, the execution time of IML-ARMCS is compared with ML-T2 that runs on the updated database 100,000 transactions, i.e., $UD = DB + db$.

According to Table 2, the proposed algorithm is definitely faster than ML-T2. This is the case because IML-ARMCS utilizes the frequent itemsets of the previous mining. Thus, there are smaller numbers of items that have to scan in the original database. Besides, CPT_m can prune out some candidate itemsets if they are not the potential frequent itemsets. Thus, the number of the candidate itemsets need to be scanned from the original database generated by the proposed algorithm is less than that generated by ML-T2 as shown in Table 2.

Table 2 Running time (s) of the maintaining frequent itemsets and the comparison number of candidate itemsets scanned in DB by ML-T2 and IML-ARMCS in DB = 90 k and db = 10 k

Threshold	I4T10 Dataset				I2T5 Dataset			
	Items rescanned in DB		Time		Items rescanned in DB		Time	
	IML-ARMCS	ML-T2	IML-ARMCS	ML-T2	IML-ARMCS	ML-T2	IML-ARMCS	ML-T2
S'1	26,679	27,797	1985.55	2054.5	11,369	11,587	519.44	524.92
S'2	14,166	15,511	1069.65	1076.58	9001	9540	421.77	428.48
S'3	9964	11,416	672.03	725.72	3056	3913	106.07	119.78
S'4	4747	6299	283.97	334.3	882	1999	22.60	38.48
S'5	2009	3742	124.42	179.42	230	1542	7.09	23.26
S'6	597	2465	33.55	91.17	76	1479	4.13	20.67
S'7	0	1893	8.08	63.59	0	1441	2.88	19.48
S'8	0	1799	7.57	59.71	0	1426	2.72	18.95
S'9	0	1491	4.74	36.11	0	1419	2.70	18.95
S'10	0	1132	2.29	15.39	0	1419	2.72	18.94
S'11	0	1117	2.25	14.74	0	1176	1.87	10.8
S'12	0	1109	2.23	14.6	0	1169	1.75	10.42

5 Discussion and Conclusions

In this paper, we have proposed IML-ARMCS algorithm which is applied in a multiple level concept and can correctly maintain a dynamic database when new transactions are inserted into the database with the different support threshold.

The experimental results show the execution time of the proposed algorithm is faster than ML-T2 because the proposed algorithm rescans fewer candidate itemsets than ML-T2 in the original database. Especially when the new support threshold is decreased from the original, the generated candidate itemsets are also definitely fewer than ML-T2. According to these 2 main reasons, our proposed algorithm is significantly faster than ML-T2.

However, the proposed algorithm cannot directly be applied when some transactions are deleted from an original database with the different support threshold. In the future, this proposed algorithm may be extended to cover the multilevel decremented database.

References

1. Han, J., Fu, Y.: Discovery of multiple-level association rules from large databases. In Proceedings of the 21th International Conference on Very Large Data Bases (VLDB'95), pp. 420–431. Morgan Kaufmann, San Francisco (1995)
2. Yinbo, W., Yong, L., Liya, D.: Mining multilevel association rules from primitive frequent itemsets. In Journal of Macau University of Science and Technology, vol. 3. Macau University of Science and Technology, Macau, China (2009)
3. Mittar, V., Ruchika, Y., Deepika, S.: Mining frequent patterns with counting inference at multiple levels. In International Journal of Computer Applications, vol. 3 (2010)
4. Cheung, D.W., Ng, V.T., Tam, B.W.: Maintenance of discovered knowledge : a case in multi-level association rules. In Proceedings of the Thirteenth National Conference on Artificial Intelligence. AAAI Press, Oregon (1996)
5. Cheung, D.W., Han, J., Ng, V.T., Wong, C.Y.: Maintenance of discovered association rules in large databases: an incremental update technique. In Proceedings of The twelfth International Conference on Data Engineering, pp. 106–114. IEEE Press, Louisiana (1996)
6. Bachtobji, M.A., Gouider, M.S.: Incremental maintenance of association rules under support threshold change. In Proceedings of the IADIS International Conference on Applied Computing. IADIS Press, San Sebastian (2006)
7. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. In Proceedings of the 20th International Conference on Very Large Data Bases (VLDB'94), pp. 487–499. Morgan Kaufmann, Santiago de Chile (1994)

Compressing Method of NetCDF Files Containing Clustered Sparse Matrix

Suntae Hwang, Gyuyeun Choi and Daeyoung Heo

Abstract Many results from scientific calculations are large-scale sparse matrices. The results of simulating volcanic ash diffusion are also a sparse matrix, and the values are clustered because of the characteristics of ash diffusion. The cost to store or transmit scientific data is usually high because such data are large scale. In this paper, we suggest a new storage format that is more efficient for storing clustered sparse matrix. Coordinate values are compressed more in the proposed format by saving only the first key value of consecutive non-zero elements and its length. The performance of the new format is the best among existing similar formats on ash diffusion simulation data, and the compressed size of the resulting file is comparable to a ZIP file. Because the new format can be applied partially to the data part of Network Common Data Form (NetCDF) files only, its header information is still readable directly from the compressed file, unlike zipped files.

Keywords NetCDF · FALL3D · Sparse matrix · Run-length · Compressing

1 Introduction

FALL3D is an application for simulating volcanic ash diffusion on the wind fields in the atmosphere, and the results are written in Network Common Data Form (NetCDF) format [1–3]. As shown in Fig. 1, the results are a clustered sparse matrix where most elements are zero, and the non-zero values are clustered consecutively. However, the size of NetCDF files is usually large because all elements are

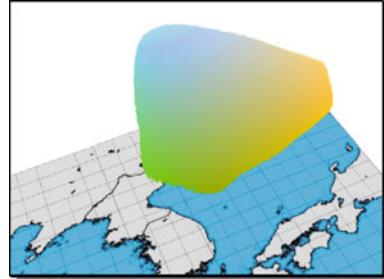
S. Hwang (✉) · G. Choi (✉) · D. Heo (✉)

Department of Computer Science, Kookmin University, Seoul, Korea
e-mail: sthwang@cs.kookmin.ac.kr

G. Choi
e-mail: gychoi@cs.kookmin.ac.kr

D. Heo
e-mail: dyheo@cs.kookmin.ac.kr

Fig. 1 Volcanic ash diffusion simulation result



recorded. For example, the data size of a four-dimension matrix whose dimensions are $190 \text{ (lon)} \times 190 \text{ (lat)} \times 67 \text{ (alt)} \times 25 \text{ (t)}$ is approximately 250 MB, and most of the space is misused with meaningless zero values. In general, the larger is the data size, the longer is the transmission time and the higher is the storage cost. When we compress the results into a ZIP format to minimize the space and transmission time overhead, the size is reduced to 40 MB, which is approximately 16 % of the original. However, it is not readable unless the file is unzipped.

In this paper, we suggest a way for reducing NetCDF file size by compressing the coordinate values of consecutive elements in the matrix without destroying the NetCDF format.

2 Background

2.1 Coordinate (COO) Format

In essence, key and values are stored in COO [4]. In this case, a coordinate value becomes a key, and nonzero values only are stored with the key.

Figure 2 illustrates an example that stores a two-dimensional matrix in a COO-structured matrix. Because the matrix is two-dimensional, it stores Row and Col values in each array. For example, 0 and 1, the key for A_{01} , are stored in the Row_Idx and Col_Idx arrays, respectively.

In this format, the size of the three arrays, Value, Row_Idx, and Col_Idx, are the same as the number of nonzero (nnz) values. In particular, because a dimension is increased by one, the number of Idx array is increased, and the space in which the

$$\begin{bmatrix} 0 & A_{01} & A_{02} & 0 \\ 0 & A_{11} & 0 & A_{13} \\ A_{20} & 0 & 0 & 0 \end{bmatrix} = \begin{array}{l} \text{Row_Idx} \quad [0 \quad 0 \quad 1 \quad 1 \quad 2] \\ \text{Col_Idx} \quad [1 \quad 2 \quad 1 \quad 3 \quad 0] \\ \text{Value} \quad [A_{01} \quad A_{02} \quad A_{11} \quad A_{13} \quad A_{20}] \end{array}$$

Fig. 2 COO format structure

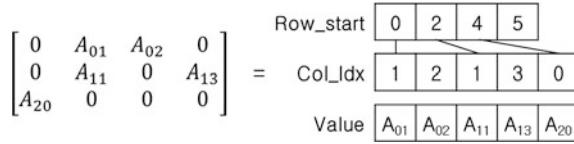


Fig. 3 CSR format structure

key is stored is also increased. Therefore, COO is not appropriate for our case because the FALL3D resulting matrices are four-dimensional systems.

2.2 Compressed Sparse Row (CSR) Format

CSR is also popular and compresses the Row index by storing only the starting point in Col_Idx for the column group of each row, instead of all the row keys.

As shown in Fig. 3, the three arrays in CSR format contain Row_start instead of Row_Idx, like COO, and Row_start contains Col_Idx pointers. For example, looking at A_{13} in Fig. 3, row key 1 becomes the index for Row_start, and Row_start: 1 contains the starting point of the column index, 2 in this case, because the column index starts from Col_Idx: 2. Column key 3 for A_{13} is stored in Col_Idx: 2 + 1 because A_{13} is the second nonzero value in the row. The last element of Row_start contains the nnz values. As shown in Fig. 3, the length of Row_start is $M + 1$ for an M by N matrix, and the sizes of Col_Idx and Value are the same as the nnz values.

2.3 Diagonal (DIA) Format

The DIA format aligns nonzeros based on the main diagonal [4]. Because nonzero values are limited to few numbers in matrix diagonals, DIA is recommended.

As shown in Fig. 4, the DIA format has Value and Idx arrays. Nonzeros are written in the Value array based on the main diagonal, and whose distance from the

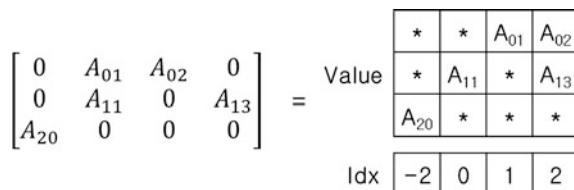


Fig. 4 DIA format structure

$\begin{bmatrix} 0 & A_{01} & A_{02} & 0 \\ 0 & A_{11} & 0 & A_{13} \\ A_{20} & 0 & 0 & 0 \end{bmatrix}$	=	<table border="1"> <tr> <td>1</td> <td>2</td> <td>A_{01}</td> <td>A_{02}</td> </tr> <tr> <td>1</td> <td>3</td> <td>A_{11}</td> <td>A_{13}</td> </tr> <tr> <td>0</td> <td>*</td> <td>A_{20}</td> <td>*</td> </tr> </table>	1	2	A_{01}	A_{02}	1	3	A_{11}	A_{13}	0	*	A_{20}	*
1	2	A_{01}	A_{02}											
1	3	A_{11}	A_{13}											
0	*	A_{20}	*											
		Col_Idx Value												

Fig. 5 ELL format structure

main diagonal is stored in the Index array. The symbol “*” in the Value array is used for padding. If the column values on the main diagonal are all zeros, nothing is written. In the main diagonal, Idx is zero. The positive numbers are placed on the right side, and negative numbers on the left side.

2.4 ELLPACK (ELL) Format

The ELL format arranges nonzeros to the left [4]. If the maximum value and average one of nnz per Row, it is efficient. Nonzero columns do not need to follow the particular pattern, which is more general than for DIA.

Figure 5 illustrates the nonzeros stored in the Value array sorted from the left, and the column coordinates stored in Col_Idx. For example, looking at the first row in Fig. 5, A_{01} and A_{02} , which are nonzeros, are stored in the first row of the Value array, and each column index (1 and 2) is written in the Col_Idx array. Next, the columns follow the same method describe above. The * mark is used for padding, similar to DIA.

3 Proposed Format

3.1 Design Considerations

ZIP compression does not allow information to be read without first decompressing such information because both the header and data are compressed. In this paper, we introduce a method for compressing the data part only.

The existing methods described in the previous section do not show good performance for storing FALL3D results because of the large index size of four-dimensional matrix values. DIA and ELL are especially not appropriate for our case because our nonzero elements are unstructured mesh. Therefore, our proposed format adopts the following two methods to manage the clustered sparse matrix characteristic of our resulting data.

3.1.1 Multi-dimensional Coordinate System is Converted to One-Dimensional System

If we convert a four-dimensional coordinate system to a one-dimensional system, we can reduce the key data size, but the scope of the coordinate system is reduced because of the limit of integer magnitude. For example, the scope is reduced to 2^{32} from $2^{(32 \times 4)}$ in a 32-bit machine. Briefly, we can represent a $256 \times 256 \times 256 \times 256$ coordinate system at most with a 32-bit integer. However, our data have a $200 \times 200 \times 100 \times 24$ coordinate system on average, and thus a 32-bit integer is sufficient for representing our coordinate system. When a larger scope is required, a 64-bit integer is used instead of a 32-bit integer.

3.1.2 Consecutive Indexes are Compressed by Run-Length Encoding [5]

In Run-length encoding, only one representative value and the total number of such values are stored if the same values follow each other consecutively. Here, Run-length means the number of consecutive same values. If we store only one value with Run-length, we can reduce data size by storing consecutive coordinate values.

In our case, we use Run-length encoding for compressing the index. If we have a starting point and the index length, all indexes can be restored if the subsequent indexes are consecutive. As shown in Fig. 1, FALL3D data are clustered, and therefore, there are many consecutive values, although the values are not the same.

3.2 Compressed Run-Length (CRL)

CRL is the proposed format to store a sparse matrix using Run-length. In this structure, all nonzero values are stored, but only the first index is stored with the consecutive length.

Figure 6 illustrates the structure of the CRL format. The first element of each run is stored on the *Idx* array, the Run-length value is stored on the *Len* array, and the nonzeros are stored in *Value* array. As shown in Fig. 6, the indexes for the nonzero elements are $A_{01} = 1$, $A_{02} = 2$, $A_{11} = 5$, $A_{13} = 7$, and $A_{20} = 8$. A_{01} and A_{02} are adjacent to each other, and thus the A_{01} index, 1, is stored in the *Idx* array, not the

	Idx	1	5	7		
	Len	2	1	2		
	Value	A_{01}	A_{02}	A_{11}	A_{13}	A_{20}

Fig. 6 CRL format structure

A_{02} index. The Run-length value 2 is stored in the Len array. For both A_{13} and A_{20} , the process is the same. Because A_{11} is not serialized, the index value 5 is stored in the Idx array and the Run-length value 1 is stored on the Len array. Consequently, the indexes for A_{01} , A_{11} , and A_{13} are stored on the Idx array, and the Run-length values 2, 1, and 2 are stored on the Len array.

4 Experiment Results

4.1 Benchmarks Description

We selected various sparse matrices to estimate CRL performance. Table 1 lists the 26 sparse matrices used in the test. Data called “fall3d1” and “fall3d2” are generated by extracting two-dimensional surface areas from FALL3D results, and “random1” includes random nonzero values. The remaining 23 matrices are selected from the data set used widely to evaluate performance in sparse matrix research, such as limited factor mesh, macroeconomics model, protein data, circuit simulation, web access, and combination problems [6–12]. They can be downloaded from the Matrix collection of the University of Florida [13].

4.2 Performance Estimation

The 26 matrices are stored in four formats, COO, CSR, DIA, and ELL, which are introduced in an earlier section, and in the proposed CRL format. For CRL, the index is stored in 8 bytes because the existing 23 matrices require a much larger scope than for our case ($200 \times 200 \times 100 \times 24$). After storing the 26 matrices set in each of the five formats, the size required to store one nonzero value is calculated by dividing the stored size by the number of nonzero values in the data. As indicated in Table 2, the best value (i.e., the smallest matrix unit size in bytes) for each matrix is shown in bold font.

For example, COO uses approximately 12 bytes per nonzero value, and CSR uses 9 bytes on average. DIA and ELL result in good performance on special cases, but show the worst performance when the matrix characteristics do not match well with their own target matrix. CRL shows the best performance for the target matrix: 4.90 bytes and 4.78 bytes for “fall3d1” and “fall3d2,” respectively, and the average value is also sufficiently acceptable, as shown in Table 2.

As the average Run-length of a matrix becomes longer, the performance of the proposed CRL format improves. For example, the CRL unit size is 8.02 for “eu-2005” when its average Run-length is 2.99, as shown in Fig. 7. To the right of “eu-2005,” the graph shows average Run-lengths longer than 2.99 and unit sizes lower than 8.0. Therefore, CRL shows good performance over three consecutive indexes.

Table 1 Benchmarks sparse matrix collection

Name	Plot	Description	Dimensions	nnz	nnz/Row	Domain
fall3d1		Fall3D	190 × 190	4.6 K	24.12	2D/3D
fall3d2		Fall3D	191 × 191	6.1 K	32.19	2D/3D
random1		Sample	190 × 190	3.6 K	19.00	Unstructural
2cubes_sphere		FEM	101 K × 101 K	1.6 M	16.23	Electro-magnetics
cage12		DNA electrophoresis	130 K × 130 K	2.0 M	15.61	Directed weighted graph
Cant		FEM/Cantilever	62 K × 62 K	4.0 M	64.17	2D/3D
circuit5M_dc		Large circuit	3.5 M × 3.5M	14.8 M	4.22	Circuit simulation
conf6_0-8 × 8-30		QCD	49 K × 49 K	1.9 M	39.00	Theoretical/quantum chemistry
Consphe		FEM/Spheres	83 K × 83 K	6.0 M	72.13	2D/3D
cop20k_A		FEM/Accelerator	121 K × 121 K	2.6 M	21.65	2D/3D
eu-2005		Small web crawl	862 K × 862 K	19.0 M	22.30	Directed graph
Hood		INDEED	220 K × 220 K	9.9 M	44.87	Structural
in-2004		Small web crawl	1.4 M × 1.4 M	16.9 M	12.23	Directed graph

(continued)

Table 1 (continued)

Name	Plot	Description	Dimensions	nnz	nnz/Row	Domain
m133-b3		Simplicial complexes	200 K × 200 K	0.8 M	4.00	Combinatorial
mac_econ_fwd500		Economics	206 K × 206 K	1.3 M	6.17	Economic
majorbasis		MCP	160 K × 160 K	1.7 M	10.94	Optimization
mc2depi		Epidemiology	525 K × 525 K	2.1 M	3.99	2D/3D
mono_500 Hz		3D Vibro	169 K × 169 K	5.0 M	29.71	Acoustics
Offshore		3D FEM	260 K × 260 K	4.2 M	16.33	Electromagnetics
pdb1HYS		Protein	36 K × 36 K	4.3 M	119.31	Weighted undirected graph
pwtk		Wind tunnel	218 K × 218 K	11.5 M	52.88	Structural
rma10		3D CFD	47 K × 47 K	2.3 M	49.73	Computational fluid dynamics
scircuit		Steve hamm	171 K × 171 K	0.9 M	5.61	Circuit simulation
shipsec1		DNV-Ex 4	141 K × 141 K	3.5 M	25.33	Structural
tp-6		Linear programming	143 K × 1 M	11.5 M	80.82	Linear programming
webbase-1 M		Web base	1 M × 1 M	3.1 M	3.11	Weighted directed graph

Table 2 Matrix unit size on storage format

Matrix	Bytes per nonzero entry				
	COO	CSR	DIA	ELL	CRL
fall3d1	12.74	8.90	20.41	34.24	4.90
fall3d2	12.55	8.67	18.91	16.95	4.78
random1	12.94	9.15	41.15	14.83	15.75
2cubes_sphere	12.00	8.47	17 K	22.29	15.11
cage12	12.00	8.26	19 K	16.92	14.56
cant	12.00	8.12	6.14	9.82	6.55
circuit5M_dc	12.00	8.73	1.58 M	39.65	10.71
conf6_0-8 × 8-30	12.00	8.10	22.36	8.00	7.35
consph	12.00	8.11	738.36	14.44	5.87
cop20k_A	12.00	8.36	39.3 K	17.09	11.04
eu-2005	12.00	8.18	133.7 K	2.5 K	8.02
hood	12.00	8.16	28.4 K	16.38	5.75
in-2004	12.00	8.33	67.3 K	5 K	6.96
m133-b3	12.00	9.00	196 K	8.00	15.90
mac_econ_fwd500	12.00	8.65	331.47	57.08	14.94
majorbasis	12.00	8.37	7.31	8.05	9.48
mc2depi	12.00	9.00	770.13	8.01	13.01
mono_500 Hz	12.00	8.14	22.7 K	193.49	12.05
offshore	12.00	8.46	47.7 K	22.16	15.58
pdb1HYS	12.00	8.07	850.46	24.47	5.09
pwtk	12.00	8.15	1.4 K	52.95	4.94
rma10	12.00	8.08	1.2 K	22.89	5.68
scircuit	12.00	8.72	93 K	503.58	11.18
shipsec1	12.00	8.14	742.15	23.80	5.83
tp-6	12.00	8.05	50 K	95.1 K	12.23
webbase-1 M	12.00	9.29	719.3 K	12.1 K	14.11
Average	12.09	8.45	116.5 K	4.4 K	9.90

4.3 Performance Comparison for FALL3D Results

By evaluating various sparse matrices in many formats, we found that the proposed CRL results in good performance for the clustered sparse matrix.

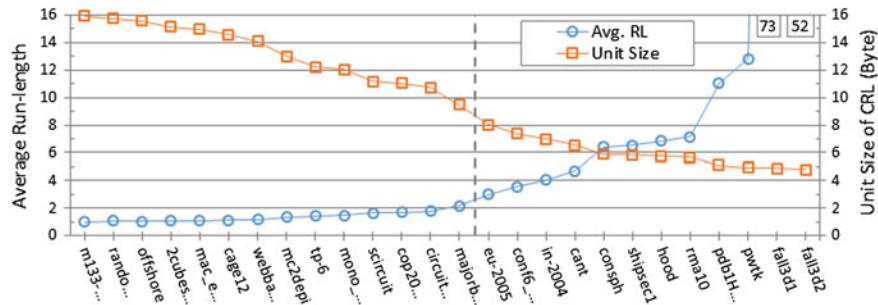


Fig. 7 CRL unit size on average run-length value

Table 3 Fall3D result files to be tested

Matrix	Dimension	nnz (K)	nnz rate (%)	File size
A	190 × 190 × 67 × 25	7.2	11.9	236 MB
B		8.3	13.8	
C		8.8	14.5	
D		9.9	16.4	
E		10.3	17.0	
F		10.9	18.0	
G		11.3	18.6	
H		11.8	19.4	
I		11.8	19.6	
J		12.0	19.9	
K		12.5	20.7	
L		12.8	21.2	
M		13.5	22.4	
N		13.8	22.9	
O		14.2	23.4	
P		16.3	27.0	
Q		16.5	27.2	
R		21.8	36.0	
S		24.6	40.6	
T		26.1	43.2	
U		26.7	44.2	
V		26.9	44.4	

Now, we evaluate the matrix compression rate in various formats using the FALL3D results only. Because DIA and ELL are not appropriate for the FALL3D results, they are excluded, and the ZIP compression method is used instead.

Table 4 Unit size on storage format

Data structure type	Bytes per nonzero entry
COO	16.00
CSR	8.10
ZIP	3.86
CRL	4.12

For testing, 22 FALL3D results are used, as indicated in Table 3. All the results are of the same matrix size, $190 \times 190 \times 67 \times 25$, and the file size is also the same, approximately 236 MB, when stored in the original NetCDF format. However, the number of nonzero values is different for each case.

For the COO, CSR, and CRL formats, the index and data are stored equally in 4 bytes. For the ZIP format, the original NetCDF file is compressed in standard mode. CRL shows almost equal performance as ZIP compression, and it is better than the other two formats, COO and CSR, for our 22 FALL3D data.

The unit size of each storage format is indicated in Table 4. The ZIP format uses 3.86 bytes on average to record one value. The CRL format results in better performance than the other storage formats at 4.12 bytes on average.

5 Conclusion

In this paper, a new storage format, CRL, was proposed for reducing the size of clustered sparse matrices, such as the FALL3D results. CRL was evaluated and compared with various existing formats, and showed very good performance in terms of unit size, i.e., the number of bytes required to store nonzero values. After compressing the results using CRL, we can store and manage them more easily, and transmit them faster via a network. Furthermore, the existing NetCDF library can still be used because the header part is not destroyed because the CRL is used only for the data part of NetCDF format.

References

1. Folch, A., Costa, A., Macedonio, G.: FALL3D: a computational model for transport and deposition of volcanic ash. *J. Comput. Geosci.* **35**(6), 1334–1342 (2009)
2. Unidata Network Common Data Form (NetCDF). <http://www.unidata.ucar.edu/software/netcdf/>
3. Rew, R., Davis, G.: NetCDF: an interface for scientific data access. *J. Comput. Graphics Appl.* **10**(4), 76–82 (1990)
4. Saad, Y.: Iterative methods for sparse linear systems. Siam, 92–94 (2003)
5. Pountain, D.: Run-length encoding. *J. Byte* **12**(6), 317–319 (1987)

6. Williams, S., Oliker, L., Vuduc, R., Shalf, J., Yelick, K., Demmel, J.: Optimization of sparse matrix-vector multiplication on emerging multicore platforms. *J. Parallel Comput.* **35**(3), 178–194 (2009)
7. Bell, N., Garland, M.: Implementing sparse matrix-vector multiplication on throughput-oriented processors. In: *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, pp. 18. ACM (2009)
8. Su, B.Y., Keutzer, K.: clSpMV: A cross-platform OpenCL SpMV framework on GPUs. In: *Proceedings of the 26th ACM international conference on Supercomputing*, pp. 353–364. ACM (2012)
9. Feng, X., Jin, H., Zheng, R., Hu, K., Zeng, J., Shao, Z.: Optimization of sparse matrix-vector multiplication with variant CSR on GPUs. In: *17th International Conference on Parallel and Distributed Systems (ICPADS)*, pp. 165–172. IEEE (2011)
10. Koza, Z., Matyka, M., Szkoda, S., Miroslaw, L.: Compressed multiple-row storage format. *CoRR*, (2012)
11. Ortega, G., Vazquez, F., García, I., Garzon, E.M.: Fastspmm: an efficient library for sparse matrix matrix product on gpus. *J. Comput.* **57**(7), 968–979 (2014)
12. Vazquez, F., Ortega, G., Fernandez, J.J., García, I., Garzon, E.M.: Fast sparse matrix matrix product based on ELLR-T and gpu computing. In: *2012 IEEE 10th International Symposium on Parallel and Distributed Processing with Applications (ISPA)*, pp. 669–674. IEEE, (2012)
13. U. of Florida. Sparse matrix collection, <http://www.cise.ufl.edu/research/sparse/matrices/>

An Approach to Discovering Weighted Service Pattern

Jeong Hee Hwang and Mi Sug Gu

Abstract It is important to provide suitable services according to user context. The previous researches are focused on service frequent pattern mining based on service history provided to user. However, service value is changing because user context varies with time. Therefore it is necessary to detect service patterns considering weight of service value. In this paper we propose a mining method by service weight based on service ontology. The method uses a weight of service significance by spatio-temporal context of user and finds out the valuable service patterns. And then the searched patterns which make a combination with existing service rule is provided to user.

Keywords Mining algorithm · Association rule · Pattern mining

1 Introduction

The previous service frequent pattern mining is performed without considering the service value, but it is common that service value is changing over time in real-world. That is, despite the same service, service value is different according to the time. The weighted pattern mining is a method to find a high weighted frequent pattern in consideration of each service weight. Even if it is the same service, service weight depends on the time or season. Therefore our proposed mining

Funding for this paper was provided by Namseoul university.

J.H. Hwang (✉)

Department of Computer Science, Namseoul University, Cheonan, South Korea

e-mail: jhhwang@nsu.ac.kr

M.S. Gu

Database/BioInformatics Laboratory, Chungbuk National University, Cheongju, South Korea

e-mail: gumisug@dblab.chungbuk.ac.kr

method is performed to extract frequent service pattern of the different weight of service in terms of user context and apply the multilevel association rule mining algorithm [1, 2] that can find out a lot of frequent patterns across different level items than in the same level, and then the discovered service patterns are the resource to provide suitable service to user in combination of the stored existing services. In this paper we propose a mining method that extracts the frequent service pattern from association across multi-level based on ontology of spatio-temporal information and service information. In general, we can provide compatible service information with user that classify the service usage pattern according to gender, age and user profile. That is, a high-quality service based on spatio-temporal information with regard to user activity and context can be provided to user. In this paper, we focus on service pattern detection by age. To do this, service history is classified by age and we predict the preferring service in terms of user activity pattern and can induce potential customer to do something by providing the suitable service. The proposed mining method that extracts age-service pattern using spatio-temporal information and service ontology is a basis of proving and recommending useful services to user.

2 Related Work

Ontology that is structured into multi-level is formalized with the definition of term and describe the concrete concept of term in specific domain [3, 4]. The context ontology with conceptual domain include the spatio-temporal information and service information considering user circumstance and activity. The ontology is a hierarchical structure of concept level to provide useful services to user.

Data mining is a process of finding out the hidden patterns that can be formalized logically on various data, which is helpful to user's decision making and predict the future activity of user. Particularly, association rule is to detect useful information based on confidence and support among the items of transaction data. Association rule is to mining frequent items. However it has a problem that can't find out the high-weighted item but infrequent item. That's why it does just consider frequency of item without regard to weight. Yun et al. [5] proposed a detection method that is an association rule technique of finding out a meaning pair of items with high occurrence probability but infrequent items. They detect association rules based on relative support that is a mechanism to find out a rare event pair of items that can't originally search them in association mining based on frequency. In [6], they propose an association rule method of weighted interesting pattern that is given different weight to each item according to item importance, so this method can find out high-quality association rules that user is interesting in.

Weighted pattern Mining is a technique of high-weighted pattern mining in case each item has different weight. Generally, in business data analysis, user shopping pattern on item depends on the season and period, so the different weight is given to item and also a weight depends on item price. There are weighted frequent pattern

mining methods based on Apriori algorithm of MINWAL [7], WARM [8], WAR [9] and so on. They have a problem to have much processing time to scan database several times. To solve this problem, WFIM [10] is proposed to improve processing time. In [10], they set to minimum weight and weight range of items and a FP-tree is composed of right toward ascending weight that is satisfied with downward closure. WIP [7] defines the feature of weight pattern by weight affinity concept and detects the useful rule of high-weight affinity. WFIM and WIP is required twice scan of database like FP-tree.

The sequence pattern mining considers only occurrence sequence of items in data set. Therefore it is easy to find out simple sequence pattern but difficult to find out interesting item in real-world application. To supplement this fault, weighted sequence pattern mining [11, 12] is proposed to detect the sequence pattern with great level of interest and concern that considers both occurrence sequence of item and importance of item in real-world.

In this paper, we detect weighted frequent service patterns based on service ontology. It is similar to generalized mining technique in specific domain with multi-class. However there is a difference that we apply the multi-level association rule mining algorithm considering the association between weighted items at different level. Accordingly, the proposed algorithm can detect a variety of service pattern to provide to user.

3 Discovering the Association Rule with Weighted Service

Each item is given weight according to importance, and the weight of item leads to detect association rule of relative meaning items. Service frequent pattern is a meaning that the service is often offered to user at the place at that time. For example, frequent service pattern (Sr7, Sp3, St2) includes that service pattern (Sr7, Sp3), (Sr7, St2), (Sp3, St2) are also frequent. And our method leads to search association rule by gender, age and so in service ontology with place and time information.

It is assumed that the example data to describe our method is the history data offered to young people in their twenties at the same place at the same time. We set the minimum support to 3 to explain the mining process. Figure 1 is a part of service ontology of 3-level. Items at lower level in ontology mean a specific service item of higher level and have greater weight than or equal to that of higher level.

To simplify the item, items of first level are represented as capital-letter and second level is represented as small letter and third level is represented as number. Service item at the same level is given in order. That is, service item in 1-level is represented as guide(A), reservation(B) and recommend(C). Each sibling item in second level of the same parent service is given the small letter in serial order. The guide(A) service has location(a), traffic(b), weather(c), sale(d), event(e) and so on. And also the third level items having the same parent are numbered. The other

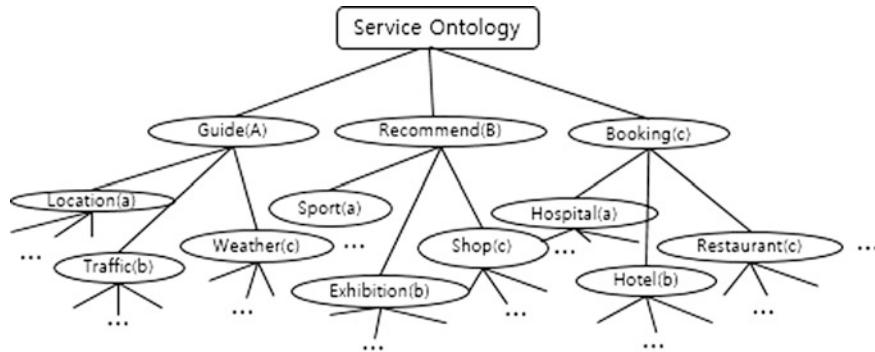


Fig. 1 Service ontology

service is given the symbol in the same way. For instance, Aa2 means the service item of guide(A)-location(a)-tourist site(2).

Table 1 is a transaction data of provided service. We assume that minimum frequent set to 3 and minimum weight set to 1.5. The weight support of item is measured by weight * item frequency. Table 2 is the filtered data from item set A**, B** and C** of level-1 meeting minimum frequency and minimum weight. Most of original data are composed of A**, B**, and C**, so all data in Table 1 is frequent in 1-level. To simplify the data, the second filtering is performed by the item frequency of level-2. The minimum frequency is 3 and the satisfied item(item: frequency) is Aa*:3, Ad*:3, Ab*:3, Ba*:3, Bd*:4, Cc*:4. An item Ab*:3 out of these meets the minimum frequency but not satisfying the weight * frequency (0.4 * 3 = 1.2). Consequently, Table 2 is the result of filtered data.

Raising the level, items associated with item A**, B**, and C** are extracted from the filtered data. The algorithm is a cross-compared method between each

Table 1 Resource data and weight support

TID	Rule items	Itemset: Wsup
100	Aa2, Af1, Ba1, Bc1, Cc1, Cd1	Aa*:2.4(0.8 * 3)
200	Ab1, Ad3, Ba1, Bd2, Cc1	Ab*:1.2(0.4 * 3)
300	Aa1, Ae1, Af2, Ba1, Bd1, Cf2	Ad*:2.4(0.8 * 3)
400	Ab1, Ad2, Bb1, Bd2, Be1, Cc2	Ba*:2.7(0.9 * 3)
500	Aa2, Ab1, Ad1, Bd2, Cc1	Bd*:2.0(0.5 * 4)
		Cc*:3.2(0.8 * 4)

Table 2 Filtered data

TID	Rule items
100	Aa2, Ba1, Cc1
200	Ad3, Ba1, Bd2, Cc1
300	Aa1, Ba1, Bd1
400	Ad2, Bd2, Cc2
500	Aa2, Ad1, Bd2, Cc1

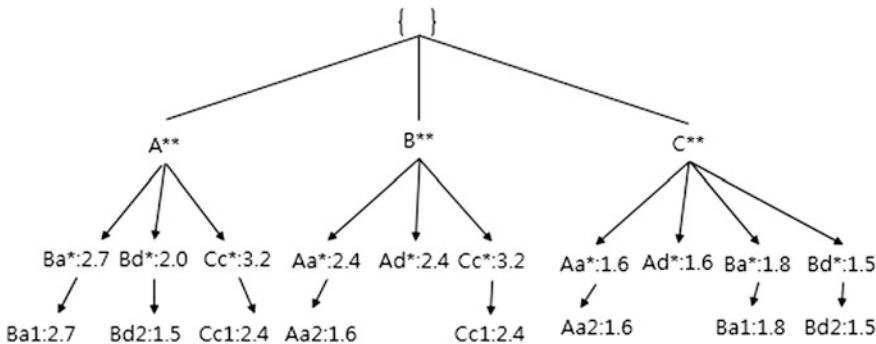


Fig. 2 Weighted level-crossing tree

level. Figure 2 is the total frequent item set of each item A^{**} , B^{**} and C^{**} . During level-crossing tree processing, some items that is not satisfied with minimum weight is deleted. For example, $Bd1:1(0.5)$, $Cc2:1(0.8)$ in the level-crossing tree based on item A^{**} are deleted and $Aa1$, $Ad1$, $Ad2$, $Ad3$, $Cc2$ based on B^{**} are deleted. The same way is applied to level-crossing tree based on C^{**} .

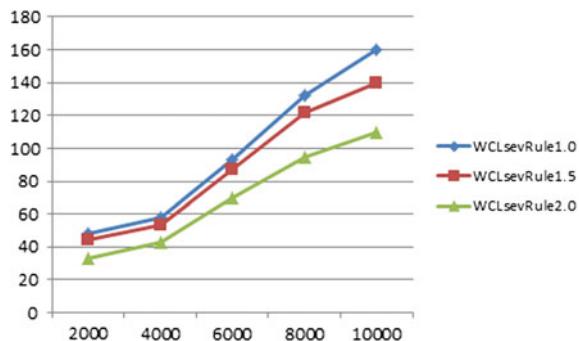
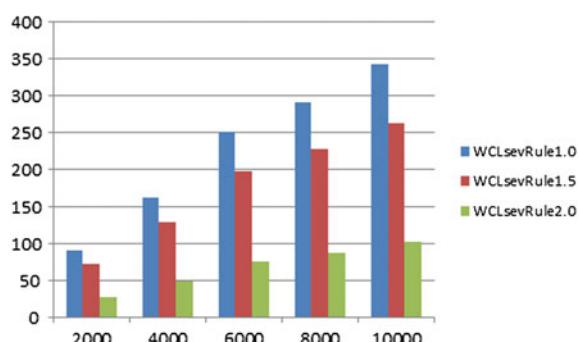
4 Experiments

To evaluate weighted service mining method, given to the minimum threshold, we conducted the level-crossing tree based mining to measure the processing time and the number of rules. In order to experiment, we produced randomly the test data and used the 10,000 transactions including 2 or 3 service items in a transaction. It is gradually tested with 2000 transaction unit to estimate the change of processing time and number of rules.

The first experiment is to compare processing time of WCservRule1.0, WCservRule1.5 and WCservRule2.0 by changing the weight 1.0, 1.5, and 2.0. As you have seen in Fig. 3, the larger the weight, the less time it take the processing. The reason is that the number of candidate frequent items to meet minimum weight is decreased, and it takes the less time.

Figure 4 compare the number of searched rules on the same condition with the first experiment. The larger the weight, the fewer the number of rules. The counting of frequent item, $Aa1$, includes the item A , Aa , $Aa1$. As a result, the small number of frequent items leads to get the fewer number of rules. Of course it is important that it's given the appropriate weight on data attribute to get useful frequent items.

On the basis of other experiments, we found out that weighted service frequent pattern mining can extract the useful items excluded from the mining method considering only the frequency of items. Generally, service rules is already stored in the ontology rulebase, though, it is important to be a combination of stored rule and new rules to provide high quality services for user.

Fig. 3 Execution time**Fig. 4** Number of rules

5 Conclusion

We have proposed a method of weighted frequent pattern mining based on service ontology. We adopt a weight of service significance by spatio-temporal context to find out the useful service patterns for user activity. The experimental results have seen that more rules than those of general mining are searched in weight based mining and also undiscovered rule can also searched. The new detected service patterns make a combination with existing service rules to offer much better service to user. The proposed method can be used to make a response to user activity and predict and analyze current requirements of user.

References

1. Thakur, R.S., Jain, R.C., Pardasani, K.R.: mining level-crossing association rules from large databases. *J. Comput. Sci.* **2**(1), 76–81 (2006)
2. Ramana, V., Rathnamma, M., Reddy, A.: Methods for mining cross level association rule in taxonomy data structures. *Int. J. Comput. Appl.* **7**(3) (2010)

3. Lee, C.H., Helal, S.: Context attributes: an approach to enable context-awareness for service discovery. In: Symposium on Applications and the Internet, pp. 22–30 (2003)
4. Brisson, L., Collard, M.: An ontology driven data mining process. In: The 10th International Conference on Enterprise Information Systems (2008)
5. Yun, H., Ha, D., Hwang, B., Ryu, K.: Mining association rules on significant rare data using relative support. *J. Syst. Softw.* **67**(3), 181–191 (2003)
6. Swargam R.J., Palakal, M.J.: The role of least frequent item sets in association discovery. In: Proceedings of International Conference on Digital Information Management (2007)
7. Ahmed, C.F., Tanbeer, S.K., Jeong, B.S., Lee, Y.K.: Mining weighted frequent patterns in incremental databases. In: Proceedings of the Pacific Rim (2008)
8. Tao, F.: Weighted association rule mining using weighted support and significant framework. In: Proceedings of the ACM SIGKDD (2003)
9. Wang, W., Yang, J., Yu, P.S.: WAR: weighted association rules for item intensities. *Knowl. Inf. Syst.* **6**, 203 (2004)
10. Yun, U., Leggett, J.J.: WFIM: weighted frequent itemset mining with a weight range and a minimum weight. In: Proceedings of the Fourth SIAM International Conference on Data Mining (2005)
11. Lo, S.: Binary prediction based on weighted sequential mining method. In: Proceedings of the International Conference on Web Intelligence, pp. 755–761 (2005)
12. Yun, U.: A new framework for detecting weighted sequential patterns in large sequential databases. *Knowl. Based Syst.* **21**, 110 (2008)

Cyber Security Modeling for the Operation of Virtualized Trusted Networks

Yong-Hee Jeon

Abstract The Virtualized Trusted Networks (VTN) is a novel network architecture which makes it possible to communicate in a secure manner between devices across unreliable networks. It provides confidentiality, availability, quality of service, mobility to users, anytime and anywhere. This paper presents the results of cyber security modeling for the reliable operation of the VTN. The security requirements are derived for security control of the system by analyzing threat elements and attack possibility in the security modeling procedure. By using STRIDE modeling, the threat elements are analyzed for each network component. Then an attack tree is constructed by analyzing attack examples for every threat. The security requirements and countermeasures are proposed to respond against cyber security attacks to the VTN.

Keywords Security modeling · Attack tree · Trusted networks · Security requirements

1 Introduction

For reducing the information disclosure and compromise by unauthorized intrusion through Internet, the Intranet may be implemented to block external access for safe operation. The network without any access point from Internet is called closed networks. Users in closed network infrastructure may only access servers in their own domain or network which they belong to. In the virtualized network environment, however, they can perform their tasks even during their business trip to other areas via network authentication. Therefore, users can securely use, communicate, and share services they want, anytime and anywhere, by using wired or

Y.-H. Jeon (✉)

13-13, Hayang-Ro, Hayang-Eup, Gyeongsan-si, Gyeongsangbuk-do, Republic of Korea
e-mail: yhjeon@cu.ac.kr

wireless terminals through the construction of Virtualized Trusted Networks (VTN).

The aim of development of VTN is to construct trusted networks so that users in the network may perform their task, anytime and anywhere, in a secure manner by using wired/wireless terminals. Therefore, we need to perform cyber security modeling and analyze the modeling results for the reliable operation of VTN.

The security policy model may be developed by using cyber security modeling of the corresponding system [1]. In order to prevent the stored and transmitting information in the VTN from illegal disclosure, modification, loss and disruption, we need to establish mechanisms to protect them. Thus it is required to develop the security policy model of VTN.

The topics for the security modeling include the precise description of security policy objective, design specification and verification, the analysis of covert channel, and the analysis of implementation conformance [2]. The security objective must include confidentiality, integrity and DoS (Denial of Service), etc. The security model precisely describes the main security points and the relationship of system activities.

The main goal of security model is to make sure the necessary level for the successful implementation of core security requirements [3]. The security requirements must be presented at the initial stage of system development process. The identification of security requirements must accompany the identification of the method to satisfy the security objective, including the higher level of security objective, automated, procedural and physical protection methods. The identification of these types of security requirements and derivation of requirements from the identified higher-level of security objectives are initial problems for the protection principle.

Attack tree model and threats modeling are frequently used as security modeling methods. Therefore these methods are also used in this paper. By deriving threat elements and attack paths through the modeling process, it is believed that the necessary requirements and countermeasures for the security control may be proposed in order to respond against the possible attacks.

2 Background

Closed networks by physical or logical separation are applied to safely use the network for businesses. The closed network only allows users in different organizations to get access to the Intranet when they are authorized and basically blocks them to access to the Intranet of other organizations.

Cloud computing is the technology or service integrating a variety of information including hardware, software and data into the central data server. Thus it enables every authorized users to access it from any place, at any time through the communication network [4]. The analysis on the vulnerability of the systems and risks and modeling are prerequisite for the network operation in such highly trusted

environment. Furthermore, it is necessary to develop the technologies to analyze security attack scenario and to respond against such attacks.

The attack tree for modeling the system security was introduced by Bruce Schneier in 1999 [5]. The basis of the system intrusion may be identified by the security analysis using this attack tree. The security modeling procedure applied to the VTN is as follows [3]:

- Preparation of system specification, network diagram and security evaluation criteria
- System identification such as system components and functions, interaction between components, users and system management, interaction of networks, and system security mechanism
- Identifying vulnerability, threat, access point to them, and listing threat sources and types
- Construction of attack tree including the attack target and the sub-attack targets. Examples of attacks include collapse or destruction of software, IDS (Intrusion Detection System), OS (Operating System), and application software.
- Refinement for the tree and check if all identified threats are defined in the tree.
- Generation of all possible attack scenarios and determination of the security risk level
- Evaluation of security threat by comparing SRL (System Risk Level) and DRL (Desired Risk Level). If $SRL > DRL$, the list of security requirements against the attack scenario is created in this step.
- Selection of security requirements to reduce SRL of the attack scenario.

3 Security Modeling

3.1 STRIDE Threat Modeling

Threat modeling aims to identify early the possible security problems that may happen during the operation of any information system. Therefore it has to be done at the system development or design phase. To identify threats, the STRIDE model proposed by Microsoft was adopted in this paper. STRIDE refers to 6 types of security properties such as spoofing, tampering, repudiation, information disclosure, DoS (Denial of Service), and elevation of privilege [6]. By the STRIDE modeling, the attack types which may occur in systems or software may be identified in advance at the design process.

In this paper, the relevant security service is mapped to each threat as follows: spoofing (S) versus authentication, tampering (T) versus integrity, repudiation (R) versus non-repudiation, information disclosure (I) versus confidentiality, DoS (D) versus availability, and finally elevation of privilege (E) versus authentication

and/or authorization. For each network device, possible attacks are listed. In this paper, main possible attacks scenarios are listed for all the components as follows:

- Spoofing (S): cross-service attack, worm virus or malicious code access, unauthorized access, man-in-the-middle (MITM) attack, harmful packet attack, hijacking, unreliable network access, intrusion attack.
- Tampering (T): cross-service attack, data and software modification, tampering personal information, modification of BGP router traffic.
- Repudiation (R): repudiation of the user account (location, access, management, security), abnormal routing attack, repudiation against malicious responses of computing service users including photo scanning and supply of malicious contents.
- Information Disclosure (I): disclosure of password, resident registration number, personal biometric information, data, and digital signature key; tapping BGP router traffic, communication information, and account/network management information; session hijacking.
- DoS (D): exhausting battery of terminal, infection of malicious code, run-away of priority AP and jamming signal attack, persistent advertisement transmission, overload and congestion on a specific link, malicious code infection, service or hardware error, network congestion, and DDoS attack.
- Elevation of Privilege (E): usage of service violating security level, access to wireless LAN management page, unauthorized user attack, unauthorized elevation of user's privilege, cloud QoS (including security and bandwidth) infringement, and user's security level/authority infringement.

3.2 Construction and Analysis of Attack Tree

Attack tree was generated on the basis of the STRIDE threat modeling by applying the attack tree modeling procedures presented by Schneier [5]. The cloud server in the VTN is designated as root node. The other devices such as terminal, wireless AP, network, and management system are positioned as intermediate nodes. Every attack elements are grouped as leaf nodes in the attack tree. Thus the attack tree describes the attack paths in accordance with data flow between devices in the VTN. The model also enables to analyze the locations and methods of threats or attacks.

OWASP (The Open Web Application Security Project) acknowledges OWASP Top 10 by investigating information exposure, malicious file and script and security vulnerability related to web and by analyzing the vulnerability of 10 web applications. Top 10 lists include injection, broken authentication and session management, cross-site scripting, unstable direct object reference, security misconfiguration, sensitive data exposure, access control to lost function level, cross-site request forging, components with known vulnerability, and nullified request and transmission. Based on the number of risk elements in the VTN, the

attack possibility was categorized into three groups such as high, medium, and low. The attack elements with high level of probability are found as injection, broken authentication and session management, and sensitive data exposure. The elements with medium level of probability are specified as access control to lost function level. The other elements are specified as ones with low level of possibility.

4 Security Requirements and Proposed Countermeasure

4.1 Security Requirements

After the identification of threats and construction and analysis of attack tree, the procedure of security threat evaluation is necessary. In the procedure, the system risk and demand risk levels are compared. In this section, the security requirements for the components of the VTN are specified based on the threat identification and attack tree in the previous sections. The cyber security services to be provided by the VTN include access control, accountability, authentication, availability, confidentiality, and integrity.

In this paper, the security requirements of VTN cloud are only presented due to the page limit. The security requirements for the virtualized cloud include encryption between VMs(Virtual Machines) because of the internal vulnerability between them. In order to cope with the attack using the internal communication in virtualization space, the communication monitoring between VMs is also required. As the key security requirements, the detection of transmission of malicious code between VMs and DDoS security inside network to use cloud service are identified. The security requirements for cloud are specified as follows [2]:

1. Access control

- Firewall between virtualization servers to prevent the intrusion of traffic
- Detection and control of malicious code infection, unreliable interface and API of cloud
- Identification and access control in cloud account management

2. Accountability

- Digital forensic function for cloud storage
- SIEM (Security Information and Event Monitoring) technology-applied cloud system monitoring against the release of cloud data and malicious users
- Backward tracing technology against malicious behaviors of cloud computing service such as port scanning and distribution of malicious contents

3. Authentication

- Secure SSO (Single-Sign On) authentication or OpenID and authorization management technology for cloud users to distributed processing resource

4. Availability

- Providing service in resource sharing and aggregation congestion
- Continuous service provisioning by fault-tolerant technology or dual system construction even in the hardware trouble
- Incorporation of resource broker concept to avoid consecutive threat by malicious code or intentional DoS (denial of service) to cloud

5. Confidentiality

- Proving secure encryption for storage data and transmission data of cloud

6. Integrity

- Verifying the effectiveness of information input and providing error checking function for the stored data and exchanged message in cloud computing.

4.2 *Proposed Countermeasure*

In order to provide six types of security services for the VTN, the following countermeasures are proposed [2].

1. Access control

- Detection and prevention of cross-service attack
- Access control technology for hybrid terminal, wireless AP and network
- Remote terminal control and malicious code detection technology
- Highly trusted gateway security technology such as SDN (Software Defined Network)
- Unreliable interface API control technology
- Virtual Machine malicious code detection and control technology
- Global control management technology based on security policy

2. Accountability

- Log-file analysis and digital forensic technology
- SIEM technology

3. Authentication

- SSO and multiple authentication technology
- Device, packet header and payload, cloud, and control system authentication technology
- Secure Wi-Fi and BGP authentication technology

4. Availability

- DDoS detection technology, Fault-tolerant technology, Flexible mobility technology
- Congestion control technology

5. Confidentiality

- Sensitive data encryption technology
- Identification and authentication information encryption technology
- Transmission data encryption technology
- IP data encryption technology
- Network security technology
- Virtual Machine transmission data encryption technology
- Control message encryption technology

6. Integrity

- Software and authentication data integrity diagnosis technology
- Hijacking prevention technology
- Cloud and control message integrity diagnosis technology
- ARP spoofing prevention technology.

5 Conclusions

In this paper, cyber security modeling was performed for the reliable operation of the VTN. First, six types of threat elements were specified according to the STRIDE threat modeling. Then the VTN attack tree was implemented using the approach proposed by Schneier. Using the implemented tree, the attack path and possibility were analyzed. Based on the analysis, the possibility of attack was categorized into 10 groups in accordance with OWASP Vulnerability Classification of 2013. It was revealed that the possibility of injection, session management and sensitive data exposure was high among them. Finally, in order to respond against these threat elements, the security requirements and countermeasures were proposed.

References

1. Gallagher, R.R.: A guide to understanding security modeling in trusted systems, NCSC-TG-010 (1992)
2. Jang, J.S., Kim, E.J., Jeon, Y.H.: Information security modeling for the operation of highly trusted networks, J. KIIT, **12**(10) (2014)
3. Khand, P.A.: Attack tree based cyber security analysis of nuclear digital instrumentation and control systems. The Nucleus **46**(4), 415–428 (2009)

4. IBM Virtualization White Paper, IBM (2006)
5. Schneier, B.: Attack trees. *Dr. Dobb's J.* **24**(12), 21–29 (1999)
6. Swiderski, F., Snyder, W.: Threat modeling, *microsoft professional*, MS, pp. 15 (2014)

Development of a Quantitative Evaluation Method for Vehicle Control Systems Based on Road Information

Jinyong Kim, Changhyun Jeong, Dohyun Jung and Byeongwoo Kim

Abstract The focus of intelligent vehicle technology is progressing rapidly from self-control to inter-vehicle control and vehicle-infrastructure control. While international standards have been established for the evaluation of inter-vehicle control systems, few studies exist on the evaluation of control systems between vehicles and infrastructure. This study provides evaluation criteria for vehicle control systems that utilize road information; it also evaluates control performance by conducting vehicle tests in the developed evaluation scenario. The results of this study will contribute to the development of advanced evaluation methods for control systems between vehicles and infrastructure.

Keywords Adaptive cruise control · Automatic emergency braking · Inter-vehicle control system · Vehicle-infrastructure control system · DGPS

1 Introduction

Intelligent vehicle technology has seen significant development with the growing emphasis on vehicle safety and convenience. The focus of intelligent vehicle technology is shifting from vehicular safety to inter-vehicle control technology, and

J. Kim · C. Jeong · D. Jung

Korea Automotive Technology Institute, 303 Pungse-ro, Pungse-myun,

Dongnam-gu, Cheonan-si, Chungnam 330-912, Korea

e-mail: kimjy@katech.re.kr

C. Jeong

e-mail: jeongch@katech.re.kr

D. Jung

e-mail: dhjung@katech.re.kr

B. Kim (✉)

School of Electrical, University of Ulsan, 102 Daehak-ro, Nam-gu,
Ulsan 680-749, Korea

e-mail: bywokim@ulsan.ac.kr

is expected to turn to communications and control technology between vehicles and infrastructure. Some examples of control systems between vehicles are Adaptive Cruise Control (ACC), Forward Vehicle Collision Warning (FVCW), and Automatic Emergency Braking (AEB). ACC is an automatic cruise control system that enables drivers to cruise at a preset speed or adjusts the speed to maintain a safe distance from vehicles ahead. FVWC alerts drivers of possible collisions, while AEB applies braking in cases of emergency after warning drivers. The smart highway project is an example of a communications and control system between vehicles and infrastructure. Long-distance radars are installed on highways to detect falling objects or stationary vehicles; collisions are prevented by delivering such information to approaching vehicles.

An effective evaluation system is essential to enhance and verify the performance of the aforementioned intelligent vehicle technology. Existing studies have proposed using hardware-in-the-loop simulation for the evaluation of ACC electronic control units [1, 2]. Driving simulations and stationary targets have been utilized in AEB evaluation [3]. In another study, targets moving at up to 70 km/h were used to assess FVCW and AEB systems [4]. While various evaluation methods have been developed for inter-vehicle control systems, few studies exist on the evaluation of control systems between vehicles and infrastructure.

Against this backdrop, this study proposes a quantitative method for objective performance evaluation of vehicle control systems that utilize road information infrastructure.

2 Selection of Evaluation Items

For a quantitative evaluation of vehicle control systems that utilize road information, this study selected the evaluation items shown in Table 1. Precise measuring equipment was used to ensure objective and quantitative evaluation of control performance. In the straight mode, warning distance accuracy was evaluated by measuring the warning time and relative distance to the stationary vehicle. More specifically, the system was considered satisfactory in the following cases: a Level 3 message was displayed at 311 m from the stationary vehicle ($TTC = 16$ s); a Level 2 message was displayed at 107 m ($TTC = 5.5$ s); and a Level 1 message was displayed at 70 m ($TTC = 3.6$ s). The evaluation criteria were set at ± 10 m based on ISO 15623 [5], which specifies the requirements for FVCW systems, and in consideration of system dependence on Wireless Access in Vehicular Environment (WAVE) and GPS. The system enters automatic control after Level 1, and longitudinal acceleration was measured to ensure that the vehicle decelerates at less than 3.5 m/s^2 after cruising at a constant speed of 70 KPH. The criteria for acceleration/deceleration was established with reference to ISO 22179 [6], which requires ACC systems to decelerate at less than 3.5 m/s^2 for a speed of 20 m/s (72 kph). The stop distance accuracy was evaluated by looking at whether the test vehicle was able to stop at exactly 7 m in front of a stationary obstacle. The range

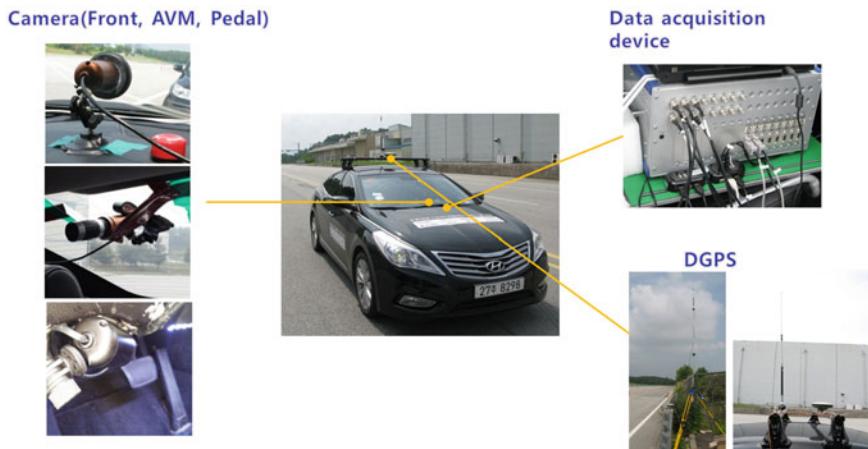
Table 1 Evaluation items for vehicle control systems

Measurement	Analysis	Criteria
Warning time and relative distance	Warning distance accuracy (approaching 70 kph: 311 m, 107 m, 70 m)	$\leq \pm 10$ m (self-established criteria) ($\leq \pm 1$ m ISO 15623: use vehicle radar)
Longitudinal acceleration	Longitudinal acceleration/deceleration below 3.5 m/s^2 at 70 kph	$\leq 3.5 \text{ m/s}^2$ @ 20 m/s (ISO 22179)
Stop distance before obstacle	Stop distance accuracy (target: 7 m ahead)	$\leq \pm 5$ m (self-established criteria)
Vehicular speed	Accuracy of attaining target speed (40 kph) at curve entrance ($R = 40$ m)	$\leq \pm 10$ kph (self-established criteria)

was set at ± 5 m to take safety into account. Vehicle control performance in the curved mode was evaluated by examining whether the vehicle was able to slow down to 40 kph when entering a curved road with a radius of 40 m. The range was set at ± 10 kph in consideration of GPS performance.

3 Test Setup

The data measuring equipment used in the evaluation was installed as shown in Fig. 1. Three cameras were attached to record the front view, warning display of the GPS navigation system, and the brake pedal. A bio-signal monitoring device was

**Fig. 1** Configuration of data measuring equipment

used to monitor the driver's brain waves. The vehicle was equipped with high-precision DGPS capable of sampling at 20 Hz to an accuracy of 1 cm. The signals of individual devices were collected via a data acquisition device and were synchronized for storage.

4 Evaluation Scenario

To evaluate vehicle control systems, we developed an evaluation scenario comprised of a straight and curved road, as shown in Fig. 2. In the straight mode, the driver steers the vehicle at a constant speed of 70 kph without operating the brake. When the distance to the stationary vehicle in front is 311 m, the navigation system displays a Level 3 warning message. When the distance becomes 107 and 70 m, the corresponding warning messages are Level 2 and Level 1, respectively. The vehicle then comes to a stop by triggering of the automatic brake. In the curved mode, involving a curved road with a radius of 40 m, the driver steers the vehicle at 70 kph without operating the brake. The system calculates the speed at which the vehicle approaches the curve and automatically applies the brake to adjust the speed to 40 kph.

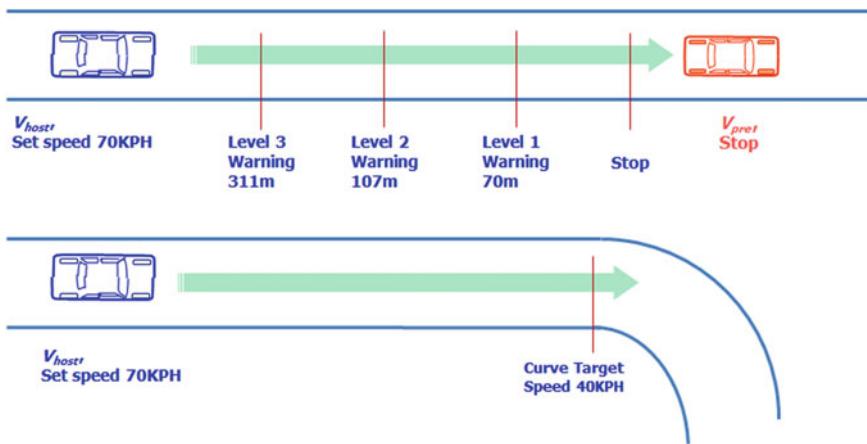


Fig. 2 Configuration of evaluation scenario

5 Evaluation Results

5.1 Warning Distance and Stop Distance Accuracy

Warning distance was calculated using DGPS signals and message displays on the navigation system. Here, the DGPS is capable of sampling at 20 Hz, and the camera stores 30 frames per second. DGPS values are retrieved for warning messages ranging from Level 1 to 3. Latitudes are converted to meter units based on the KATECH coordinate conversion technique, and relative distance to the stationary vehicle is derived. Table 2 shows the results for a total of 11 tests. The average warning distance and stop distance fell within ± 10 m of the reference value, thus satisfying the evaluation criteria. A one-sample t-test was applied for statistical analysis.

5.2 Longitudinal Acceleration/Deceleration and Curve Speed Accuracy

Longitudinal acceleration/deceleration was calculated by averaging the acceleration/deceleration over 2 s from the moment of entering the Level 1 warning stage. The speed at which the vehicle approaches the curve was derived by averaging the speed for one second before and after the steering angle changes to 20°. Table 3 shows the longitudinal acceleration/deceleration and the curve speed. The average longitudinal acceleration/deceleration for 10 trials was found to satisfy the evaluation criteria by falling below 3.5 m/s^2 , and this was found to be statistically significant. The average curve speed was also satisfactory at $40 \pm 10 \text{ kph}$.

Table 2 Evaluation of warning/stop distance accuracy

Classification	Warning distance accuracy			Stop distance accuracy
	Level 3 position	Level 2 position	Level 1 position	
Average	315.03	106.72	72.66	7.185
Standard deviation	7.07	7.58	5.57	0.543
95 % confidence interval	310.28– 319.78	101.63– 111.81	68.92–76.41	6.821–7.550
Criteria	311	107	70	7
Results	Satisfactory	Satisfactory	Satisfactory	Satisfactory

Table 3 Evaluation of longitudinal acceleration/deceleration and curve speed accuracy

Classification	Longitudinal acceleration/ deceleration (m/s^2)	Curve speed (kph)
Average	2.63	42.06
Standard deviation	0.38	1.24
95 % confidence interval	2.35–2.90	41.17–42.94
Criteria	3.5	40
Result	Satisfactory	Satisfactory

6 Conclusion

This study proposed a method for quantitative evaluation of vehicle control systems built upon road information. The evaluation items were selected with reference to existing evaluation methods for inter-vehicle control systems; modifications were made to evaluation criteria in consideration of road information measurement characteristics. After developing an evaluation scenario, statistical techniques were applied to various factors such as warning time, stop distance, deceleration, and curve speed. The proposed evaluation method is expected to contribute to the development of an international standard for vehicle control systems based on road information. For future work, we will further investigate these systems by utilizing bio-signals to evaluate driver acceptance.

Acknowledgments This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the CITRC (Convergence Information Technology Research Center) support program (NIPA-2015) supervised by the NIPA (National IT Industry Promotion Agency) and the Industrial Convergence Foundation Building Project by the Ministry of Trade, Industry and Energy.

References

1. Hwang, Y., Lee, H., Ki, S., Yang, I.: Evaluation of collision avoidance system (ACC/PCS) with hardware-in-the-loop simulation. In: Proceedings of KSAE, pp. 2358–2365 (2009)
2. Kim, J.Y., Jung, D.H., Jeong, C.H., Choi, H.J.: Worst-case scenario development based on sine with dwell test and evaluation for vehicle dynamics controller in UCC HILS. *Int. J. Automot. Technol.* **15**(6), 961–966 (2014)
3. Song, I., Yu, S., Kim, M., Kim, B.: A study on development of sensor evaluation methods for AEBS & ACC and its unmanned target object prototype. In: Proceedings of KSAE, pp. 1016–1021 (2012)
4. Woo, J., Kim, M., Lee, S.: Study on the test method of AEB and FCW system. In: Proceedings of KSAE, pp. 1160–1163 (2013)
5. Intelligent transport systems—forward vehicle collision warning systems—performance requirements and test procedures. ISO 15623 (2013)
6. Intelligent transport systems—full speed range adaptive cruise control (FSRA) systems—performance requirements and test procedures. ISO 22179 (2009)

Design Challenges and Implementation of a Shipborne Gateway for Safe and Secure Navigational Networks

Kwangil Lee and Moonsub Song

Abstract As the shipborne system becomes more complex by integrating more shipborne equipment through a network, the safety and security for the navigational system becomes an important issue. In this paper, we address some design challenges and issues for the design of safe navigational networks including cyber security. In addition, we introduce the safety and security system for navigational networks. Finally, we implement a security shipborne gateway system to interconnect with safe and secure navigational networks.

Keywords Gateway · Security · Safety · Navigational networks · Monitoring

1 Introduction

The safety of the ship has a significant impact on the protection of the environment and human life. The ship operation must be safely protected especially for navigation. This leads the shipborne system and networks to be independent and isolated system in tradition [1, 2]. However, the shipborne system is evolved into an automated and integrated system as the communication and digital technology evolves significantly and the demand for the eco-navigation increases [1]. To meet this requirement, a new Ethernet based international standard was published, called Light-Weight Ethernet which provides a method by which navigational and radiocommunication equipment can be safely interconnected in a single Ethernet network [3, 4]. However, many ships require more complex bridge systems due to heavier use of remote diagnostics and maintenance, tighter integrations with other

K. Lee (✉) · M. Song
Electronical and Telecommunication Research Institute, Gajeongro 218,
Daejeon, Republic of Korea
e-mail: leeki@etri.re.kr

M. Song
e-mail: sirius@etri.re.kr

systems onboard of the ship or off-ship or for any other reason, e.g. to support new e-navigation services [5, 6]. This threat the safe and secure operation of the ship since the malfunction and failure of equipment affect other equipment in the network and increases the possibility of the cyber security threats. Thus the protection of the shipborne navigational networks from various threats needs to be provisioned to guarantee the safe operation and navigation of the ships [7, 8]. This paper addresses some safety and security issues for shipborne networks and explains the shipborne gateway system suitable for protecting shipborne networks from external networks.

2 Shipborne Network Architecture

2.1 Safety and Security Requirements

The followings are summary of the safety and security requirements for the shipborne network.

- Virus and worm: virus and worms in some infected equipment may paralyzed the network and system availability
- Mis-configuration or mis-operation: Configure and operate the equipment in improper way by unauthorized persons or non-experts.
- Mal-functioning: Some equipment may mal-functioning under a certain circumstances and it effects the operation of other equipment.
- Hacking and malicious attack: This could be malware, backdoors or other hostile software.
- Eavesdropping: This could be data from data files or technical data that can be used to predict or compromise system functions or plans.

2.2 Safe Shipborne System Architecture

The shipborne network architecture to provide safety and security is illustrated in Fig. 1 [4]. The figure shows some required physical and software components which add more functionalities than that of basic shipborne equipment. The logical software functions are represented as a pentagon in Fig. 1.

As illustrated in Fig. 1, the network infrastructure provides the security and safety functions and application level gateways are defined so as to protect from the threats from the external networks. This paper also introduces some safety functions such as network monitoring function and system management function. The networks monitoring function monitors all activities and behaviors of the equipment

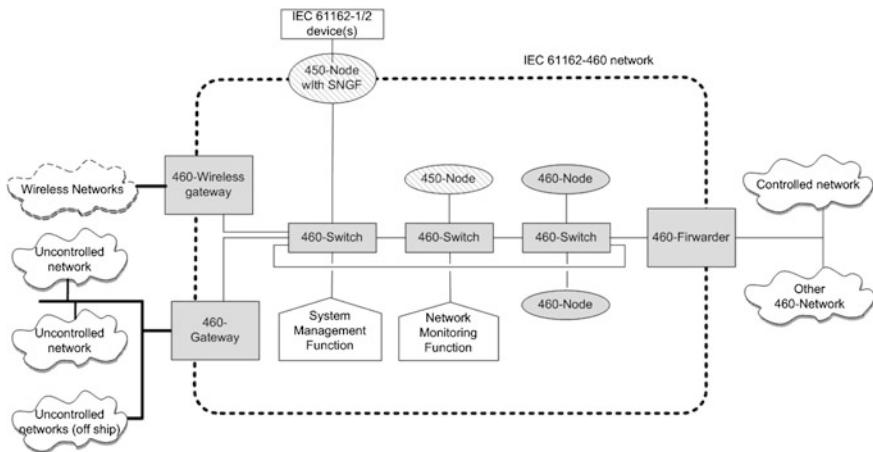


Fig. 1 Overview of a secure shipborne network

and networks and generate warning or alarms to the navigator if the unexpected or suspicious behavior was detected.

3 Shipborne Safety and Security

3.1 Shipborne Network Security

The shipborne network shall be protected from various threats within a network itself. The followings are some of scenarios and protection mechanism from the internal threats.

- Access Control: Two types of access controls are defined: Device and network access control. Device access control is to physically authenticate a user before any changes in equipment can be made. Network access control is to prevent any un-authorized equipment and traffic from accessing the network. These authorizations are performed using ACL (Access Control List) in network infrastructure equipment. The network infrastructure devices themselves are authorized by their MAC address and data traffic in the network is authorized by originator IP address and UDP/TCP port number.
- Denial of Service: Protection from a Denial-of-Service (DoS) attack is a common safety and security function. Two mechanisms are defined to protect from DoS in shipborne networks. The first method is to limit the maximum traffic volume for each data traffic stream using QoS (Quality-of-Service) mechanisms. The other method is to provide prevention from ICMP and IGMP DoS attacks.

- Removable External Data Source (REDS): Arguably, the most likely security threat for any control system is removable storage devices such as USB memory sticks. To protect from such threats, All automatic execution at a 460-Node from REDS including USB auto-run shall be prohibited. Manual execution of any type of files from REDS shall only be possible after the device and its programs have passed an authentication check. The manual execution shall be possible only for the files which are verified before execution using digital signature or special keys.

3.2 External Network Security

The external communication security is to provide a security when any equipment in a navigational network is going to communicate with outside of the networks. Since the outside of navigational networks is not guaranteed of the security, all traffics from outside of the networks shall be passed or processed through a secure shipborne gateway. Figure 2 illustrates an example of external communication scenario and functions of a shipborne gateway.

The followings are the functions required for the external security:

- Firewall: The shipborne gateway is required two types of firewalls: external and internal firewall. An external firewall is to blocks all traffic from outside of the network unless it is registered and destined only to equipment in the DMZ. This means that in principle all direct communication to a network is not allowed. An internal firewall is to blocks all traffic unless it is destined to equipment in a shipborne network and it originates from equipment in the DMZ. All traffic passing through the internal firewall is registered in advance.
- Communication Security: By default, any equipment in navigational networks cannot communicate directly with external networks. However, when a communication is required, explicit permission from an administrator or supervisor is required with the monitoring. Also, a list of all activated direct connections shall be recorded by the gateway. All direct connection shall be terminated automatically after a pre-defined time period.

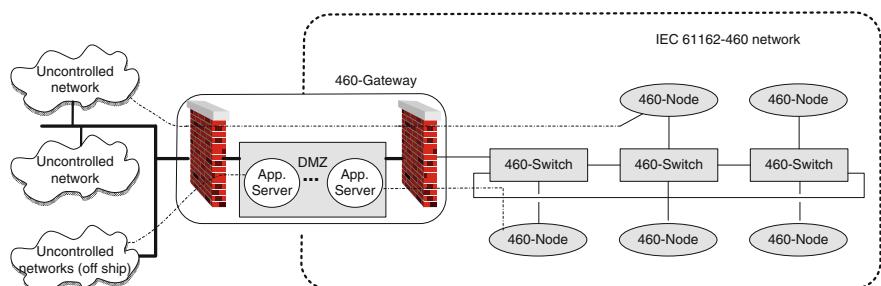


Fig. 2 External communication security with a shipborne gateway

3.3 Additional Security Functions

The network monitoring is also important for providing safety and security. By detecting unexpected behaviors of network and equipment, system administrator can cope with the potential threats. For this, monitoring the network load, redundancy and topology monitoring is required. When detects some violations or unexpected behaviors, then it generate alerts to the administrator.

Also, each network infrastructure need to provide the roll-back mechanisms to revert its configuration or operation whenever network administrator wants. When it is not provided by the system, it needs to be provided by the system management system in a network.

4 Implementations for Shipborne Gateway

In this paper, we implement a shipborne gateway which implements communication with DMZ. We also implement a gateway and network monitoring function. All these functions are operated in a shipborne gateway.

Figure 3 illustrates the system monitoring function. The system monitoring function monitors the network traffic node. Based on the threshold with the configuration, it generates warnings to the administrator. Also, it also reports the various summary information related with a shipborne gateway including the filtering status and connection time and connection duration.



Fig. 3 Example of a shipborne gateway monitoring functions

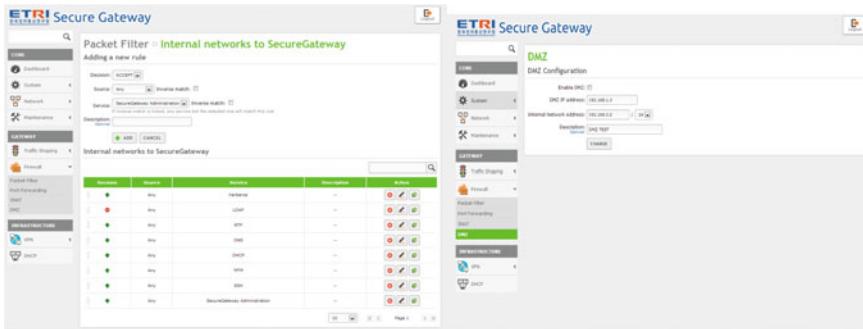


Fig. 4 Example of a shipborne gateway security functions

Figure 4 illustrates the gateway function. The gateway provides a internal and external firewalls by configuring the rules to deny and to accept traffic. It also provides the DMZ functions within a shipborne gateway. With the configuration, the DMZ can be easily provided for the protection of the networks.

5 Conclusion

In this paper, we consider the design challenges for the provisioning of safety and security for shipborne navigational networks. Also, the requirements and functions are analyzed. In this paper, a shipborne gateway is defined so as to provide security with monitoring function. The monitoring and security function is essential for the realization of the security for the shipborne networks. These works are included as an international standard, IEC 61162-460, which is under development by IEC and it is expected to be published in 2016.

The works for the quantitative analysis for the various security threats and implementation of all security components are left for the future works. Finally, the security for all components in maritime domain shall be studied in the near future as the e-navigation implement stages comes near and the cyber security becomes more important issues.

Acknowledgments The IEC 61162-460 standard has been jointly developed by several members of IEC TC80/WG6 of which the author of this paper was the project leader. This research was partially supported by ICT Standardization program of MSIP (The Ministry of Science, ICT and Future Planning).

References

1. Lee, K., et al.: Trends of international standards for shipborne networks. *TTA J.* (2009)
2. Luft, L.A., et al.: NMEA 2000: a digital interface for the 21st century. In: Institute of Navigation's 2002 National Technical Meeting, San Diego (2002)
3. Rødseth, Ø.J., et al.: Design challenges and decisions for a new ship data network. *ISIS 2011* (2011)
4. IEC 61162-450: Maritime navigation and radiocommunication equipment and systems—digital interfaces—part 450: multiple talker and multiple listeners—ethernet interconnection
5. Kim, J., et al.: Integrated shipborne networks design and implementation for digital ship. In: Summer Workshop of Korea Institute of Communication and Information Science, pp. 1008–1011 (2005)
6. Kim, G.: Multi-interface design and implementation for ships. *J. Adv. Inf. Technol. Convergence* **8**(7), 1–6 (2010)
7. ENISA: Analysis of cyber security aspects in the maritime sector. The European Network and Information Security Agency, Heraklion, Greece (2011)
8. Rødseth, Ø.J., Kvamstad, B., Poratne, T., Burmeister, H.-C.: Communication architecture for an unmanned merchant ship. In: Proceedings of IEEE Oceans 2013, Bergen, Norway (2013)

A Design and Implementation of Lightweight ENC for Android App

Moonsub Song, Kwangil Lee, Byungtae Jang and Soonghwan Ro

Abstract A large ship has an expensive electronic chart display and information system (ECDIS) installed on its bridge for the safety of navigation and uses a variety of mobile equipment for auxiliary purposes. In contrast, small- and mid-sized vessels, passenger ships and fishing boats employ mobile equipment instead of such ECDISs, as their primary devices for navigation. Small-sized vessels only require some information and features of the ECDIS owing to the performance and resource restrictions of mobile devices and the characteristics of inshore or offshore navigation. This study proposed information lightening and indexing methods for electronic navigational charts to fit mobile equipment and implemented a tailored display configuration feature for navigators.

Keywords S-57 · ECDIS · Mobile device · Optimization · ENC app

1 Introduction

Recently, the ICT convergence technologies are applied on various industry domains. In the maritime domain, the two major directions for the ICT convergence techniques are digital ship-yard and smart ship services [1]. ICT technology and communication infrastructure has been developed for the realization of digital ship

M. Song · K. Lee · B. Jang
ETRI, Daejeon, Korea
e-mail: sirius@etri.re.kr

K. Lee
e-mail: leeki@etri.re.kr

B. Jang
e-mail: jbt@etri.re.kr

S. Ro (✉)
Kongju National University, Gongju, Korea
e-mail: rosh@kongju.ac.kr

[2–4]. Also, application of ICT technology on ship and on-board equipment has been contributed to the intelligent ship. Due to catastrophic accidents and pollutions, the maritime safety gets more importance and thus IMO (International Maritime Organization) makes a strategic plan for the e-navigation implementation [5]. e-Navigation leads many research and developments in world-wide on the requirements for the safe navigation and broadband communication infrastructures in coast [6]. The main objectives of these approaches are to provide enough information and to help make the right decision to the navigators. This requires the exchange of all maritime information with a standard format and an equipment to display such information [7].

S-57 is an international data exchange standard in maritime by IHO (International Hydrographic Organization) [6]. ECDIS (Electronic Chart Display and Information System) shall be comply with S-57 standard [8]. ECDIS is an equipment which provides not only an electronic chart but also various maritime navigation aids information such as current, temperature, weather, traffic conditions to the navigator. ECDIS is mandatory for all vessels over 300 tons but is not required for small vessels [5] even though it is essential for the safe navigation. This leads most of accidents are caused by small vessels [8]. Thus, small vessels are also required an equipment which display electronic chart similar to ECDIS. Since smart devices including smartphones and pads get more popularity, various apps have been developed to display an electronic chart system for smart devices [9].

In this paper, we propose an electronic chart display app which is suitable for smart devices. We design a database and implement it for electronic chart information so as to improve the user usability. The app can be used as an electronic chart display system for small vessels and is expected to improve the safety of small vessels significantly.

2 Related Works

2.1 *e-Navigation*

IMO defines e-navigation as “the harmonized collection, integration, exchange and analysis of marine information on board and ashore by electronic means to enhance berth to berth navigation and related services for safety and security at sea and protection of the marine environment”. Recently, IMO approved the strategic implementation plan of e-Navigation [5]. Korea is launched a Smart-Nav project (e-navigation for Korea) which implements not only IMO e-navigation services but also services for fishery and small vessels. The major cause of the ship conflicts is from small vessels. The safety of small vessels is most important factor to reduce the accidents. This needs to consider the environment of near ocean and navigation information so as to achieve the goals of e-navigation. The characteristics of Korea e-navigation is not a simple extension to small vessels. By utilizing the ICT

infrastructure in Korea, e-Navigation services will be implemented with maritime mobile devices and services.

Recently, more high-end mobile devices are introduced but still mobile device has memory and graphic processor constraints to be used for the processing of high-volume of electronic chart information. Therefore, data conversion of S-57 and a data structure to rendering for mobile device is required [10, 11].

2.2 S-57

S-57 is an international standard by IHO for the hydrographic information exchange in 1992 and newest version (version 3.1) is published in 2000 [8]. It specifies the record field and structure for the specification of hydrographic data exchange but it should be translated into machine-readable data structure first to be used in electronic chart display system. S-57 consists of features and spatial information which shall have a relationship with feature. The spatial information includes node, edge and face to present point, line and area. Electronic chart represents all space information with node and edge by connecting surface and edges. S-57 stores all hydrographic data in records which consist of fields and subfields. Hydrographic data is stored in the files [12]. A service consists of multiple files and exchange set for meta-data information among files.

2.3 *ECDIS (Electronic Chart Display and Information System)*

ECDIS is a system which replaces paper chart with electronic chart for the navigation and display the chart in the display devices. It also displays additional information such as route planning and route monitoring by providing the ship location information and other ship's location from the sensors. The ECDIS performance standard and product specification is specified by IMO (International Maritime Organization) and IHO (International Hydrographic Organization) accordingly. IEC has a role to define the test standards for ECDIS which is published as IEC61174 [13]. The main function of ECDIS is to generate alarm to the navigation for the collision of the ship in advance, provide optimal voyage plan, analyze the automatic voyage record for the accidents, and provide various navigational information to the navigator [9].

3 Lightweight ENC Data and Indexing

3.1 ENC Data Re-structuring

Figure 1 (left) is a data scheme for S-57 data. This complies the international standard but is not applicable for ECDIS system. The main reason is that several tables are required to be referred to display an object. Thus, each ECDIS manufacturer develops their own system ENC which filters some unnecessary information and reduces the volume of electronic chart information. Figure 1 (right) illustrates the table scheme for database in mobile devices so as to store cell-based electronic chart information which is extracted necessary object information in advance.

Objects in local cell are categorized into dot, line and surface objects. Dot object is an object which has only one location information of bathymetry, lighthouse and rock. Line object is an object to present a line for the shoreline, river and sea boundary. Surface object is used for displaying land, ocean, lake and river. Table 1 shows the type of the S-57 object, the first attribute and second attribute information.

We propose the data table scheme which does not require several queries for rendering information and improves the performance. The rendering information is attribute information required to display each object. In principle, referencing larger volume table may lead to performance degradation since mobile device has the resource constraints, especially such as memory constraints.

3.2 DB Table Indexing

The other cause for the performance degradation is a search method after S-57 data is stored in a well-designed table scheme. This paper proposes a DB indexing approach since there are tens of thousands objects even cell data for simple region.

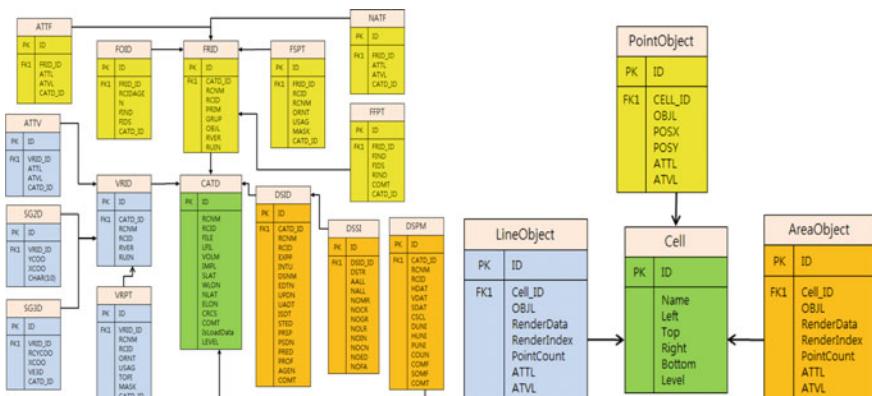


Fig. 1 S-57 data table (left) and re-structured data table (right)

Table 1 S-57 object type (a partial list)

	Object type	ATTL	ATVL
Point	LNDMRK	CATLMK	Tower
	LIGHTS	COLOUR	
	LNDRGN	NOBINM	
	LNDEVL	ELEVAT	
	CURENT	ORIENT CURVEL	
	WRECKS	CATWRK	
	BOYCAR	BOYSHP	Pillar
	BOYINB		
	BCNAR	BCNSHP	Pilebeacon
	DEPTH	DEPVAL	
Line	COALNE		
	SLCNS		
	RIVERS		
	DEPCNT		
	ROADWY		
Area	LNDARE		
	DEPARE	DRVAL	
	RIVERS		
	LAKARE		

We apply this approach to the highly referenced four tables which is illustrated in the following:

- SG2D: point information of line and surface for land and river
- SG3D: point information of line and surface for river and bathmetry
- FSPTObject: table for all objects which use indexed with CATD_ID (Catalogue ID). When information is searched, CATD_ID is used for object extraction.
- Coordinate table: point drawing object. This table uses ID and FSPTObject ID as an index.

(Applying example) `CREATE INDEX idx_SG2D ON SG2D (VRID_ID COLLATE BINARY ASC);`

According to the above usage, using the VRID_ID of the SG2D table as an index and collating sequences in ascending order.

4 Conclusion

This research aims to re-structure object information of S-57 as proposed in this paper, extract and stores binary information for rendering as a file. Figure 2 illustrates a result after the information is applied to SQLite DB in android platform.

Uses of electronic chart display app for mobile devices is usually not navigators but normal users who don't want to display various and complex information.



Fig. 2 S-57 data display using proposed method in android platform

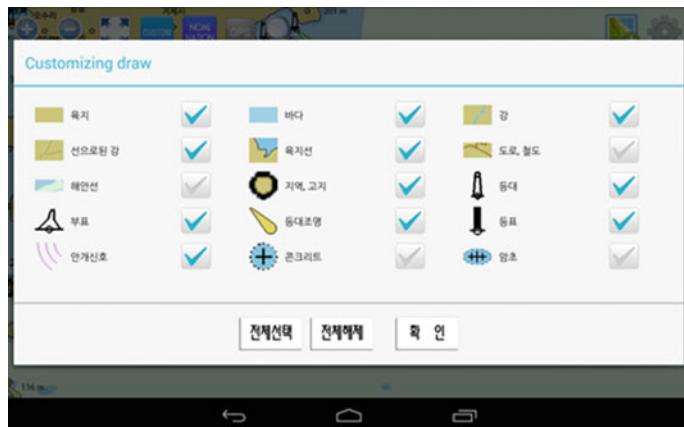


Fig. 3 Object configuration for customizing user interface

In addition, due to size constraints, it is not possible to display all available information. Thus, it is required to design the display information as user needs. This may conflict the requirements of international standard and user needs. This might needs to provide some guideline can be applicable both to navigators and to normal users.

Figure 3 is an example of some configurable object information for the mobile device users. For users who are not expert navigators, it is difficult to understand desired regional map information or distributed file information.

In the future, the quantitative analysis for the proposed scheme and methods of optimal mobile DB rendering are left for future works. Also, the study on the standardization to display simplified ENC lightweight object for mobile device is necessary.

References

1. Lee, W.H., Min, S.W., Jeong, M.A., Lee, S.R.: A proposal for a femto-cell local gateway and packet off-loading for remote shipyards. *J. KICS* **38C**(4), 387–393 (2013)
2. Cho, K.M., Yun, C.H., Lim, Y.K., Kang, C.G.: Real-time transmission scheme for ad hoc self-organizing (ASO) TDMA in multi-hop maritime communication network. *J. KICS* **39B**(5), 260–270 (2014)
3. Bekkadal, F.: Emerging maritime communications technologies. In: 9th International Conference on Intelligent Transport Systems Telecommunications 2009, pp. 358–363 (2009)
4. Kim, Y.G., Lee, S.R., Jeong, M.A., Kim, B.M., Min, S.W.: Efficient data transmission scheme with data fusion inside a smart vessel. *J. KICS* **39C**(11), 1146–1150 (2014)
5. IMO. <http://www.imo.org>
6. Peters, S.W., Heath, R.W., Jr.: The future of WiMAX: multihop relaying with IEEE 802.16j. *IEEE Commun. Mag.* **47**(1), 104–111 (2009)
7. Yang, C-h., Wang, R-f., Wang, S., Wang, X.: The relationship between symbol index and showing content in electric navigational chart based on GIS. *Geosci. Remote Sens. (IITA-GRS)* 564–566 (2010)
8. IHO: IHO transfer standard for digital hydrographic data version 3.1. Special publication no. 57 (2000)
9. IHO: Specification for chart content and display aspects of ECDIS. Special publication no. 52 (2010)
10. Xing, S., Zhang, Y.: PDA based electronic chart pilotage system. In: Asia-Pacific Conference Wearable Computing Systems (APWCS), pp. 159–162 (2010)
11. Jo, G.J., Lee, J.H.: Study on how to display S-57 ENCs in a embedded mobile-platform. *J. Korean Inst. Intell. Syst.* **22**(3), 334–340 (2012)
12. Lee, H.Y.: A design and implementation of SENC structure for efficient storage of S-57 spatial data. *J. Korean Navig. Port Res.* **28**(8), 673–678 (2004)
13. IEC. <http://www.iec.ch>

Study of Censorship in Named Data Networking

Xingmin Cui, Lucas C.K. Hui, S.M. Yiu and Yu Hin Tsang

Abstract Named Data Networking (NDN) (Zhang et al. in Named data networking (ndn) project. Relatório Técnico NDN-0001, Xerox Palo Alto Research Center-PARC, 2010) focuses on content rather than hosts as in today's IP-based network. While NDN aims to incorporate security into its design, some security issues such as censorship in NDN has not been studied. This paper aims to investigate the robustness of file transmission in NDN in the face of censorship. We mainly focus on two types of censorship techniques: name-watchlist attack and deep packet inspection (DPI). We show by simulation that file transmission in NDN is robuster than that in IP-based network since NDN allows multiple outgoing faces when forwarding an Interest packet. Despite of this, we show that censorship in NDN still affects users to get their requested data and therefore anti-censorship is necessary. We further propose an anti-censorship system against these two censorship techniques. We give the system design along with the security analysis and performance evaluation.

Keywords Named data networking · Internet censorship · Anticensorship

X. Cui (✉) · L.C.K. Hui · S.M. Yiu · Y.H. Tsang

Department of Computer Science, The University of Hong Kong, Pokfulam, Hong Kong
e-mail: xmcui@cs.hku.hk

L.C.K. Hui
e-mail: hui@cs.hku.hk

S.M. Yiu
e-mail: smyiu@cs.hku.hk

Y.H. Tsang
e-mail: yhtsang@cs.hku.hk

1 Introduction

In order to deal with scalability, mobility and security issues of today's Internet, new network architectures have been proposed. Named data networking (NDN) [1] is one promising candidate among them. NDN uses names instead of IP addresses to identify packets. There are two types of packets in NDN: Interest packet and Data packet. Communication is driven by data consumers. A consumer sends out an Interest packet identified by his interested name. Intermediate routers forward this packet towards the data producer until a Data packet with the same name is returned. The Data packet is routed back to the consumer taking the reverse way of the Interest packet.

Internet censorship is used to control and suppress the access to, or the publishing of certain information on the Internet [2]. In this paper we aim to study the performance of NDN under two types of censorship techniques: *name-watchlist attack* [3] and *deep packet inspection (DPI)* [2]. In a name-watchlist attack, the censor has a list L , which contains the names he wishes to block or eliminate. He monitors all packets in the network and once he finds a packet whose name is in L , he will suppress or delete this packet. In NDN intermediate routers with bigger computational and memory resources make deep packet inspection more easily to employ.

Unlike in IP-based network where packets are routed according to a pre-calculated spanning tree that gives the shortest path from source to destination, in NDN an Interest can be forwarded to multiple outgoing faces. We explored the robustness of file transmission in NDN against censorship via simulation. The result shows that NDN benefits a lot from the multiple outgoing faces forwarding design. We also showed by simulation that anti-censorship is necessary when the percentage of censors reaches a certain bound. We further proposed a censorship resistant scheme to defend against the name-watchlist attack and deep packet inspection.

In our design, censored names are disguised into valid names to defend against the name-watchlist attack. One censored name corresponds to a large set of valid names. The consumer only needs to randomly choose one from this set and make a request. We assume the existence of the Public Key Infrastructure (PKI) to distinguish the capability of the producer and censors. The producer can recover the original name using his private key, but censors cannot. After receiving the Interest packet, the producer can get a secret with which the replied Data packet is covered to defend against deep packet inspection. The security analysis and performance evaluation show that our system can defend against censorship while keeping an acceptable average file retrieval time.

2 Censorship in NDN

2.1 System Model and Attack Model

Three parities exist in the network: data producers, data consumers and censors. All parties are attached to a public network. There are intermediate routers to link these parties together. The actions each party can take are as follow:

- *Data Producers*: Producers publish the list of names they want to serve and prepare data packets for these names.
- *Data Consumers*: Consumers send out Interest packets to request for the corresponding Data packets.
- *Censors*: Censors can act as normal consumers. Besides, they can observe all packet deliveries on the network. We assume that they can take the role of intermediate routers and block the packets passing by. We call these routers *censor routers* and the remaining routers *normal routers*. We assume the existence of multiple censors in the network to make the censorship more effective.

In this work, we mainly focus on the following two types of attacks:

- *Name-watchlist attack*: Name-watchlist attack [3] blocks packets with censored names. The censor monitors all packets in the network and once he finds a packet whose name is in his censored name list L , he will suppress or delete this packet.
- *Deep Packet Inspection (DPI)*: We assume censor routers have been installed with the deep packet inspection device to check whether the sniffed Data packets contain sensitive content.

Notice that censorship should NOT violate *data availability* [4]. For example, a censor cannot simply delete all the packets in the network since he may delete legitimate packets as well.

2.2 Robustness of File Transmission in NDN Against Censorship

The forwarding strategy in NDN allows an Interest to be forwarded to multiple outgoing faces. This enables an Interest packet to be routed via the most suitable path and can automatically try other paths when disconnection or congestion occurs on the current path. This makes NDN an ideal platform for information dissemination and makes it difficult to regulate the spread of information. In this section we will compare the situations of single outgoing face and multiple outgoing faces to see to what extent NDN benefits from this forwarding strategy to defend against censorship.

We use ndnSIM [5] as our simulation platform. NdnSIM provides several forwarding strategies including *BestRoute* which allows only one outgoing face and *SmartFlooding* which allows multiple outgoing faces. We use a selected Rocket-fuel topology AS-1755 [6]. Leaf nodes are randomly assigned as producers and consumers. Backbone and gateway nodes are assigned as routers. We vary the percentage of censor routers from 0 to 100 % to represent the effectiveness of censorship.

Consumers send Interest packets in the first 10 s. The time to start these requests follows uniform distribution. 30 % of these Interests are with censored names and the Data packets of the remaining Interests are with censored content. Therefore when the censorship is effective enough, no files can be successfully received. We considered both situations of sparse traffic and dense traffic since the significance of Content Store is not the same in these situations. Each consumer sends out 5 and 60 Interests in the situations of sparse and dense traffic respectively. Each file is separated into several blocks and we say a file is successfully received if all blocks have been received. The simulation lasts for 500 s. The percentage of successfully received files using different forwarding strategies is given in Fig. 1.

Figure 1 shows that with the same percentage of censor routers, multiple outgoing faces forwarding enables consumers to receive much more files than single outgoing face forwarding. This indicates that the multiple outgoing faces forwarding strategy in NDN weakens the effectiveness of censorship and favours the spread of information.

Figure 1 also indicates that even if SmartFlooding is used, the percentage of successfully received files decreases with the increase of censor routers. In other

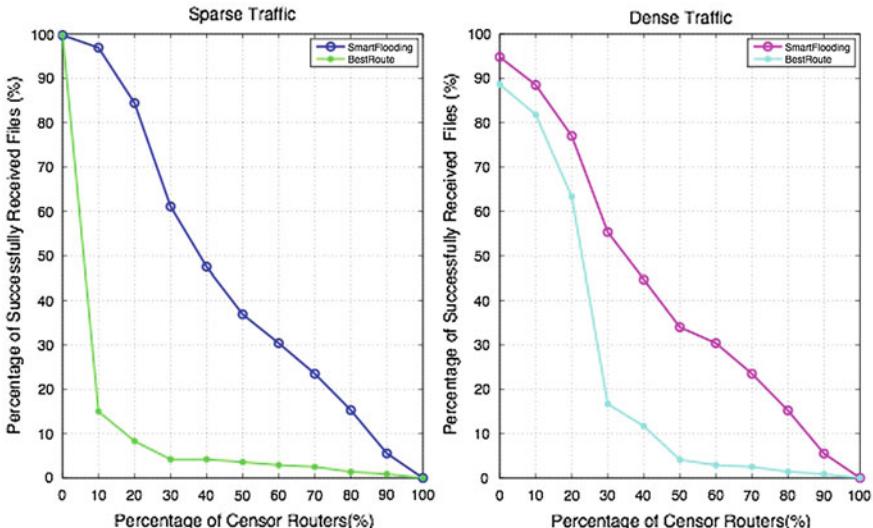


Fig. 1 SmartFlooding (multiple outgoing faces) versus BestRoute (single outgoing face)

words, only relying on limited *normal routers* to forward packets would render a poor network performance. Therefore anti-censorship is still necessary in NDN.

3 A Censorship Resistant System

From now on we will propose our censorship resistant system. We use Smart-Flooding as the forwarding strategy in later simulations to take advantage of the multiple outgoing faces design.

Our proposed censorship resistant system aims to defend against name-watchlist attack and deep packet inspection. DPI can be circumvented with the help of encryption or steganography. The problem is how to negotiate a secret key between the data producer and consumer. In reality the producer cannot differentiate normal users and censors, therefore we aim to construct a system which relies on computation asymmetry rather than information asymmetry. We lay our emphasis on the defense against name-watchlist attack and aim to circumvent DPI at the same time.

To achieve this goal, our censorship resistant system makes efforts on the following two dimensions:

Against name-watchlist attack: Disguise requested names. We assume the existence of an underlying PKI. Censors and normal consumers know the public key of the data producer in advance. The name N is disguised like this: N is concatenated with a *salt* and then encrypted using the producer's public key. To make use of cache, we use a deterministic encryption algorithm such as RSA. Therefore when two consumers happen to choose the same salt value, the cached Data packet can be sent back right away. The encryption result is then attached to a prefix that provides information for routing. To provide replication, the *salt* is within a large range R which is specified by the producer. For example, R can be defined as all binary strings of the length 64 bits. We call a name calculated in this way a *valid name*. For example, one valid name of the censored name “*hku/CS dept/sensiN*” could be “*hkuServer/CS dept/Enc(PK, hku/CS dept/sensiN||0)*” where 0 is the salt value chosen by the consumer and the prefix “*hkuServer/CS dept*” is used by intermediate routers for routing.

The producer broadcasts the correspondence between *name* and R in advance. For a normal consumer, he randomly selects one salt in R , constructs a valid name and sends out an Interest with this name. When the data producer receives the Interest, he recovers the original name by decrypting the suffix and then sends back the requested data. For the censor, however, he cannot do decryption to check whether the original name is a censored name. He has to traverse R to calculate possible valid names until a match with the sniffed name is found. In the worst case, he needs to try all possible values in R . If $|R|$ is very large, this process will take quite a long time.

Against DPI: Cover replied Data packets. After receiving the consumer's request, the data producer generates the covered Data packet on the fly by

encrypting the content using the *salt* value extracted from the Interest packet. After receiving the replied Data packet, the consumer can decrypt it to get the plaintext since he knows the *salt* value which is the decryption key. The censor can figure out the content of the Data packet only after he has got the *salt* value in the packet name. This results in the same computational complexity as in the name-watchlist attack, which requires the censor to try all the values in R in the worst case. Therefore if we can prove that our scheme defends against name-watchlist attack, it will automatically defend against DPI.

4 Analysis

4.1 Security Analysis

Static Version In our design the original names are disguised into valid names whose suffix is of the form $Enc(PK, N||salt)$, therefore the censor needs to figure out whether the sniffed disguised name corresponds to a censored name on his list L . He cannot decrypt the suffix since he doesn't know the data producer's private key. He cannot tell the disguised censored names from disguised uncensored names either. Therefore, he can only calculate possible valid names until a match with the sniffed name is found. In the worst case, he needs to try all possibilities in the range R . Suppose the calculation of one valid name takes t seconds (which is acceptable for a data consumer), the worst case time consumption of the censor is $T = t * |R|$. The producer needs to carefully choose R to make the time difference between the consumer and censor distinctive. For example when he sets R to represent all binary strings of 64 bits, then $|R| = 2^{64}$. Suppose $t = 10$ s, then $T = 5.85 \times 10^{12}$ years. Even if the censor can computer 10,000 times faster than a normal consumer, he still needs 5.85×10^8 years. When the match is found, the consumer has already got the replied Data packet and recovered the content!

Pre-calculate and Match As an option, the censor can pre-calculate all possible suffixes and replace the original censored list with a list of disguised names. Suppose at a certain time, the censor has already calculated a subset of R , say, R_s , then the consumer can only choose a salt from the set $R - R_s$ to avoid being blocked.

The censor needs time $T/2$ to cut the feasible range down to $R/2$. In the above setting, $T/2 = 2.93 \times 10^8$ years. Within this time range, the consumer can select a feasible salt with a probability of at least $1/2$. We call this time range a *feasible window*. To get his interested data, a consumer can select more than one *salt* and request more than one valid name at a time. If he calculates 10 valid names and sends out 10 Interests at a time within the feasible window, the probability that at least one of these Interests arrives at the producer is $1 - (1/2)^{10}$ which approaches 1.

Dynamic Version In the above setting, if the consumer makes a request outside of the feasible window, there is a high probability that the valid name he uses has been

recorded by the censor. The data producer can change R periodically to cope with this challenge. When R is changed, the already calculated names would be useless and the censor needs to re-calculate valid names using the new range. In this case, the producer needs to publish the valid time interval of a certain range R . The consumer needs to include the generation time of the Interest packet for the producer to verify.

The above analysis shows that our scheme effectively defends against name-watchlist attack. As mentioned in Sect. 3, it automatically defends against deep packet inspection.

4.2 Performance Evaluation

In our scheme, the data consumer needs to calculate valid names before sending out the Interest packet and the data producer needs to encrypt the Data packet on the fly after receiving the consumer's Interest packet. We simulated it on ndnSIM [5] to evaluate its performance overhead. The file receive ratio and the average retrieval time of successfully received files with and without our scheme are shown in Table 1. Without loss of generality, we set the percentage of censor routers as 30 %.

Table 1 shows that when each user requests less than 30 interests, our scheme outperforms the original NDN on both aspects of file receive ratio and file retrieval time. This is because without anti-censorship an Interest may need to try several outgoing faces before finding the right way to the producer. When traffic becomes denser, the cost of our scheme also becomes larger. But it keeps a file receive ratio over 97 %, which is much higher than that in the situation without anti-censorship (around 61 %). Therefore this is a tradeoff between efficiency and file transmission quality. The data producer can determine whether to adopt the anti-censorship scheme or not based on the network flow condition and service requirement.

Table 1 Performance overhead of our anti-censorship scheme

IPC*	File receive ratio		Average retrieval time	
	Original NDN (%)	With our scheme (%)	Original NDN (s)	With our scheme (s)
5	61.6	99.2	26.6	24.0
10	61.1	99.2	28.4	27.5
20	61.1	97.5	36.8	29.1
30	61.1	98.7	42.0	41.8
40	61.1	97.8	42.6	53.7
50	61.0	99.5	43.1	71.9
60	61.0	99.1	50.2	88.9

IPC*: Number of Interests sent by each consumer

5 Conclusions

In this paper we studied the robustness of file transmission in NDN under two types of censorship techniques: name-watchlist attack and deep packet inspection. The simulation results show that NDN's multiple outgoing faces forwarding strategy favours the dissemination of information and weakens the effectiveness of censorship. Despite of this, we further showed that anti-censorship is still necessary in NDN.

We propose our anti-censorship scheme which relies on computational asymmetry and does not require the data producer and consumers to pre-share any secret. The security analysis shows that our system can defend against the name-watchlist attack and deep packet inspection. The performance evaluation shows that when the traffic of the network is not dense, our scheme outperforms the original NDN scheme on both aspects of file receive ratio and average file retrieval time.

Acknowledgments The work described in this paper was partially supported by the HKU Seed Fundings for Applied Research 201409160030, and HKU Seed Fundings for Basic Research 201311159149 and 201411159122.

References

1. Zhang, L., Estrin, D., Burke, J., et al.: Named data networking (ndn) project. Relatório Técnico NDN-0001, Xerox Palo Alto Research Center-PARC (2010)
2. Leberknight, C.S., Chiang, M., Poor, H.V., et al.: A taxonomy of internet censorship and anti-censorship. In: Fifth International Conference on Fun with Algorithms (2010)
3. Arianfar, S., Koponen, T., Raghavan, B., et al.: On preserving privacy in content-oriented networks. In: Proceedings of the ACM SIGCOMM Workshop on Information-Centric Networking. ACM, pp. 19–24 (2011)
4. Stubblefield, A.B., Dagster, W.D.S.: Censorship-resistant publishing without replication. Rice University, Department of Computer Science, Tech. Rep. TR01-380 (2001)
5. Afanasyev, A., Moiseenko, I., Zhang, L.: ndnSIM: NDN simulator for NS-3. University of California, Los Angeles, Tech. Rep. (2012)
6. Rocketfuel. <http://research.cs.washington.edu/networking/rocketfuel/>

Research on Design of LTE-Based High-Speed Railway Networks in Korea: Current and Emerging Issues

**Jin-Kyu Choi, Hanbyeog Cho, Hyun-Seo Oh, Kyong-Ho Kim
and Heung-Gyo Ryu**

Abstract In this paper firstly current state and progress of LTE-based public safety network construction project is described, secondly emerging issues with designing LTE high-speed railway network are discussed, and finally a conclusion and future works are addressed. Two LTE networks are several similar features of LTE for public safety based radio access technology, 700 MHz frequency band, 10 MHz bandwidth, and nationwide deployment. Since the inter-network interference can cause the call disconnection with wireless train control which results in a fatal train accident, it is the most important thing at present for designing and planning LTE railway network to eliminate—not to mitigate or reduce—the radio interference, and therefore in our view it urgently needs to secure the reliability and survivability of the wireless train control system when we design and deploy both LTE networks together.

Keywords LTE railway network · LTE-R · Public safety network · PS-LTE

J.-K. Choi (✉) · H. Cho · H.-S. Oh · K.-H. Kim

Industries IT Convergence Department, Electronics and Telecommunications Research Institute (ETRI), 183, Doandong-ro, Seo-gu, Daejeon 1505-2502, Korea
e-mail: jkchoi@etri.re.kr

H. Cho
e-mail: hbcho@etri.re.kr

H.-S. Oh
e-mail: hsoh5@etri.re.kr

K.-H. Kim
e-mail: kkh@etri.re.kr

H.-G. Ryu
Department of Electronic Engineering, Chungbuk National University, Cheongju, Korea
e-mail: ecomm@cbu.ac.kr

1 Introduction

In South Korea the national disaster safety network construction project worth over 1.8 billion dollars has already been started in 2014. This innovative and dedicated network of the public safety will be established, which is based on the following main features: LTE for public safety based radio access technology, 700 MHz frequency band, 10 MHz bandwidth, and nationwide deployment including the railway and the e-navigation network. And target users are about 200,000 people from 330 mandatory agencies, including military, police, fire, EMS, coast guard, provincial administrative offices, electricity, and the gas service. According as this plan, the commercialization project of LTE-based railway network of the conventional and high-speed railway has also been started in 2014.

However, several important problems facing these network construction projects are still being discussed. One of them is how the reliability of the wireless train control system could be technically guaranteed in the case where these two networks should coexist on the same frequency bands. In this paper, we describe current states and progress of these two LTE-based network construction projects, firstly, and explain the network design problem of LTE railway network to coexist with LTE public safety network secondly, and finally discuss our conclusions and the future works.

2 Related Works

The need to upgrade the existing wired and wireless railway systems and the new interference cancellation techniques for LTE-based heterogeneous networks have already been addressed in many notable research projects. As the speed of train trialled by the French National Rail Corporation has reached more than 500 km/h, the secure and continuous connection to the train control system has attracted interest from railway operators [1–4].

Traditional high-speed railway communication engineering focuses its main attention to the frequent handover latency problem as well as network reliability such as the duplication of network elements [5–7]. According to Future Railway Mobile Telecommunication System (FRMTS) project launched in 2013 under International Union of Railway (UIC), it is necessary to examine the exclusive voice call function of railway, QoS management for train control, the performance of wireless communication at high-speed mobility of more than 350 km/h [2–4].

3 Current State and Progress of Two LTE-Based Network Construction Projects

In fact the railway operators in Europe, United States, Japan, and Korea have been trying to secure the dedicated frequency for the railway communication networks over the past decade since the wireless train control urgently needs the reliability

Table 1 Key features of two LTE-based networks

Parameter	LTE public safety network	LTE railway network
Frequencies	718–728 MHz for uplink, 773–783 MHz for downlink	
Target users	Military, police, fire, EMS, coast guard, provincial administrative offices, electricity, gas and the forest service	Automatic train control, railway personnel
Standard	3GPP LTE Rel.12, 13	UIC and 3GPP LTE
Data rate	25 Mbps Max. for UL, 75 Mbps Max. for DL	>GSM-R
Coverage mobility	0.55–3.3 km Less than 350 km	About 0.2–1 km More than 350 km
Features	GCSE, D2D ProSe, IOPS, Relay, MCPTT	Group call, broadcast call, railway emergency all, etc.

Table 2 Functional requirements for LTE based railway communication system

Item	Functional requirements
Voice call service	Individual voice call, public emergency call, broadcasting emergency call, group call, multilateral voice call
Data service	Multimedia message service, general data service, train control service
Video call service	Individual video call, group video call, video image transmission
Call function	CID filtering, call priority, exclusive user group
Railway service	Function addressing, position-based addressing, railway emergency call, shunting mode, direct communication

and safety of the railway operation [1–3]. Nevertheless, LTE-based railway network of Korea should share the same frequency bands with the public safety network on 718–728 MHz for uplink and 773–783 MHz for downlink, as shown in Table 1.

Though the public safety network will be compliant with the release 12 or 13 of 3GPP specification, the LTE railway network has not yet been internationally standardized. The railway network has its own features of high-speed mobility more than 350 km/h, data communication for train control, and railway emergency call as shown in Table 2. These railway-intrinsic requirements and reference model have been being standardized by Korean standard body, TTA as shown in Fig. 1 and Table 2 [8, 9].

4 Emerging Issues to Coexist with LTE Public Safety Network

The previously introduced several challenges of the LTE railway communication system toward next generation railway network must be addressed as follows [2–4].

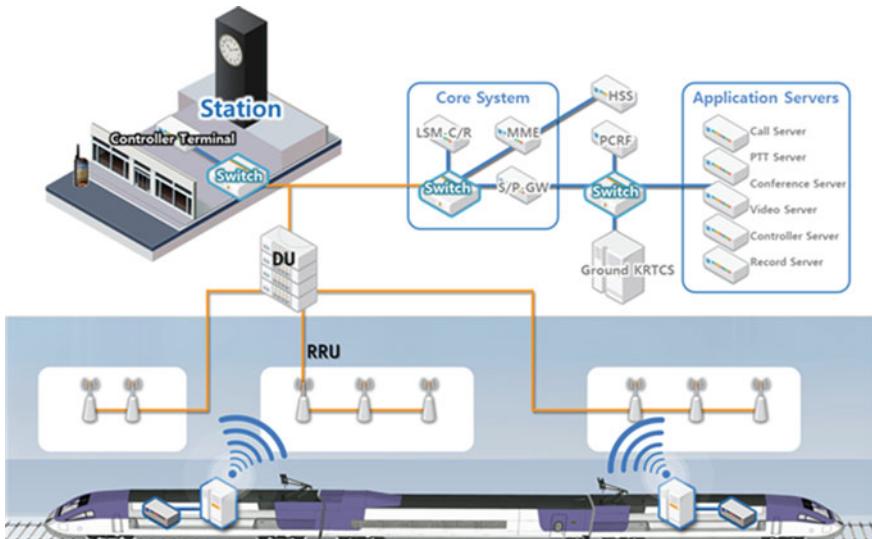


Fig. 1 Reference model of LTE-based railway communication system

- Network reliability and availability.
- Handover and network performance for high-speed train.
- QoS management to support the required railway functionalities.
- Exclusive voice service provision for railway.

In spite of what was expected if two networks would share the same frequency band, great concerns and practical researches about inter-network interference coordination between two LTE networks are urgently needed since the inter-network interference can cause the call disconnection with wireless train control which results in a fatal train accident. Thus it is most important to secure the reliability and survivability of the wireless train control system when we design and deploy both LTE networks together. While focusing on the coexistence and cooperation between two LTE networks of the faultless and continuous connection to train control, the next-generation LTE-based high-speed railway network topology could be one of four candidates as shown in Fig. 2.

If two LTE networks could be simply integrated, and had common e-NodeB such as Fig. 2a, b, inter-cell interference can be easily coordinated. But as previously mentioned railway network has exclusive user requirements and functional requirements and as a matter of fact the source of investment finance and network operator are all different each other. Hence, unfortunately there is no choice but to rule out the possibility of Fig. 2a, b. Subsequently, if the both networks were deployed such as Fig. 2c and connected via X2 interface, LTE heterogeneous network solutions can be considered to mitigate the radio interference among the adjacent cells, such as ICIC, eICIC and feICIC which mean inter-cell interference coordination, enhanced ICIC, and further enhanced ICIC respectively [10–15].

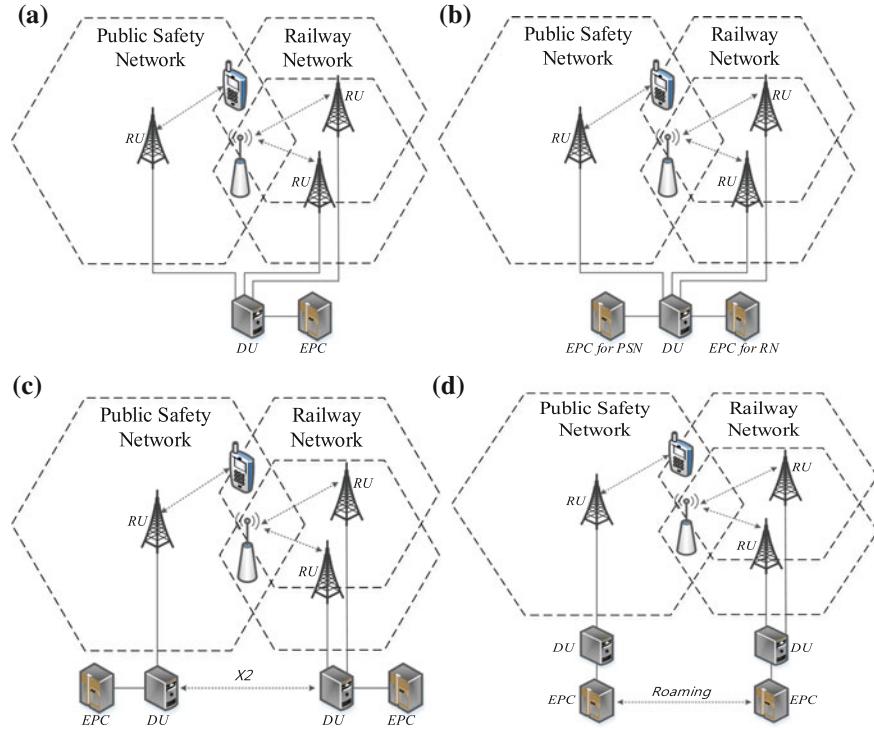


Fig. 2 Network topology candidates for LTE-based railway network interworking with LTE public safety network

However, these LTE heterogeneous network solutions via the X2 interface are not about the elimination of radio interference but reduction of it. Hence it may not be a good solution to design the faultless wireless train control network. In case of Fig. 2d which means there are no X2 interface between two networks, inter-network interference cannot be resolved automatically and consequently no one can be sure of the reliability of the train control.

Furthermore, the duplication of network elements and cell coverage in railway should be considered, which means the redundancy for each network element of the core network to RU and DU to avoid any single point of failure and ensure high reliability and survivability. As previously mentioned, in our view, it is the most important thing at present for designing and planning LTE railway network to eliminate—not to mitigate or reduce—the radio interference, especially from the adjacent LTE public safety network to LTE-based wireless train control on the same frequency band.

The QoS requirements for the high-speed train can be much higher than the existing commercial mobile communication network such as LTE or LTE-A and even the LTE public safety network, and should be satisfied with reliability, availability, safety, and security. And moreover bandwidth requirements for railway

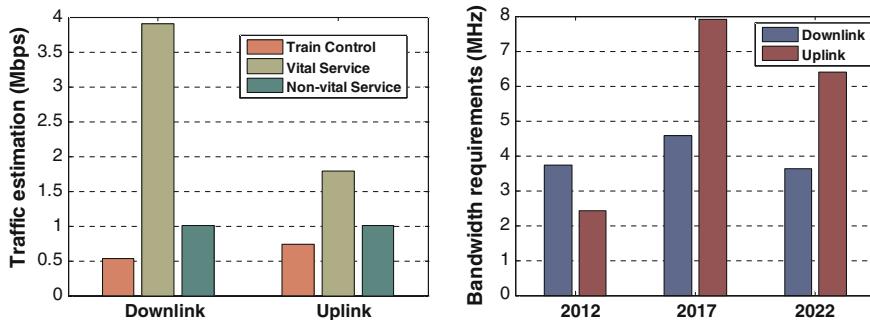


Fig. 3 Traffic estimates for railway network in 2012 (*left*) and bandwidth requirements (*right*)

network could be increased in 2017 as shown in Fig. 3, QoS management policies will need great concern together with interference management of these public safety and railway coexisting networks in near future [1].

5 Conclusions and Future Work

In this paper current status and progress of the public safety and railway coexisting networks are introduced, and several emerging issues of designing these two networks of the same frequency bands are discussed. In our view, interference management, bandwidth requirements, duplication of network elements, and QoS management policies between two networks will be the main emerging issues, and it is the most important thing at present for designing and planning LTE railway network to eliminate—not to mitigate or reduce—the radio interference especially from the adjacent LTE public safety network to LTE-based wireless train control on the same frequency band. And finally, QoS management policies will need great concern together with interference management of these public safety and railway coexisting networks in near future.

Acknowledgments This research was supported by a grant (14RTRP-B088444-01) from the Railway Technology Research Project funded by the Ministry of Land, Infrastructure and Transport (MOLIT) of the Korean government and by the Korea Agency for Infrastructure Technology Advancement (KAIA).

References

1. Jeong, M., Yoon, H., Park, D., Kim, K., Lee, S.: Bandwidth requirements estimation method for future wireless railway communication systems. *J. Korean Soc. Railway* **16**(6), 540–550 (2013)
2. Calle-Sanchez, J., Molina-Garcia, M., Alonso, J.I.: Top challenges of LTE to become the next generation railway communication system. In: Proceedings of the 13th computers in railways, pp. 85–96 (2012)
3. Ai, B., et al.: Challenges toward wireless communications for high-speed railway. *IEEE Trans. Intell. Transp. Syst.* **15**(5), 1–16 (2014)
4. Durk, J.: How to better utilise rail communications to improve the on-board customer experience. In: Special pre-conference LTE workshop (2012)
5. Ning, B., et al.: An introduction to parallel control and management for high-speed railway systems. *IEEE Trans. Intell. Transp. Syst.* **12**(4), 1473–1483 (2011)
6. Li, Z., et al.: Handoff performance improvements in MIMO-enabled communication-based train control systems. *IEEE Trans. Intell. Transp. Syst.* **13**(2), 582–593 (2012)
7. Karimi, O.B., et al.: Seamless wireless connectivity for multimedia services in high speed trains. *IEEE J. Sel. Areas Commun.* **30**(4), 729–739 (2012)
8. TTAK.KO-06.0369. Functional requirements for LTE based railway communication system (2014)
9. TTAK.KO-06.0370. User requirements for LTE based railway communication system (2014)
10. 3GPP TR 36.814, E-UTRA; Further advancements for E-UTRA physical layer aspects
11. Damnjanovic, A., et al.: A survey on 3GPP heterogeneous networks. *IEEE Wirel. Commun.* **18**, 10–21 (2011)
12. Lei, W., et al.: Heterogeneous network in LTE advanced system. In: IEEE international conference on communication systems, pp. 156–160 (2010)
13. Ha, J., et al.: Performance analysis of dynamic spectrum allocation in heterogeneous wireless networks. *ETRI J.* **32**(2), 292–301 (2010)
14. Hussein, Y.S., et al.: Reduction of outage probability due to handover by mitigating inter-cell interference in long-term evolution networks. *ETRI J.* **36**(4), 554–563 (2014)
15. Jang, U., et al.: Interference management with block diagonalization for macro/femto coexisting networks. *ETRI J.* **34**(3), 297–307 (2012)

Alternative Way to Manage the Research Documents

Jeong Ah Kim, Jae-Young Choi, Jong-Won Ko, Sun-Tae Kim and Young-Hwa Cho

Abstract For competing with other companies, IT companies are struggling to innovate their capabilities. Every companies invested research or research and development project for developing new technology and new products. Through the research process, there are so many valuable research deliverables. In this paper, we suggested to integrate the Product Data Management (PDM) system with Research Document Management (RDM).

1 Introduction

“Technology development” refers to a special type of development projects. This project produces new knowledge, new technology, a technical capability, or a technological platform [1]. Basic research or fundamental research project belongs to technology development. Technology development projects are special since there cannot be strict and complete business or technical requirements. Also this project are

J.A. Kim (✉)

Department of Computer Education, Catholic Kwandong University, 579 beongil-24,

BeomIlRo, KangNeungsi 210-701, Korea

e-mail: clara@cku.ac.kr

J.-Y. Choi · J.-W. Ko · Y.-H. Cho

College of Information and Communication Engineering, Sungkyunkwan University,

Suwon-si, South Korea

e-mail: jaeychoi@skku.edu

J.-W. Ko

e-mail: jwko0820@skku.edu

Y.-H. Cho

e-mail: choyh2285@skku.edu

S.-T. Kim

Department of Software Engineering, Chonbuk National University, Jeonju-si, South Korea

e-mail: stkim@jbnu.ac.kr

not predictable and have many high-risk unknowns. It also have great technical uncertainties. It is the reason that traditional development process or traditional project management cannot be applied into the technical development projects.

Product Data Management (PDM) is the discipline of controlling the evolution of a product design and all related product data during the full product life cycle, historically with the focus upon hardware product design. The aim to control and manage projects and products in PDM is to provide different disciplines with the accurate product information at the right time in the right format [2]. These feature are very close to the requirements of Research Document Management (RDM) system.

For research project management, we should clear the meaning of data, file, and document [3]. The smallest discrete entities which describe something in the real world and properties, characteristics, attributes, relationships, and objects are all “data” objects. We can also refer to data as “contents.” When logically collected together in an envelope or “container” of some sort, data elements are said to be contained within a file. Document is logical or conceptual object. What differentiates a document from a file is metadata. Metadata are the key to identify the documents. These features can be supported with PDM data model with business item or business data and part structure.

In this paper, we describe how much analogous between PDM and RDM so we identified the possibility of adoption of PDM for RDM. In Sect. 2, we discuss the R&D project management scope and previous approach for effective R&D project management systems. In Sect. 3, we explained the analogy between PDM and RDM in 2 perspectives: (1) scope of application, (2) data model. In Sect. 4, we describe the new approach of RDM based on PDM and describe the future implementation plans in Sect. 5.

2 R&D Project Management

2.1 R&D Process

Activities for R&D are performed iteratively and its purpose is to develop new products. These activities are (1) opportunity identification, (2) design, (3) prototyping, (4) release to market, and (5) feedback to each activities [4]. The most R&D organizations have adopted Stage-Gate process suggested in [5]. Stage-Gate has become a popular process for driving new products to market and it is a conceptual and operational map for moving new product projects from idea to launch. Stage-Gate process consists of 2 elements: (1) a series of states, each state defines the activities to gather the information by project team, the results of activities is analyzed, and produce the deliverables. (2) Followed by gates, where Go/Kill decision are made based on the deliverables of stages. This process is so flexible. Inside the stage, we can iterate the activities and back-and-forth play as the project proceeds [6]. Also, stage can be overlapped. Stage-Gate is a macro process—an overarching process and project management is micro process. Therefore

Table 1 R&D project management scope [8]

Area	Sub activities
Goal management	Establish the goals for each stage Manage the goals of each stage Approve the goals at each state
Progress management	Procedure of progress control Review the progress
Resource management	Manage the organization Allocate the human resources to project Manage the budgets Control the cost
Standard management	Establish the process Manage the research document Make the research assets Improve and deploy the process
Risk management	Identify the risks Analysis the risks Establish the risk management plan Control and monitor the risk
Evaluate the outcome	Establish the evaluation system Develop the culture or evaluation Track the assessment methods

Stage-Gate is not a substitute for project management. Rather Stage-Gate and project management are used together [6].

2.2 *PMS for R&D Projects*

R&D Project management systems support the R&D activities by providing the functionality of project planning, progress monitoring, resource management, data management as well as integrated management to increase the productivity and quality [7]. Actually Korean R&D management institute identified the R&D project management scopes as follows defined in Table 1.

2.3 *Limitation of PMS in Perspectives of Research Document Management*

To be more systematic and operate efficiently R&D administration, it should be converted by on-line administration way combined with IT technology rather than the existent off-line administration way [9]. Young [9] identified the difficulties of PMD in R&D organizations such as followings: (1) misunderstanding of administrative officer to PMS, (2) refusal against the automation system to researchers, and

(3) adoption strategy of PMS. In this paper, we identified the importance of adoption strategy of PMS to research laboratory since researchers want the advantages from the PMS not control mechanism of PMS first. Byun [10] considered the importance of data management in research laboratory and suggested the integration management system through introduction of research project management, raw data management, electronic laboratory notebook management and design of primary experimental process. In this paper, we also identified the importance of research data management since research data are not well-formed nor pre-defined. So we suggested new alterative way for management R&D project as to integrate with Product Data Management (PDM) system.

3 How Much Analogical Similarities Between PDM and RDM?

3.1 *In Perspectives of Purpose and Scope*

According to the definition by CimData [11], “Product Data Management (PDM) is a tool that helps people manage both product data and the product development process. PDM systems keep track of the masses of data and information required to design, manufacture or build, and then support and maintain products—whether your product is an aeroplane, petrochemical plant, highway, railway system, pharmaceutical, automobile, consumer product, or service. PDM is used effectively in a multitude of industries.” Products are often complex systems consisting of hardware, software, and related documents, developed by several groups. PDM have been focused on mechanical design, component classification and retrieval, revision tracking, workflow, sign-off control, version handling and parts of configuration management.

Research Data and Document Management (RDM) is also the discipline of controlling the evolution of research design and all related research data during the full research life cycle. For researcher, it is very important to provide correct and right information at right time with various suitable format. Research process produces various research document format such as word, presentation, spreadsheet, image, or well-structured document so on. These are very similar with elements of product which are hardware, software, and related documents. Also, these research deliverables should be identified and retrieved, tracked by revision, controlled with version, and also be the configuration item.

There are two function groups in PDM. First group is user functions what are Data vault and document management, workflow and process management, product structure management, part and component management (classification and retrieval), and program management. Second group is utility function group what consists of communications and notification, data transport and translation, image services, administration, and application integration. These functionalities defined in Fig. 1 can be the main features of RDM system.

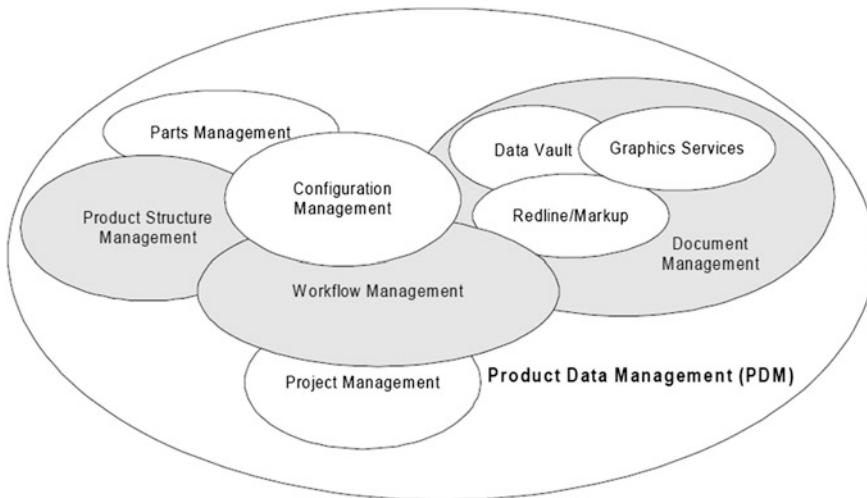


Fig. 1 Scope of PDM application [Nguyen, CIM]

3.2 *In Perspective of the Data Model*

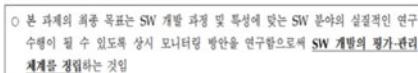
The information in a PDM system is based on an object oriented product data model. Objects used to represent parts, assemblies, documents and other kinds of objects. In PDM, these are called Business Items or Business Objects. Of course PDM system manages file as well and file can be considered as another kinds of Data Item. Relationships are used between objects to relate them to each other. These relationship can be represented by product structure which is another class of Business Items. A product structure defines a hierarchy of assemblies and components. An assembly might compose of other assemblies.

In research, it should be clear that files and documents are just container of data not data. Therefore, research data management system should be data-centric system and should manage the data directly. It is very similar features with those of PDM. In this paper, we call the research objects or research items which are analogy with business objects or business items (Fig. 2).

4 How to Integrate the PDM with RDM

To manage the research documents or artifacts, we need almost the functionalities of PDM. We need project management, configuration management, and workflow management of PDM what are the same functionalities in RDM. We also need document management, part management, and part structure management of PDM but their concept or management strategies are different in RDM. As described in 3,

Research File: Project Plan

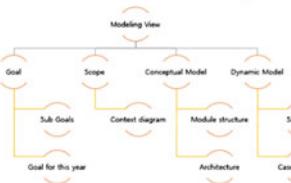


- 달해연도 목표는 「제반 연구정보의 유지관리 및 활용을 위한 Knowledge base 연구」로서 이를 위한 세부 목표는 (1) SW 연구개발 평가-관리를 위한 상시 모니터링 시스템 기능 추가-보완, (2) Knowledge 기반 정보 활용을 위한 통합 구조화 방안, (3) 도메인 인지형 SW 개발 프로젝워크 인프라 체계 구축, 그리고 (4) Integrated Test-bed 기반의 상호 연계형 시설 철차 및 기법 연구인

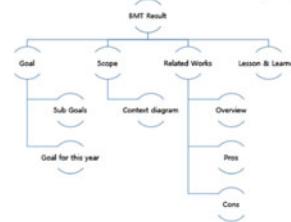
Research File: Conceptual Model



Research Document: Solution Reports



Research Document: Benchmarking report



Research Item

<Goal scope="final">
본 과제의 최종 목표는 SW 개발 과정 및 특성에 맞는 SW 분야의 실질적인 연구 수행이 될 수 있도록 상시 모니터링 방안을 연구함으로써 SW 개발의 평가·관리 체계를 정립하는 것임
</Goal>

『Goal, scope: "this year"』
단행예도 역시 "제반 연구정보의 유통관리 및 활용을 위한 Knowledge base 연구,로서 이를 위한 세부 목표는 (1) SM 연구개발 평가·관리를 위한 시스템 모니터링·설계·시행 기능 추가·보완, (2) Knowledge 기반 정보 활용을 위한 통합 구조화 방안, (3) 도메인 지향적 SM 개발 프레임워크 연구로 세부 구체화, 그리고 (4) Integrated Test-bed 기반의 상호 연계성 시험 체계 및 기법 연구원

```
<Model type="static architecture">
    <!-- 관리자 화면은 2개 화면의 복 기반 시스템으로 구축된다. 시스템 개발을 위해 Spring Framework는 기반으로 Presentation Layer, Business Layer, Data Access Layer로 구성된다.
    Presentation Layer는 사용자와 상호작용하는 계층으로, 사용자에게 맞는 화면을 제공하는 역할을 한다.
    Business Layer는 시스템을 제어하고 있는 Spring Web MVC를 기반으로 359장 357로 개발된 View, 사용자의 요청을 처리하기 위한 Controller 클래스를 포함된다. 그 외에도 시스템에 보안(Authentication & Authorization)을 위한 ace framework, 혼재, 레이아웃을 구성하는 위 시스템, Ajax 기반으로 웹을 확장하는 Ajax 기반 웹 서비스를 포함된다.
    Data Access Layer는 비즈니스로 제공하는 모듈을 포함하는 모듈로 포함된다.
    Business Layer는 비즈니스로 제공하는 모듈을 포함하는 모듈로 포함된다.
    Data Access Layer는 데이터베이스에 접근 가능한 기능을 제공하는 모듈을 포함하여, 이는 iBatis SQL Mapper나 MySQL JDBC 드라이버와 같은 데이터 접근 API를 포함한다.
    Data Access Layer는 데이터베이스에 접근 가능한 기능을 제공하는 모듈을 포함하여, 이는 iBatis SQL Mapper나 MySQL JDBC 드라이버와 같은 데이터 접근 API를 포함한다.
    Manager 클래스는 관리권 및 개인의 대화에 대한 관리를 담당할 수 있도록 적절한 코드로 분할하여 설계된다. 예를 들어 ProjectManager 클래스는 프로젝트 기본 정보 및 팀원정보를 처리하기 위한 서비스를 제공하는 모듈로 설계될 수 있다. DAO 클래스는 테이블과 1:1 관계로 설계하여, 해당 테이블에 대한 기본 CRUD 기능을 제공한다.
</Model>
```

Research Item

<Goal scope="final">
본 과제의 최종 목표는 SW 개발 과정 및 특성에 맞는 SW 분야의 실질적인
연구 수행이 될 수 있도록 상시 모니터링 방안을 연구함으로써 SW 개발의
평가·관리 체계를 정립하는 것임
</Goal>

<Goal, scope> this year>
단행해도 목표는 <제반 연구정보의 유지관리 및 활용을 위한 Knowledge base 구조>, 여기서是以 위한 세부 목표는 (1) SM 연구개발 평가·관리를 위한 상시 모니터링 시스템 기능 추가·보완, (2) Knowledge 기반 정보 활용을 위한 통합 구조·방안, (3) 도메인 지향적 SM 개발 프레임워크 양립·제공에 계획·구축, 그리고 (4) Integrated Test-bed 기반의 상호 연계성 시스템 철자 및 기법 연구임
</Goal>

Fig. 2 Research document, research item, research structure

2, research document management should be data-centric not document centric. To meet this features, we revise the document management, part management, and part structure management like in Fig. 3.

Data-centric document management consists of research item repository for storing the research item and research item markup for identifying the research item from the research file. Research item is the basic unit to classify the content, to retrieve the content, to reuse the content, to track the progress, and to test the research. Figure 4 described the revised data model for RDM. We classified the research files and research documents. Research item and research document should be identifiable with meta data and should be under the configuration management. Research structure define the structure composing the research item which is analogous with part structure of PDM.

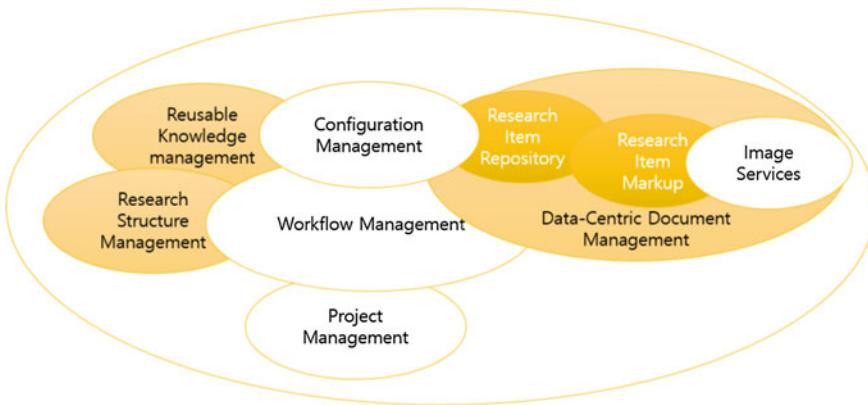


Fig. 3 Similarities and differences between PDM and RDM

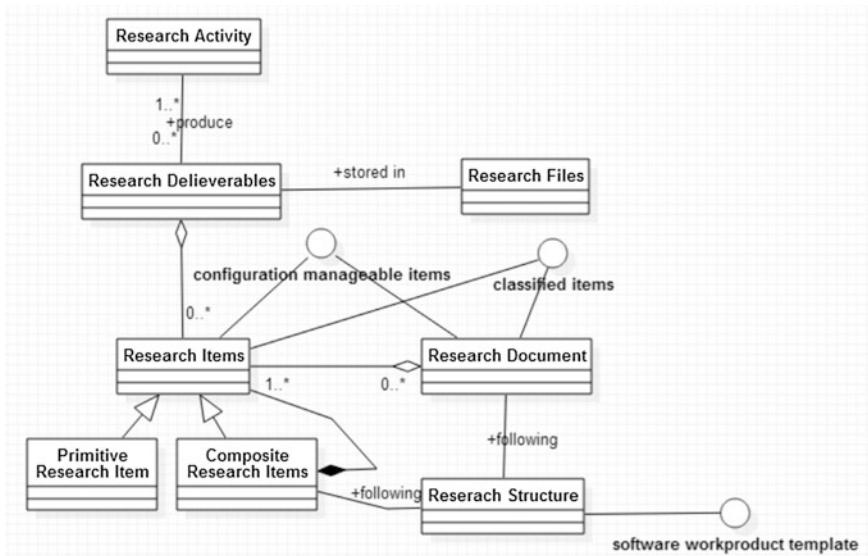


Fig. 4 Revised data model of RDM

5 Conclusions

In the paper, we suggested to integrate PDM for management the search document. For this, we analyze the similarities and differences between PDM and RDM. Almost functionalities of PDM can be the scope of RDM application. Also data model of PDM including of business item can satisfied the features of research data management. Based on the PDM data model, we revised the model for data-centric

management rather than document-centric management. With this suggestions, next step will be implementation of RDM based on PDM systems.

Acknowledgements This research was supported by the Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (NRF-2014M3C4A7030503).

References

1. Cooper, R.G.: Managing technology development projects. *IEEE Eng. Manag. Rev* **35**(1), 67 (2007)
2. Nguyen, T.N.: A unified model for product data management and software configuration management. In: *IEEE/ACM international conference on automated software engineering* (2006)
3. InterGraph. Don't be "content" with just containers. Document management in a data-centric world white paper, can be access at <http://www.intergraph.com/assets/plugins/ppmcollaterals/files/DocumentManagement.pdf>
4. Jeong, W., Moon, S.: Introduction to execution strategy and case study for R&D process innovation. *SAMSUNG SDS Consulting Review*, No. 1 (2007)
5. Cooper, R.G.: *Winning at new products: accelerating the process from idea to launch*, 3rd edn. Perseus Books, Reading (2001)
6. Cooper, R.G.: The stage-gate idea-to-launch process-update, what's new and nexgen systems. *J. Prod. Innov. Manag* **25**(3), 213–232 (2008)
7. Korea Industrial Technology Association. *R&D Project management* (2007)
8. Cho, Y.R.: On the development of R&D quality management system through project management approach. *Master Degree of Industrial Engineering at Ajou University* (2012)
9. Young, H.K.: Development and implementation of R&D project management system. In: *Thesis of Master Degree at The Graduate School of Industrial Engineering, Management, and Design, Hanyang University* (2008)
10. Byun, J.: (A) Study on design of a laboratory information management system for R&D knowledge management. *Thesis of Master Degree at The Graduate School of Industrial and Information System Engineering, Hanyang University* (2010)
11. CIMdata. Inc. <http://www.cimdata.com/>

R&D Project Management Using Context in Document: Research Descriptor

**Jong-Won Ko, Jae-Young Choi, Jung-Ah Kim, Sun-Tae Kim
and Young-Hwa Cho**

Abstract Recently developed requirement management tools or configuration management tools identify the requests within the project and offer traceability for the configuration items related to requests throughout the cycle or offers links to tasks or issues related to configuration items in addition to the identified configuration item, or to change requests or versions of specific files so that a more software engineer approach is possible in project management. This paper used XML to define the classification criteria for the types of data, the traceability of the Research Descriptor, and the document RD and management RD samples as a form of the Research Descriptor. The Research Descriptor based project management approach that can offer semantic analysis of the generated WorkProducts and their content, a semantic-based test, traceability and monitoring of project quality metric allows it to be applied to even R&D projects that are ambiguous in their goals or solutions.

Keywords Project management approach · Research descriptor · Software traceability

J.-W. Ko · J.-Y. Choi (✉) · Y.-H. Cho

College of Information and Communication Engineering, Sungkyunkwan University,

Suwon-si, South Korea

e-mail: jaeychoi@skku.edu

J.-W. Ko

e-mail: jwko0820@skku.edu

Y.-H. Cho

e-mail: choyh2285@skku.edu

J.-A. Kim

Department of Computer Education, Catholic Kwandong University, Gangneung-si,
South Korea

e-mail: clara@chu.ac.kr

S.-T. Kim

Department of Software Engineering, Chonbuk National University, Jeonju-si, South Korea
e-mail: stkim@jbnu.ac.kr

1 Introduction

Approaches traditionally taken to project management have been developed in various forms, from the waterfall model that works under the assumption that there are minimal changes and agile methodology that allows for a prompt response to changes.

R&D projects, unlike existing system integration projects or software development projects, are ambiguous in their goal or the resolution methods in order to reach that goal. It also has limits in having applied the waterfall model, iteration and gradual spiral models, or agile methodology which are usually applied to system integration projects or software development projects. In order to overcome these constraints, this paper defined the various forms of work products generated across the entire cycle of the R&D project, their major keywords and core sentences as the Research Descriptor, and through a project management based on the content, an analysis of the semantics of the content that composes the WorkProducts, a semantic-based test and traceability were supported. By monitoring the project quality metric, the paper suggests a Research Descriptor based project management approach that can be applied to R&D projects with ambiguous goals or resolution methods.

Chapter 2 describes the project management approach by project type used in existing studies, followed by Chap. 3 where one of the core concepts of Research Descriptor based Project Management Approach suggested by this paper, the Research Descriptor, is described. Chapter 4 concludes the suggestions in this paper and describes follow-up studies.

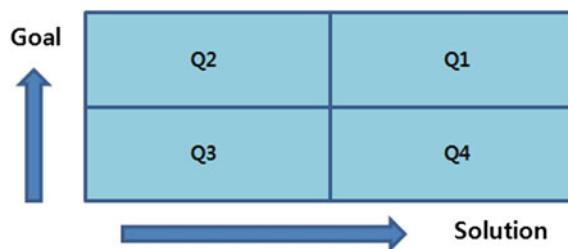
2 Classification of Project Management Approaches by Wisoki

This chapter reviews existing studies on project management methods and the four categories of project types by the clearness of the goal and the solution.

As seen in Fig. 1, project types are classified from Q1 to Q4, with Q1 type having a clear goal and solution, low uncertainty and small-scale improvements. Management focused on planning is appropriate for this type and since the goal and solution are clear a detailed plan for the timing of the launch of the project can be established. The project can be controlled based on whether the plan is being adhered to.

Q2 type has clear goals but the solution is not clear. Unexpected events can happen or unverified technology can be adopted, and there are many links both internally and externally, leading to high complexity. Therefore, a flexible and light process that can cover various situations is effective. High uncertainty and a lot of re-work means that plans change frequently and product-related risks are higher than Q1 project. It is often referred to as Just In Time (JIT) Plan.

Fig. 1 Project type: goal and solution by Wisoki



Q3 type projects have ambiguous goals and solutions and can often be seen when items for venture businesses are identified or R&D projects are undertaken. Lastly, Q4 projects apply to R&D projects where a new technology with a clear solution has been discovered but a business area where the technology can be applied isn't clear.

3 Conceptual Model of Research Descriptor

This chapter looks at the research descriptor, which is the core concept of Research Descriptor Based Project Management Approach suggested by this paper. The reason why the concept of research descriptor is needed in R&D Project Management, what types of Research Descriptors there are, how traceability is supported and how XML expression is implemented are described in this chapter.

3.1 *Why the Research Descriptor Is Required for R&D Project Management?*

The various text files generated throughout the entire R&D project cycle are recognized as configuration items in units of files and were managed as part of configuration management along with source codes. However, management in units of files led to the files not being supported from a system perspective through verification on consistency and completeness, but rather the researcher himself who is an expert in the given field of the R&D project had to conduct the verification himself. Therefore, the work products generated throughout the entire cycle of the R&D project is managed not based on files but based on keywords or core sentences, and based on the technology that analyzes the semantics, a verification on consistency and completeness is carried out. Moreover, the work products of the verification, too, need to be reflected onto the process of the R&D project by being stored as data in the system, and as such, the various forms of work products generated throughout the R&D project and various forms of information to test these WorkProducts are defined as the Research Descriptor. Therefore, unlike

existing project management approaches, by applying the concept of Research Descriptor as suggested in this paper, the texts generated during the project and their content can be sufficiently reflected onto the development of source codes and quality indices of the project, which was not possible with previous methods. Although this is an approach from the perspective of software engineering, it offers a more effective R&D project management approach compared to existing methods.

3.2 *Classification of Research Descriptor*

As noted in Sect. 3.1, the concept of Research Descriptor, which is at the core of the Research Descriptor based project management approach, receives various information throughout the entire cycle of the R&D project. Depending on the information that the Research Descriptor must have, the Research Descriptor can be classified as seen in Table 1.

First, Document RD refers to the various forms of text files generated over the entire course of the project including major research WorkProducts such as research plan or annual performance reports. Contents RD refers to the content of the text file, that is, the major keywords or core sentences that can be the subject of the semantic-based tests or traceability. Contents RD can be defined as the term RD Item.

Research Descriptor that stores the management information that needs support in order to carry out the Research Descriptor based project management approach, such as the rules on traceability, quality metrics for monitored quality metrics, indices for the consistency and completeness required to conduct semantic-based tests, and RD template that is supported by the system, are defined as Management RD.

Table 1 Classification of research descriptor

Research descriptor	Description
Document RD	Various forms of text files generated throughout the project including major research results ex. research plan, meeting logs, compilation of research content, research notes, etc.
Contents RD	Major keywords or core sentences that can be the subject of semantic-based analysis, semantic-based tests or traceability. Defined as RD Item from the perspective of Document RD
Management RD	RD that compiles the support information required for the R&D project to be implemented smoothly. ex. Rules on traceability, metrics rules for monitored quality metric, quality metrics rules in semantic-based tests, information on the RD template
Process RD	Research descriptor that defines the project's WBS (Phase-Activity-Task)
Software RD	Requirement ID, use case, class, User Interface form ID and test case generated in relation to the development of software within the R&D project

3.3 Research Descriptor Traceability

The Research Descriptor generated in the R&D project has traceability to support the semantic analysis technology, similarity analysis, or semantic-based tests. The traceability for Research Descriptor can be defined as follows from four perspectives.

- Traceability from the perspective of work products: Traceability between Document RD which is the result in the form of document files Tracing of various versions for one document file and traceability for other related document files
- Traceability from the perspective of process: Tracing of relation between the generated Document RD and any tasks or activities on the project WBS
- Traceability from the perspective of the monitoring quality metrics: Tracing of Document RD that affects monitoring quality metrics
- Traceability from the perspective of software development: Tracing of components or class related to software development associated with generated Research Descriptor or source files

Figure 2 describes the traceability of the Research Descriptor from the perspective of WorkProducts.

3.4 Research Descriptor and XML Expression

The Research Descriptor based project management approach suggested in this paper stores data on the Research Descriptor generated during the R&D project

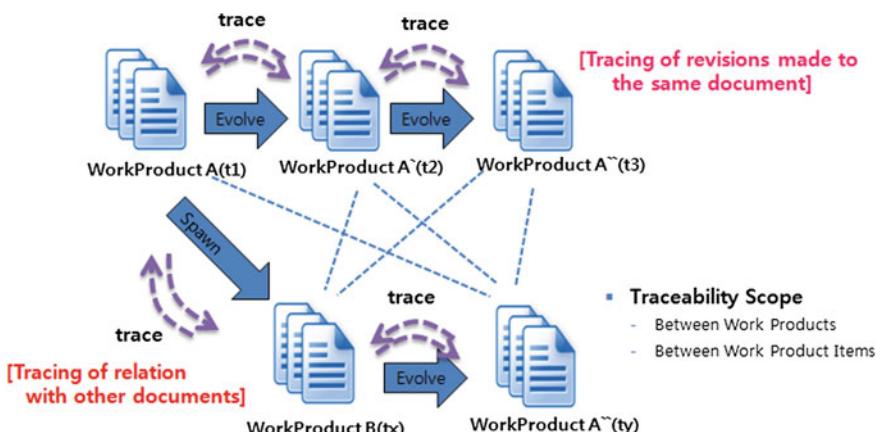


Fig. 2 Research descriptor traceability: WorkProduct perspective

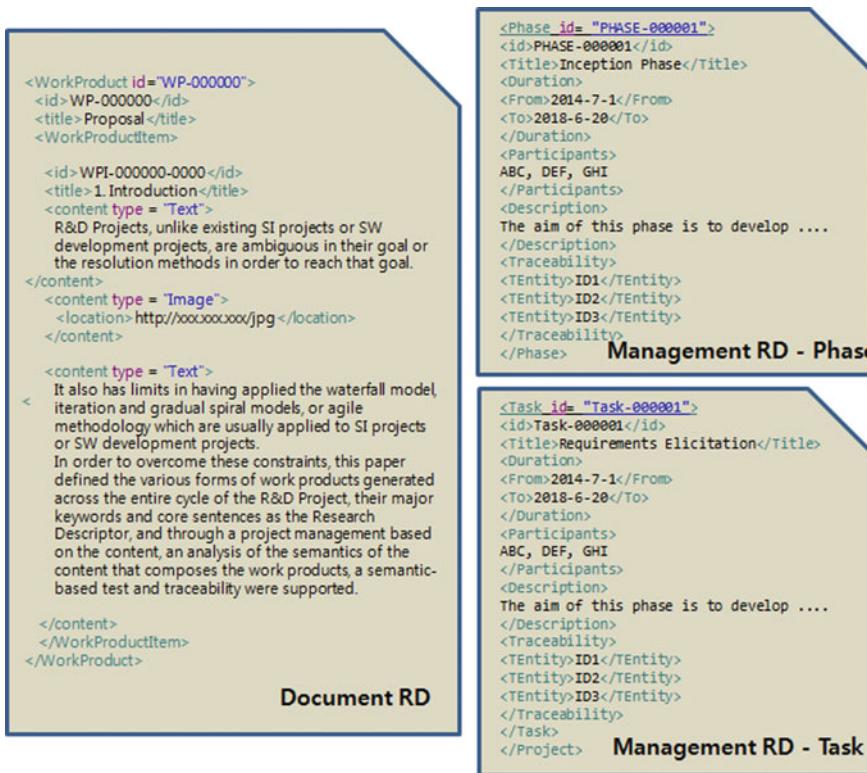


Fig. 3 XML expression for research descriptor: document RD and management RD

onto its Knowledge Repository. To do this, the content of the Research Descriptor and related information are expressed as XML data. Figure 3 shows how the document in the form of a proposal, as an example of a Document RD, can be expressed using an XML Tag.

Using a content type Tag, contents and images in the form of text are expressed. In addition, components of documents such as a table using a content type can be defined. By distinguishing between WorkProduct and WorkProduct Item, they can be identified as Document RD and Contents RD, respectively. Tags such as ID or Title are used in both WorkProduct and WorkProduct Item.

In Fig. 3, as an example of Management RD, the Research Descriptor that stores the information of the phase and task on the WBS is expressed using an XML Tag. Here, as is the case with Document RD, the ID and Title Tags are used to identify the basic information on the phase and task and such processes are expressed using the 'from' and 'to' of the Duration Tag while they are carried out. In addition, using the Description Tag, the explanation on the process is stored, and in order to support traceability, an ID for traceability can be saved using the TEntity tag of the Traceability Tag.

4 Conclusion and Further Works

This paper discussed the various forms of WorkProducts generated throughout the entire cycle of R&D projects, their major keywords and core sentences. The Research Descriptor is where such information is stored. This paper used XML to define the classification criteria for the types of data, the traceability of the Research Descriptor, and the document RD and management RD samples as a form of the Research Descriptor. The Research Descriptor based project management approach that can offer semantic analysis of the generated WorkProducts and their content, a semantic-based test, traceability and monitoring of project quality metric allows it to be applied to even R&D projects that are ambiguous in their goals or solutions. For this approach, more in-depth research on the Research Descriptor generator that converts and generates the Research Descriptor, on the R&D project quality metric, on Research Descriptor traceability, and on the recommendation of appropriate processes through the monitoring of, R&D project quality metrics are currently underway. Also being conducted are researches on semantic-based tests on the Research Descriptor, the Knowledge Repository that stores the test and Research Descriptor, and re-use framework including the Research Descriptor search engine.

Acknowledgements This research was supported by the Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (NRF-2014M3C4A7030503).

Research Descriptor Based Project Management Approach for R&D Projects

**Jong-Won Ko, Jae-Young Choi, Jung-Ah Kim, Sun-Tae Kim
and Young-Hwa Cho**

Abstract R&D projects, unlike existing System Integration (SI) projects or software development projects, are ambiguous in their goal or the resolution methods in order to reach that goal. It also has limits in having applied the waterfall model, repeated and incremental spiral models, or agile methodology which is usually applied to System Integration projects or software development projects. In order to overcome these constraints, this paper defined the various forms of work products generated across the entire cycle of the R&D project, their major keywords and core sentences as the Research Descriptor, and through a project management based on the content, an analysis of the semantics of the content that composes the WorkProducts, a semantic-based test and traceability were supported. By monitoring the project quality metric, the paper suggests a Research Descriptor based project management approach that can be applied to R&D projects with ambiguous goals or resolution methods.

Keywords Project management approach · Research descriptor · Software traceability

J.-W. Ko · J.-Y. Choi (✉) · Y.-H. Cho

College of Information and Communication Engineering, Sungkyunkwan University,

Suwon-si, South Korea

e-mail: jaeychoi@skku.edu

J.-W. Ko

e-mail: jwko0820@skku.edu

Y.-H. Cho

e-mail: choyh2285@skku.edu

J.-A. Kim

Department of Computer Education, Catholic Kwandong University, Gangneung-si,
South Korea

e-mail: clara@chu.ac.kr

S.-T. Kim

Department of Software Engineering, Chonbuk National University, Jeonju-si, South Korea
e-mail: stkim@jbnu.ac.kr

1 Introduction

Approaches traditionally taken to project management have been developed in various forms, from the waterfall model that works under the assumption that there are minimal changes and agile methodology that allows for a prompt response to changes. These approaches have been applied to various forms of software development project. Along with such methodology, case tools are used in various fields such as requirement management, configuration management, time management, scope management, risk management, quality management and comprehensive management. Recently developed requirement management tools or configuration management tools identify the requests within the project and offer traceability for the configuration items related to requests throughout the cycle or offers links to tasks or issues related to configuration items in addition to the identified configuration item, or to change requests or versions of specific files so that a more software engineer approach is possible in project management.

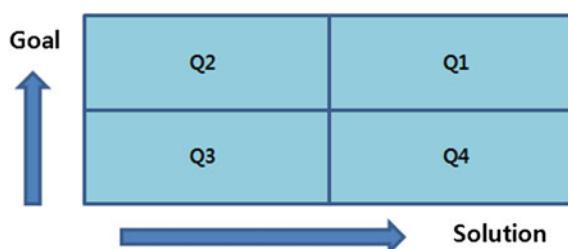
Chapter 2 describes the project management approach by project type used in existing studies, Chap. 3 describes the scenario from the perspective of support for the system, the researcher, and the proposed approach method. Chapter 4 concludes the suggestions in this paper and describes follow-up studies.

2 Classification of Project Management Approaches by Wisoki

This chapter reviews existing studies on project management methods and the four categories of project types by the clearness of the goal and the solution.

As seen in Fig. 1, project types are classified from Q1 to Q4, with Q1 type having a clear goal and solution, low uncertainty and small-scale improvements. Management focused on planning is appropriate for this type and since the goal and solution are clear a detailed plan for the timing of the launch of the project can be established. The project can be controlled based on whether the plan is being adhered to.

Fig. 1 Project type: goal and solution by Wisoki



Q2 type has clear goals but the solution is not clear. Unexpected events can happen or unverified technology can be adopted, and there are many links both internally and externally, leading to high complexity. Therefore, a flexible and light process that can cover various situations is effective. High uncertainty and a lot of re-work means that plans change frequently and product-related risks are higher than Q1 project. It is often referred to as Just In Time (JIT) Plan.

Q3 type projects have ambiguous goals and solutions and can often be seen when items for venture businesses are identified or R&D projects are undertaken. Lastly, Q4 projects apply to R&D projects where a new technology with a clear solution has been discovered but a business area where the technology can be applied isn't clear.

3 Research Descriptor Based Project Management Approach

This chapter reviews how the researcher or evaluator, while conducting an R&D project, can use the Research Descriptor as explained in Chap. 3, can carry out content-based project management. The overall concept of Research Descriptor based project management, scenarios from the perspective of the researcher and system support are reviewed, major usecases are identified and a detailed diagram for each usecase are discussed.

3.1 Research Descriptor Based Project Management: Overview

The Research Descriptor based project management approach suggested in this paper is linked with the re-use framework that is based on the Research Descriptor or semantic-based tests that are based on semantic analysis technology, support for traceability and configuration management based on the content of the same file or major keywords or core sentences. This is in contrast with existing project management systems or configuration management systems that supported the version management and traceability between files, as well as configuration management based on files such as document files or source codes generated while the project is carried out.

Figure 2 shows an overview of the Research Descriptor based project management approach. The upper part of the figure shows the work breakdown structure for R&D planning, R&D itself, R&D verification, and R&D management, and defines major artifacts, while carrying out each stage that are iterative and incremental processes.

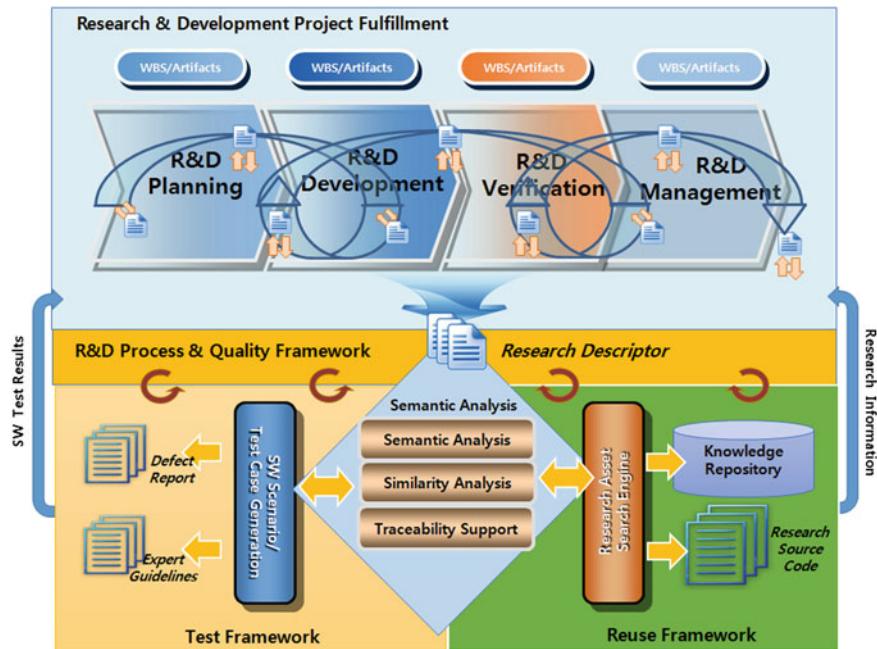


Fig. 2 Research descriptor based project management: overview

When each stage is implemented, the Research Descriptor is generated in the form of a research note, and the generated Research Descriptor, too, carries out an iterative and incremental process, through which it evolves as shown in Fig. 2.

As with the test framework that carries out a semantic-based test on the Research Descriptor generated through semantic analysis, similarity analysis or traceability support, the research asset search engine for re-use through semantic analysis technology, and the re-use framework that conducts re-use searches on the Knowledge Repository, Research Descriptor and the source code are linked with the R&D Process & Quality Framework to carry out the R&D project.

3.2 *Research Descriptor Based Project Management: Researcher and System Support Perspective Scenario*

The scenario for the Research Descriptor based project management from the perspective of the researcher and System Support is as seen in Fig. 3. First the researcher registers as a user on the system, logs in and establishes a plan for the Work Breakdown Structure of the R&D project. At this point, the information on

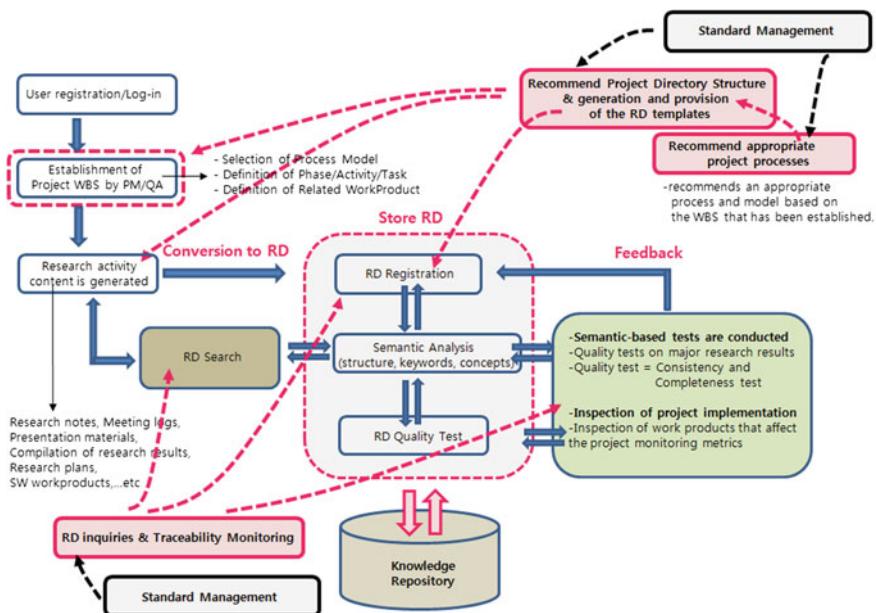


Fig. 3 Research Descriptor (RD) based project management: researcher and system support perspective scenario

the process model to be applied to the system and the structure information for the phase, activity and task are received. If major WorkProducts are generated as a result of carrying out an activity or a task, the components or the text format of major WorkProducts are also received.

Once all of the WBS of the R&D project are established, the researcher generates the content of the research activity. The content of the research activity refers to the research notes, meeting logs, presentation materials, research plans or WorkProducts of the software development.

The research activity content thus generated is converted to Research Descriptor and registered as Research Descriptor to be stored in the Knowledge Repository. The methods of converting to Research Descriptor are as follows: the hwp file, doc file or ppt file that has already been generated as part of the research content can be converted into Research Descriptor, an Research Descriptor generator can be used to directly generate Research Descriptor, or major content can be selected from existing files to be stored as Research Descriptor. When registering as Research Descriptor, an Research Descriptor quality test can be conducted based on the semantic analysis on the Research Descriptor structure, major keywords or concepts in order to enable content-based Research Descriptor management, traceability and tests.

3.3 *Research Descriptor Based Project Management: Design Model*

Based on the scenarios of the Research Descriptor based project management from the perspectives of the researcher and system support, this chapter identifies the major actors and use cases and presents their relationships using an UML usecase diagram. In addition, for major usecases, through the activity diagram, a detailed scenario for the interior of the usecase is defined.

Figure 4 shows that as in the perspective applied in the scenario of the Research Descriptor based project management, the actors are identified as the researcher and the evaluator, PM, QA, the researcher that carries out standard management and the system administrator that carries out the processing of user registration and authorization management. Figure 5 shows an activity diagram that describes the detailed scenario for the establishment of the project WBS.

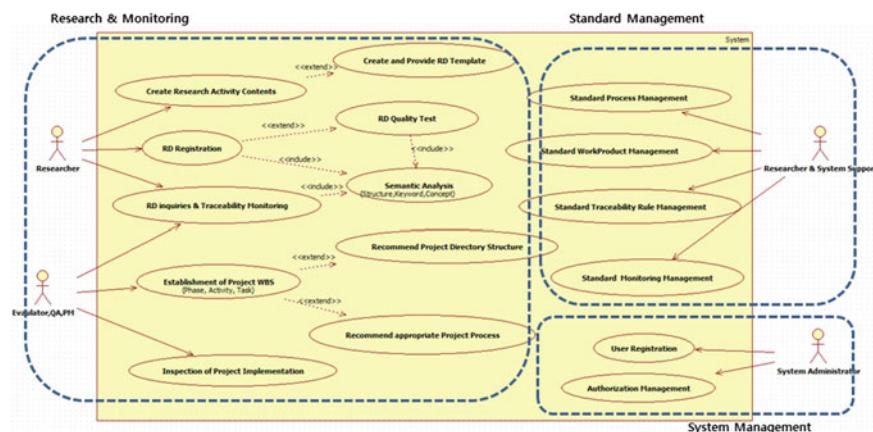


Fig. 4 Research descriptor based project management: usecase diagram

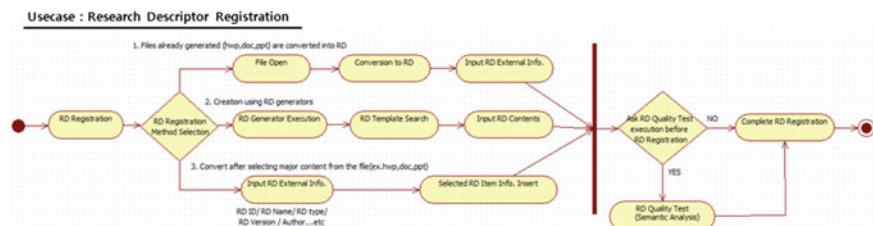


Fig. 5 UC—research descriptor registration: activity diagram

4 Conclusion and Further Works

The Research Descriptor based project management approach that can offer semantic analysis of the generated WorkProducts and their content, a semantic-based test, traceability and monitoring of project quality metric allows it to be applied to even R&D projects that are ambiguous in their goals or solutions. For this approach, more in-depth research on the Research Descriptor generator that converts and generates the Research Descriptor, on the R&D project quality metric, on Research Descriptor traceability, and on the recommendation of appropriate processes through the monitoring of, R&D project quality metrics are currently underway. Also being conducted are researches on semantic-based tests on the Research Descriptor, the Knowledge Repository that stores the test and Research Descriptor, and re-use framework including the Research Descriptor search engine. Going forward, further research is scheduled on the Research Descriptor based project management approach as suggested in this paper for the design and implementation of a framework that supports the R&D process.

Acknowledgements This research was supported by the Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (NRF-2014M3C4A7030503).

Attributes for Characterizing Java Methods

Illo Lee, Suntae Kim, Sooyong Park and Younghwa Cho

Abstract This paper proposes the list of attributes for characterizing Java Methods. The list consists of 67 attributes with three categories of vectors: a method signature, a behavior of a method and an association between the signature and the behavior. In addition, this paper introduces an approach to extracting the characteristic vector from the source code corpus, and presents representative vectors after examining all methods from 10 open source projects to show how well the vector characterizes the validity of method identifiers.

1 Introduction

In a programming language, a method is the minimal unit of naming, executing the intended behavior of the system [4]. A method declaration consists of two sections: a method signature containing method identifier, parameters and a return type, and a method body section that realizes its intention. Particularly, a developer as an author of the method tries to clearly deliver the method's intended behavior by

I. Lee

Defense Agency for Technology and Quality, Seoul, South Korea
e-mail: 215-b@hanmail.net

S. Kim (✉)

Department of Software Engineering, Chonbuk National University, 567 Baekje-daero,
Deokjin-gu, Jeonju-si, Jeollabuk-do 561-756, South Korea
e-mail: stkim@jbnu.ac.kr

S. Park

Department of Computer Science and Engineering, Sogang University, Seoul, South Korea
e-mail: sypark@sogang.ac.kr

Y. Cho

College of Information and Communication Engineering, SungKyunKwan University, 2066
Seobu-ro, Jangan-gu, suwon, Gyeonggi-do, South Korea
e-mail: choyh2285@skku.edu

appropriately naming the method signature [6]. In contrast, a collaborator as a code reader tries to comprehend the author's intention based on the method signature in a short period of time.

There are some work for characterizing the method behavior based on the method signature and implementation (see [1, 3, 4, 7]). Most of the research is based on the nano pattern, which is a collection of attributes of a method implementation. However, most of the approaches tend to focus on attributes of the method body without dealing with the relationship between the method signature and the method body. Thus, the vectors has a limitation for applying it to mining relationships between them.

In order to address the issue, we suggest an enhanced characteristic vector (hereafter we call it as C-Vector). We defined the C-Vector with three categories: the C-Vector for a method signature, a method body, and relationships between them. Use of the C-Vector enables one to express the characteristics of the method, also it is utilized to cluster methods and compare a characteristics of a method with others. In this paper, we illustrate the list of C-Vectors after extracting them from ten popular Java based open sources.

The remainder of this paper is structured as follows. Section 2 introduces related works for characterizing method attributes. Section 3 proposes three types of C-Vectors. Section 4 introduces representative C-Vectors extracted from 10 open source projects. Section 5 concludes a paper.

2 Related Work

Some researchers tried to study for characterizing the method behavior based on the method signature and body. Gil and Maman [3] proposed 27 micro patterns to analyze the characteristics of the Java classes and interfaces. Although they newly suggested the term 'nano pattern' for defining the traceable patterns between classes/interfaces and their contained methods, it is insufficient to handle the implicit relationship between a method signature and its implementation.

Host and Ostvold [4] suggested a method for detecting naming bug by applying association rules into mining relationship between a verb phrase of a method signature and characteristics in the method implementation section. They introduced 31 method attributes with six categories according to types of the method behaviors. In addition, the attributes can be extracted from the Java byte code.

Singer et al. [7] enhanced Host and Ostvold' approach with removing redundant patterns and adding new 17 patterns for coping with array types. Arnaoudova et al. [1] proposed linguistic anti patterns in the code level for hindering a developer from improving its quality.

3 C-Vectors of Java Methods

A method is composed of the method signature containing a return type, method identifier and parameters, the method body implementing the goal of the method. A method identifier can be separated into the verb part and the object part. For example, the method identifier `createSomething()` can be divided into the verb part `create` and the object part `Something`. The verb part is generally tightly related to the behavior of the method, while the object part is directly related to the data that the method manipulates. Thus, the behavior of the method `createSomething()` is a step-wise process for creating an object, and the target for creating should be `Something` as the object part mentioned.

3.1 C-Vectors for a Method Signature

The method signature is composed of a return type, a method identifier and a list of parameters. In order to characterize the method signature, we propose the following C-Vector as shown in Table 1. In the method identifier, we assume that it observes the Java Naming Convention [2] suggesting that the method identifier should be a verb (phrase). In addition, it might have an adjective or objects. Although it might contain diverse POS (Part-Of-Speech)es (e.g., adverb), we ignored it due to its frequency. The data type of all vectors are the boolean type (true/false).

3.2 C-Vectors for a Method Body

The method body contains the behavior of the method, which is represented in the verb part of the method identifier. Our paper selectively applied the Nano-Pattern [5] for defining the C-Vector for the method body. The nano pattern is categorized into four, containing 17 attributes as shown in Table 2. The data type of each attribute is boolean.

Table 1 C-Vectors for a method signature

Method identifier			Number of parameters		
Verb	Adjective	Object	None	Single	Multi
T/F	T/F	T/F	T/F	T/F	T/F
<i>Return type</i>					
Void	Boolean	Integer	Reference		
T/F	T/F	T/F	T/F		

Table 2 C-Vectors for a method body [5]

Category	Attributes
Calling	NoParams, NoReturn, Recursive, SameName, Leaf
Object-orientation	ObjectCreator, FieldReader, FieldWriter, TypeManipulator
Control flow	StraightLine, Looping, Exceptions
Data flow	Local Reader, Local Writer, ArrayCreator, ArrayReader, ArrayWriter

3.3 C-Vectors for Relationships Between a Method Signature and a Method Body

In the method signature, the object part is generally associated with the data that the method body manipulates. Also, the parameters must be used in the specific position in the method body. As an example, the method `createSomething()` is intended to carry out the behavior for creating an object, which is generally expressed in `Something s = new Something();` in Java. In the example, we can recognize that the object part `Something` is positioned after the `new` operator. Based on this observation, we defined the C-Vector capturing the relationship between the object part and the method body, and the relationship between the parameters and the method body.

The C-Vector for the relationships is classified into two : the *include* C-Vector showing if the object part and the parameters are included in which part of the method control sequence, and the *target* C-Vector for capturing where it is located among diverse readers and writers as shown in Tables 3 and 4. Particularly, the target C-Vector is based on the nano-pattern, as it is defined in the behaviors of the method body.

Table 3 C-Vectors for *Include* relationships

Category	Attributes
Method identifier's object	<code>Obj_InReturnVariable, Obj_InReturnType,</code> <code>Obj_InReturnVariable, Obj_InReturnType,</code> <code>Obj_InParameter, Obj_InParameterType,</code> <code>Obj_InSubMethodName, Obj_InSubMethodParam,</code> <code>Obj_InIfBranch_Condition, Obj_InIfBranch_Body,</code> <code>Obj_InForLoop_Condition, Obj_InForLoop_Body,</code> <code>Obj_InWhileLoop_Condition, Obj_InWhileLoop_Body</code>
Parameters	<code>Param_InReturnVariable, Param_InReturnType</code> <code>Param_InSubMethodName, Param_InSubMethodParam</code> <code>Param_InIfBranch_Condition, Param_InIfBranch_Body</code> <code>Param_InForLoop_Condition, Param_InForLoop_Body</code> <code>Param_InWhileLoop_Condition, Param_InWhileLoop_Body</code>

Table 4 C-Vectors for *Target* relationships

Category	Attributes
Method identifier's object	Obj_OfObjectCreator, Obj_OfFieldReader, Obj_OfFieldWriter, Obj_OfTypeManipulator, Obj_OfLocalReader, Obj_OfLocalWriter, Obj_OfArrayCreator, Obj_ofArrayReader, Obj_ofArrayWriter
Parameters	Param_OfObjectCreator, Param_OfFieldReader, Param_ofFieldWriter, Param_ofTypeManipulator, Param_ofLocalReader, Param_ofLocalWriter, Param_ofArrayCreator, Param_ofArrayReader, Param_ofArrayWrite

For example, the *Obj_InIfBranch* Condition attribute indicates if the object part exists in the condition part of the `If` statement. Also, the object part exists in the body part of the while statement, the *Obj_InWhileLoop* Body attribute is set to be true. Similarly, the *Param_InIfBranch* Condition attribute checks if the parameters existed in the condition part of the `If` statement. For the target C-Vector, the *Obj_ofObjectCreator* attribute indicates that the object part is discovered in the statement containing the new operator. Also, if the object is used in the statement writing fields, the *Obj_ofFieldWriter* is set to be true.

4 Applying C-Vector to Open Source

This section introduces several cases of the C-Vectors. We established the C-Vectors from 10 popular Java based open source projects. Among all methods from the projects, we only selected the methods composing of a verb and object part. Due to the space limit, we present the C-Vectors of the two representative methods: the 2663 methods structured as `create + Object()` and the 19,025 methods structured as `get + Object()` as shown in Fig. 1.

For the methods shaped like `create + Object()`, a developer tends to name it when he/she creates a new object with the passed parameters. Figure 1a presents the C-Vector to show the tendency. The *isReturnReference* and *existCreateCustomObject* attributes present that the method creates a new object with the reserved word `new` and return the created object. According to the figure, about 85.3 % of the methods returns a reference, and 59.5 % creates a new object. In addition, 57.9 % of the methods are observed that the object part of the method identifiers exists in the return type, meaning that the methods are intended to create an object and they are returned to the method invokers as a return type. Other attributes do not characterize the methods.



Fig. 1 The Frequency of the C-Vector of the *createO* and *getO* methods. **a** Frequency of the *createO* method's C-Vector. **b** Frequency of the *getO* method's C-Vector

When it comes to the methods structured as `get + Object()` as shown in Fig. 1b, most of the methods (72.27 %) returns a custom reference (see the `isReturnReference` attribute), and many methods (67.7 %) do not obtain parameters. The figure also says that 74.8 % of the methods do not declare new local variables.

5 Conclusion

In this paper, we have defined the C-Vector with three categories: the C-Vector for a method signature, a method body, and relationships between them. Also, we have examined 10 popular open source projects written in Java and analyzed distribution of the C-Vector for the representative two methods. Use of the C-Vector enables one to express the characteristics of the method, also it is utilized to cluster methods and compare a characteristics of a method with others. As a future work, we are planning to build tool support for mining relationships between the method signature and the method implementation by applying a machine learning approaches into it.

Acknowledgments This research was supported by the Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (NRF-2014M3C4A7030505).

References

1. Arnaoudova, V., Penta, M.D., Antoniol, G., Guhneuc, Y.-G.: A new family of software antipatterns: linguistic anti-patterns. In: The European Conference on Maintenance and Reengineering, pp. 187–196. Genova, Italy (2013)
2. Sun Microsystems: Code conventions for the java programming language: why have code conventions. <http://www.oracle.com/technetwork/java/index-135089.html> (1999)
3. Gil, Y., Maman, I.: Micro patterns in java code. In: The 20th Annual ACM SIGPLAN Conference on Object-Oriented Programming, Systems, Languages, and Applications, pp. 97–116 (2005)
4. Host, E.W., Ostvold, B.M.: The programmers lexicon: the verbs. In: The Seventh IEEE International Working Conference on Source Code Analysis and Manipulation, vol. i, pp. 193–202 (2007)
5. Kim, S., Pan, K., Whitehead, E. Jr.: Micro pattern evolution. In: Proceedings of the International Workshop on Mining Software Repositories, pp. 40–46. Shanghai, China (2006)
6. Robert, C.: Martin. Prentice Hall, Clean Code (2008)
7. Singer, J., Brown, G., Lujn, M., Pocock, A., Yiapanis, P.: Fundamental nano-patterns to characterize and classify java methods. Electron. Notes Theor. Comput. Sci., **253**(7):191–204 (2010)

A Study on Real Time Circular Motion in Robots Using Kalman Filters

Malrey Lee, Suntae Kim and Younghwa Cho

Abstract This paper presents a more detailed design method of Q and R when the mobile robots move in circular motions: It through measuring and comparing the displacement difference of broken-line motions and circular motions within a relatively short time interval t to determine the value of Q at the same time using the mean of measurement error as the value of R. The results show that this way of design can effectively reduce the error of the trajectory.

Keywords Kalman filters • Circular motions • Process noise • Measurement noise

1 Introduction

In trajectory prediction of mobile robot one of the key problems is how to estimate the position. In order to obtain this estimated position Kalman filter has become one of the most commonly used technical methods. In the work by von der Hardt and Wolf [1], an Extended Kalman Filter (EKF) algorithm has been developed to calculate the distance of robot in different time. Martinelli [2] and Larsen [3] introduced an Augment Kalman Filter (AKF) to estimates the robot configuration.

M. Lee

Department of Computer Science and Engineering, Chonbuk National University, 567 Baekje-Daero, Deokjin-gu, Jeonju-si, Jellabuk-do 561-756, South Korea
e-mail: mrlee@chonbuk.ac.kr

S. Kim (✉)

Department of Software Engineering, Chonbuk National University, 567 Baekje-daero, Deokjin-gu, Jeonju-si, Jellabuk-do 561-756, South Korea
e-mail: stkim@jbnu.ac.kr

Y. Cho

Colleage of Information and Communication Engineering, SungKyunKwan University, 2066 Seobu-Ro, Jangan-Gu, Suwon, Gyeonggi-Do, South Korea
e-mail: choyh2285@skku.edu

Identically, there still exist other approaches for example Villacotra-Atienza and Makrarov [4] constructed a Neural Network Architecture (NNA) to implement Trajectory modeling of roving robot.

Kalman filter is a set of mathematical equations that provides an efficient computational solution of the least-squares method. The filter is very powerful in several aspects: it supports estimations of past, present, and even future states, and it can do so even when the precise nature of the modeled system is unknown. Because of the advantages and simplicity of the design that be mentioned above the applications of Kalman filters have been gotten a long-term and extensive development meanwhile in order to realize a higher level of accuracy researchers have already made a lot of efforts. In the work by Borenstein and Feng [5], a calibration technique called, “UMBmark” has been developed to calibrate for systematic errors. Tu and Kiang [6] proposed a Robust Kalman Filter (RKF) algorithm to improve the estimation accuracy.

Compared with other researches, our job is focused on the calculation of Q and R. Through reasonable selection of Q and R the precision of Kalman filter can be enhanced markedly. In the design of Q this paper uses the displacement difference of broken-line motions and circular motions within a relatively short time interval t and R uses the average error of measurement. This paper is organized as follows: Sect. 2 reports the EKF method. Section 3 reports the experiment results. Section 4 gives the conclusions.

2 Extended Kalman Filter Method

2.1 Mobile Robot Model

Mobile robots can be modeled in many ways at many different levels. The methods for modeling a mobile robot can be classified into three categories. The first one is velocity approach, which models the robots using the relationship of multi-rigid body locomotion. This method is commonly applied into planar mobile robots [7–9]. The second one is the geometric method, which models the robots’ kinematics using the geometric constrains of multi rigid. For example, the robot “SRR” made by MIT is modeled based on the different expression of the height of wheel center in variant wheel-leg chains [10]. The third one is the coordinate transformation approach, which is firstly introduced by Muir and Newman [11] based on the analysis of the multi chain and closed characters of a mobile robot. This transformed method is also used here as well. Assuming the translational and rotational displacement of mobile robot can be measured as follows:

$$\delta d = \frac{dr + dl}{2} \quad (1)$$

$$\delta\theta = \frac{dr - dl}{b} \quad (2)$$

where δd , $\delta\theta$, dr , dl and b represents the translational displacement, the rotational displacement, the distance covered by right wheel, the distance traveled by left wheel and the width of the mobile robot respectively. In this paper we use the MATLAB (R2010b) to simulate the circular motions of robot and use the White Gaussian Noise (WGN) to instead of the actual measurement error. The simulation result is shown in Fig. 1.

$$Rl = r + \Delta r \quad (3)$$

$$Rr = r + \Delta r + b \quad (4)$$

where Rl , Rr , r and Δr represents the motion radius of left wheel, the motion radius of right wheel, the initialization radius of left wheel and WGN respectively. We set $r = 100$ cm, $b = 10$ cm, $\Delta r = 0.1 * \text{randn}$ cm. When we know the Rl and Rr the value of dl and dr can be calculated through the Euclidean distance. Simultaneously, with the above calculated translation δd_n and rotation $\delta\theta_n$ at an instant time $t = n * T$ ($T = 0.1$ s, $n = k = 0, 1, 2 \dots 10$) the robot location (in 2D plane) update can be estimated as follows:

$$X_{n+1} = X_n + \delta d_n \cos\left(\theta_n + \frac{\delta\theta_n}{2}\right) \quad (5)$$

$$Y_{n+1} = Y_n + \delta d_n \sin\left(\theta_n + \frac{\delta\theta_n}{2}\right) \quad (6)$$

$$\Theta_{n+1} = \theta_n + \delta\theta_n \quad (7)$$

Fig. 1 The simulation of circular motions in MATLAB

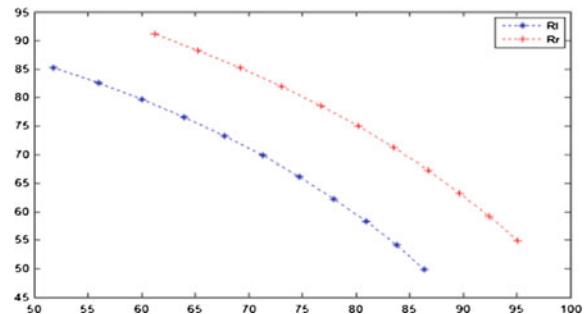
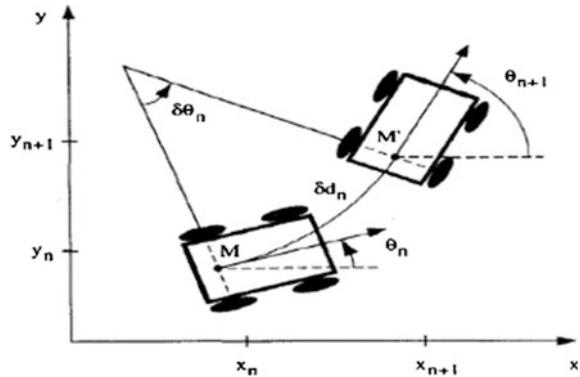


Fig. 2 The position of the robot in 2D coordinates



where (X_n, Y_n) and θ_n represents the location and tangent angle of robot at time t while (X_{n+1}, Y_{n+1}) and θ_{n+1} represents the location and tangent angle of robot at time $t + 1$. The position of robot in 2D is shown in Fig. 2.

2.2 Extended Kalman Filter

The traditional Kalman filter is optimal when the model is linear [12]. However in many practical applications this Kalman filter is very limited because most of the state estimation problems like tracking of the target are nonlinear. In order to deal with this situation the extended Kalman filter provides a more effective solution.

Different from Kalman filter, extended Kalman filter deals with nonlinear process model. In the extended Kalman filter, the state transition and observation models need not be linear functions of the state but may be differentiable functions. The nonlinear process model (from time $k-1$ to time k) is described as:

$$x_k = f(x_{k-1}, u_{k-1}) + w_{k-1} \quad (8)$$

$$z_k = h(x_k) + v_k \quad (9)$$

where x_{k-1} and x_k ($k = 0, 1, 2 \dots$) represents the system state vectors at time $k - 1$ and k , f is the system transition function, u_k is the system input, z_k represents the system measurement vector, h is the observation function, w_k and v_k represents the process noise and measurement noise, in this paper they are assumed to be independent of each other and with normal probability distributions with covariance matrices Q and R . The same as the working principle of Kalman filter, there are also two update processes (the time update and measurement update) in the extended Kalman filter: Time update:

$$x_{k|k-1} = f(x_{k-1|k-1}, u_{k-1}) \quad (10)$$

$$P_{k|k-1} = A_{k-1} P_{k-1|k-1} A_{k-1}^T + Q_{k-1} \quad (11)$$

Measurement update:

$$K_k = P_{k|k-1} H_k^T (H_k P_{k|k-1} H_k^T + R_k)^{-1} \quad (12)$$

$$x_{k|k} = x_{k|k-1} + K_k (z_k - h(x_{k|k-1})) \quad (13)$$

$$P_{k|k} = (1 - K_k H_k) P_{k|k-1} \quad (14)$$

where $P_{k|k-1}$ and $P_{k|k}$ represents the predicted and updated estimate covariance, K_k is the optimal Kalman gain, A_{k-1} and H_k are the state transition and observation matrices. Now, considering this nonlinear process model in circular motions of mobile robot the extended Kalman filter can be implemented as follows: Assuming:

$$x_{k+1} = A_k x_k + B_k u_k + w_k \quad (15)$$

$$z_k = H_k x_k + v_k \quad (16)$$

$$x_k = z_k = [X_k, Y_k, \theta_k]^T \quad (17)$$

$$u_k = [\delta d_k, \delta \theta_k] \quad (18)$$

Here, H_k is set to an identity matrix, while A_k and B_k can be obtained by Jacobian matrices:

$$A_k = \begin{bmatrix} \frac{\partial f1}{\partial X_k} & \frac{\partial f1}{\partial Y_k} & \frac{\partial f1}{\partial \theta_k} \\ \frac{\partial f2}{\partial X_k} & \frac{\partial f2}{\partial Y_k} & \frac{\partial f2}{\partial \theta_k} \\ \frac{\partial f3}{\partial X_k} & \frac{\partial f3}{\partial Y_k} & \frac{\partial f3}{\partial \theta_k} \end{bmatrix} \quad B_k = \begin{bmatrix} \frac{\partial f1}{\partial \delta d_k} & \frac{\partial f1}{\partial \delta \theta_k} \\ \frac{\partial f2}{\partial \delta d_k} & \frac{\partial f2}{\partial \delta \theta_k} \\ \frac{\partial f3}{\partial \delta d_k} & \frac{\partial f3}{\partial \delta \theta_k} \end{bmatrix} \quad (19)$$

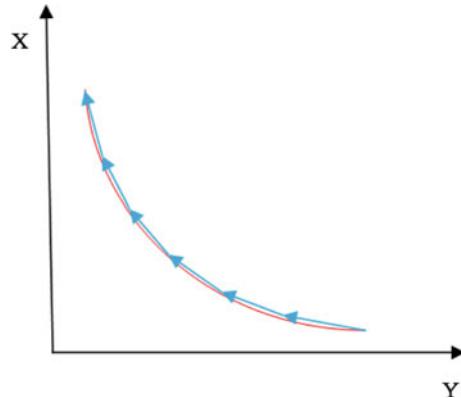
where $f1$ and function (5) are the same, $f2$ and function (6) are the same and $f3$ and function (7) are the same. After calculation the A_k and B_k can be defined as:

$$A_k = \begin{bmatrix} 1 & 0 & -\delta d_k \sin(\theta_k) \\ 0 & 1 & \delta d_k \cos(\theta_k) \\ 0 & 0 & 1 \end{bmatrix} \quad B_k = \begin{bmatrix} \cos(\theta_k) & -\frac{1}{2} \delta d_k \sin(\theta_k) \\ \sin(\theta_k) & \frac{1}{2} \delta d_k \cos(\theta_k) \\ 0 & 1 \end{bmatrix}$$

$$\theta_k = \theta_k + \frac{\delta \theta_k}{2} \quad (20)$$

The solving of Q we use the difference of δd , $\delta \theta$ in circular motions (c) and in broken-line motions (l), here we make the displacement (δd) of c and l is the same

Fig. 3 The difference of δd and $\delta\theta$ in c and l



in this way this difference comparison can be more accurate than they are diverse, this illustration is shown in Fig. 3. It can be seen that: when we select the time interval t is too short then the difference is too small so the EKL simulation result of c is approximate to l we can say this result is not ideal, while when we select the time interval t is too long then the difference is too big so the EKL simulation result of c will have a big wave we can say this result is not very accurate. In order to achieve a high accuracy we should search a suitable time t and when we know it the solving of R can be known as well because we can calculate the mean error of δd , $\delta\theta$ at time t to as R.

$$Q = \begin{bmatrix} \delta d_k \cos(\theta_k) - \delta d_k \cos(\theta_k) \\ \delta d_k \sin(\theta_k) - \delta d_k \sin(\theta_k) \\ \theta_k - \theta_k \end{bmatrix} = \begin{bmatrix} Q_{kx} \\ Q_{ky} \\ Q_\theta \end{bmatrix} \quad (21)$$

$$R [\bar{e}_x, \bar{e}_y, \bar{e}\theta] \quad (22)$$

where \bar{e}_x , \bar{e}_y , and $\bar{e}\theta$ represents the mean error of δd on the x axis, mean error of δd on y axis and the mean error of $\delta\theta$ respectively.

3 Simulation Results

Figure 4 shows the value of Q_{kx} and Q_{ky} . It can be concluded from the Fig. 4 that though the Q_{kx} and Q_{ky} showed a negative correlation, the absolute value of both are become bigger and bigger with the increase of time t ($t = k * T$, $T = 0.1$ s, $k = 0, 1, 2 \dots 10$), because of this reason we select the value of δd and $\delta\theta$ in middle time $k = 5$.

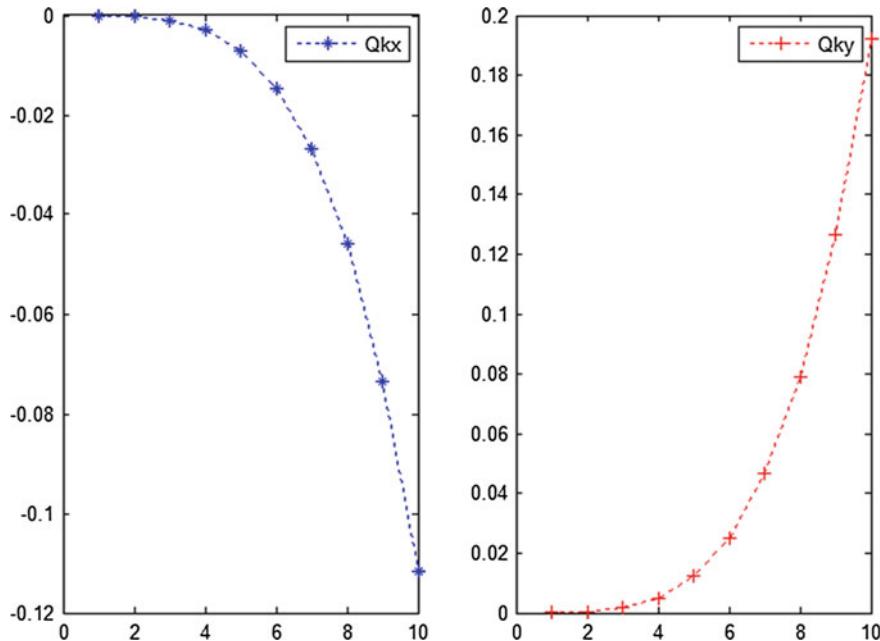


Fig. 4 The value of Q_{kx} and Q_{ky}

In the process of simulation the velocity and start angle of mobile robot is consistent $v = 50$ cm/s, $\theta_k = \frac{\pi}{6}$ and to ensure the continuity of data in WGN every time interval t we randomly generate 500 values and run many times.

The value of Q_{kx} , Q_{ky} , \bar{e}_x and \bar{e}_y are chosen as follows:

$$[Q_{kx}, Q_{ky}] = [0.007 \ 0.0122] \quad (23)$$

$$[\bar{e}_x, \bar{e}_y] = [0.005 \ 0.005] \quad (24)$$

In Table 1 we also compare the value of Q and R which calculated by our method with other people's research in this filed.

Through this comparison we can get the conclusion: our method has a higher accuracy than EKF2. In Figs. 5 and 6 illustrates the estimated errors of position X and position Y.

Table 1 The value of Q and R in EKFs

	Q (cm)	R (cm)
EKF1	(0.007, 0.0122)	(0.005, 0.005)
EKF2	(0.01, 0.02)	(0.005, 0.005)

EKF1: the design method of extended Kalman filter in this paper
EKF2: the design method of extended Kalman filter in [13]

Fig. 5 Estimation errors of position X

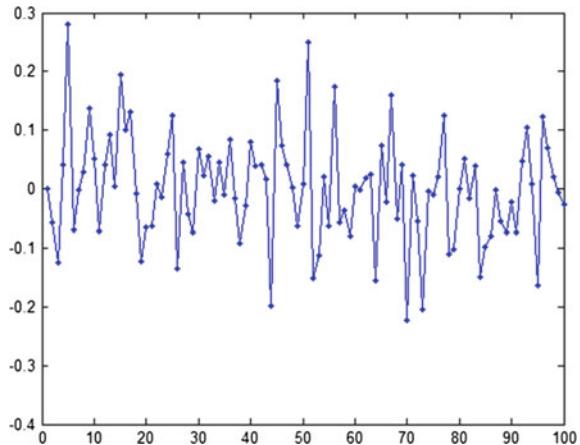
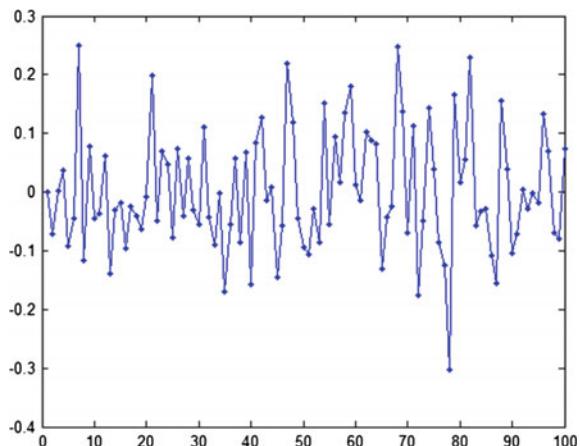


Fig. 6 Estimation errors of position Y



4 Conclusion

Compared with other people's research in this paper we propose a more detail and feasible method to obtain the process noise Q and measurement noise R . In addition, the estimation errors produced by this method are relatively small (no bigger than 0.4 cm). In the design of Kalman filter some people ignore the importance of Q and R which affects the accuracy of Kalman filter to a large degree, however, how to select and confirm the value of Q and R is still a kind of challenge, here, as for Q we use a comparative method: when the mobile robots move in circular motions the broken-line motions are a main interference. But we should say sometimes this method is very powerful especially when we know the movement rule of the object, on the other hand if the rule cannot be found for example the random motions the

measurement of Q will face with great difficulties and this method must be changed or no longer applicable. As for R its value is relatively easier to determine than Q.

Acknowledgments This research was supported by the Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (NRF-2014M3C4A7030503).

References

1. von der Hart HJ, Wolf, D., Husson, R.: The dead reckoning localisation system of the wheeled mobile robot romane. In Proceeding of the 1996 IEEE/SICE/RSJ International Conference on Multisensor Fusion and Integration for Intelligent Systems, 1996
2. Martinelli, A., Tomatis, N., Siegwart, R.: Simultaneous localization and odometry self-calibration for mobile robot. *Auton. Robots* **22**(1), 75–85 (2007) (Springer, Netherlands)
3. Larsen, T.D., Hansen, K.L., Andersen, N.A., Ravn, O.: Design of Kalman filters for mobile robots; evaluation of the kinematic and odometric approach. *IEEE. Int. Conf. Control Appl.* **2**, 1021–1026 (1999)
4. Villacorta-Atienza, J.A., Makarov, V.A.: Neural network architecture for cognitive navigation in dynamic environments. *IEEE Trans. Neural Netw. Learn. Syst.* **24**(12), 2075–2085 (2013)
5. Borenstein, J., Feng, L.: Measurement and correction of systematic odometry errors in mobile robots. *IEEE Trans. Robot. Autom.* **12**(6), 869–880 (1996)
6. Tu, P.J., Kiang, J.F.: Estimation on location, velocity, and acceleration with high precision for collision avoidance. *IEEE Trans. Intell. Transp. Syst.* **11**(5), 374–379 (2010)
7. Campion, G., Bastin, G., Dandrea, N.B.: Structural properties and classification of kinematic and dynamic models for wheel mobile robots. *IEEE Trans. Robot Autom.* **12**(1), 47–62 (1996)
8. Sreenivasan, S.V., Nanua, P.: Kinematic geometry of wheeled vehicle systems. In: Proceedings of the ASME Design Engineering Technical Conferences and Computers in Engineering, vol. 8, pp. 18–22 1996
9. Wheekuk, K., Byund, J.Y., Dong, J.L.: Kinematic modeling of mobile robots by transfer method of augmented generalized coordinates. *J. Robotic Syst.* **21**(6), 301–322 (2004)
10. Iagnemma K, Genot F, Dubowsky S. Rapid physics-based rough-terrain rover planning with sensor and control uncertainty. *Proc. IEEE Int. Conf. Robotics Autom.* **3**, 2286–2291 (1999)
11. Muir, P.F., Neuman, C.P.: Kinematic modeling of wheeled mobile robots. *J. Robotic Syst.* **4** (2), 281–333 (1987)
12. Kalman, R.E. A new approach to linear filtering and prediction problems. *ASME Trans. J. Basic Eng. Ser. D.* **82**, 35–45 (1960)
13. Hassanzadeh, I., Fallah, M.A.: Design of Augment Extended Kalman Filter for Real Time Simulation of Mobile Robots Using Simulink. Proceeding of ISMA sharjah, UAE (2009)

A Study on Traceability Between Documents of a Software R&D Project

Suntae Kim, HyunYoung Kim, Jeong Ah Kim and Younghwa Cho

Abstract During the R&D project, a huge amount of documents are produced and shared with each other, and also the documents are related to software systems. Without careful management of the relationships, researchers cannot recognize an impact of the changes, which degrades the research quality. This paper presents new types of traceability for the software R&D project including traceability between documents, traceability between a document and its items, traceability between different versions of documents. We exemplify the traceability of documents in a R&D project. Based on the traceability, the stakeholders of the R&D project are able to recognize and monitor relationships between documents and research outcomes also, it can also eventually contribute to improving the quality of research outcomes and research progress.

S. Kim · H. Kim

Department of Software Engineering, Chonbuk National University, 567 Baekje-daero,
Deokjin gu, Jeonju-si, Jeollabuk-do 561-756, Republic of Korea
e-mail: stkim@jbnu.ac.kr

H. Kim

e-mail: kyrte@jbnu.ac.kr

J.A. Kim (✉)

Department of Computer Education, Catholic Kwandong University, 24, Beomil-ro
579 beon-gil, Gangneung-si, Gangwon-do, Republic of Korea
e-mail: clara@cku.ac.kr

Y. Cho

College of Information and Communication Engineering, SungKyunKwan University,
2066 Seobu-Ro, Jangan-Gu, Suwon, Gyeonggi-Do, Republic of Korea
e-mail: choyh2285@skku.edu

1 Introduction

A software research and development (R&D) project indicates a project for developing an innovative software systems, or enhancing previous systems [7]. During the R&D project, a huge amount of documents such as research plan, research note and software requirements specification are produced and shared with each other, and also the documents may be directly related to software systems. Without careful management of the relationships, researchers involved in the project cannot recognize an impact of the changes, which degrades the research quality. In addition, evaluators or a research project manager cannot monitor the progress of the project, which eventually makes the project failed.

In the software engineering field, the relationship between software artifacts is defined as ‘software traceability’, which is the ability to maintain the relation between artifacts created in developing a software system [6]. Many approaches are suggested to establish traceability (see [1–3, 6]). These approaches only focus on traceability between artifacts for software development (e.g., traceability between requirements and design documents, or that between design and source code). Therefore, it should be tailored to apply it into building and managing traceability of documents in the software R&D project.

In order to address the issue, we suggest new types of traceability for the software R&D project. First, we introduce the general process of the R&D project and identify main documents that can be elaborated in each process. Then, we propose several types of research traceability, which is categorized into three: traceability between documents, between document and its items, between different versions of documents. Based on the traceability, the stakeholders of the R&D project are able to recognize and monitor relationships between documents and outcomes (e.g., a software system, papers and patents). It can eventually contribute to improving the quality of research outcomes and research progress.

The remainder of this paper is structured as follows. Section 2 presents a software research and development process and main documents at each activity. Section 3 proposes research traceability between research documents including artifacts from software development process. Section 4 concludes a paper.

2 A Process for a Software R&D Project

This section introduces participants and a process of software R&D project, also it presents diverse research documents produced in the project. A software research and development (R&D) project indicates a project for developing an innovative software systems (methods), or enhancing previous systems (methods) [4, 7]. Diverse stakeholders are involved in the project, including a research director who is a leader of the project and researchers who conduct the research. Evaluators are also one of the stakeholders, as most of the projects are funded by the government

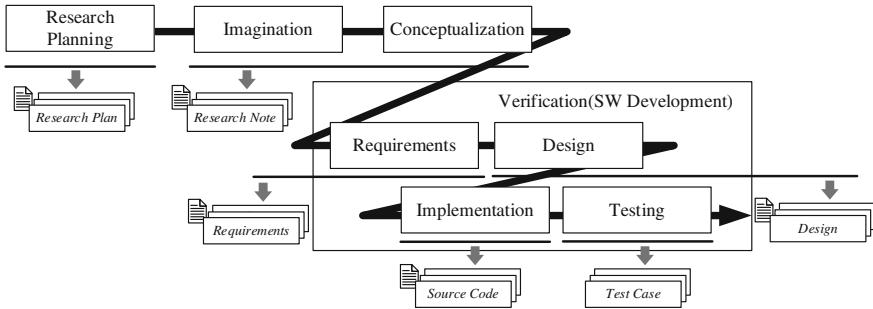


Fig. 1 Process and documents in the software R&D project

or companies. The evaluators monitor the progress of the project, give advices on the projects, and decide whether the project can be proceeded.

A process for carrying out the software R&D project is shown in Fig. 1. According to [7], the software R&D project is executed in four major phases: Research Planning, Imagination, Conceptualization and Verification. The research planning phase establishes a research direction and produces the research plan document containing a research goal, rough research contents, role assignment, a yearly/monthly plans and so forth. In the imagination and conceptualization phases, researchers including a director build their idea from scratch or based on their/other's previous research outcomes. While conducting these phases, their idea gets concrete and realizable. The progress or the output of the phases are recorded in the Research Note in various formats.

As the last verification phase, the idea in the project is likely to be realized into a software system because a software project tends to produce a software system as a final outcome. Thus, the process of this phase generally follows a software development process composed of mainly four activities: Requirements, Design, Implementation and Testing [5]. Each activity produces artifacts such as Requirements Specification, Design Document, Source Code and Test Case at the end of each activity respectively. In addition, diverse publications such as Papers, Patents, and Reports are produced during all phases of the project. Figure 2 shows the hierarchical relation of the research documents. All documents inherit the

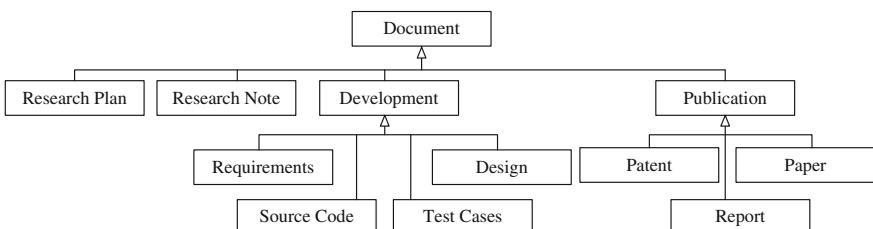


Fig. 2 Software R&D document hierarchy

Document entity that defines common attributes of research documents such as title, versions, authors, last revision date and so on.

It should be noted that we selected the essential outputs of each phase in the software R&D project. Thus, there must be more diverse research documents in various forms depending on the research area in software. In addition, we assume the general software development process such as Waterfall and V-Model [5] for verifying the research. However, various software development processes (e.g., Agile methods) can be applied in carrying out software development in the R&D project. These assumptions are intended to narrow down the discussion scope for defining traceability between research documents.

3 Traceability Between Documents of a Software R&D Project

This section presents traceability between research documents by categorizing it into three: traceability between documents, traceability between document and its items, traceability between different versions of documents. Figure 3 shows the hierarchy of the traceability. *TR-BTW-Doc* defines traceability between research documents, containing *is refined by/conform*, *produce/is based on*, *is satisfied by/satisfy* and *is referred by/refer*. *TR-IN-Doc* defines traceability between a document and its items. Also, *TR-IN-Ver* indicates traceability in terms of revision history.

Inherently, a traceability is bi-directional and each direction has a traceability identifier. Thus once one traceability from an entity e_1 to another entity e_2 is defined, the opposite traceability from e_2 to e_1 is automatically derived from it. In this sense, the traceability *produce/is based on* can be understood as the traceability from e_1 to e_2 is the produce traceability, and the opposite traceability from e_2 to e_1 is the *is based on* traceability. In the following subsection, we discuss three types of traceability in more detail.

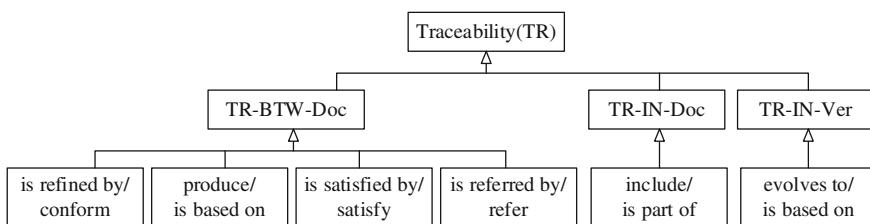


Fig. 3 Type of traceability in a SW R&D project

3.1 Traceability Between Research Documents

In the software engineering field [5], traceability is generally defined as relationship between documents. Although Spanoudakis and Zisman [6] collected the diverse use of traceability in software engineering, it cannot directly apply them into the software R&D project field without modification. Thus, we modify their research in order to reflect the software R&D project environment. In order to show the use of traceability in the R&D project, we only present documents in the project, and categorize them into three: Support, Research and Software Development (SW Dev.) as shown in Fig. 4. Based on the documents, traceability between research documents can be defined as below:

- *is refined by/conform*: In this type of traceability, a document $d1$ is refined by a document $d2$ and $d2$ conforms the $d1$. $d2$ is a result of efforts to realize $d1$. Thus, the contents of $d2$ should be in line with the contents of $d1$ (e.g., goals). In the R&D project, the traceability is discovered inbetween the research plan and research note documents. In terms of the contents, this traceability does not strictly constraint contents-conformance. As an example of Fig. 4, sometimes the research note cannot conform the research plan, but generally do.
- *produce/is based on*: A document $d1$ produces a document $d2$, also $d2$ must be based on $d1$. When defining this traceability, the outcome should be officially

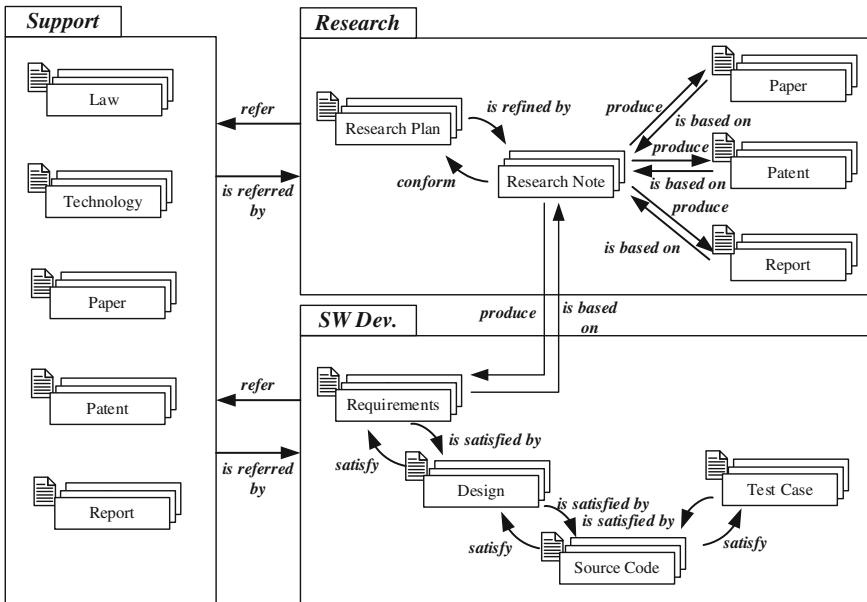


Fig. 4 Traceability definition between R&D documents

published. Papers, Patents, Reports and Requirements Specification based on the contents of the research note are examples of defining this traceability.

- *is satisfied by/satisfy*: This traceability indicate the case of two documents $d1$ and $d2$ has a relationship for realization. All contents of $d2$ should be a way of realization of $d1$. Relationships between Requirements and Design, and Design and Source Code are examples of this traceability. Although it seems to be similar to the *is refined by/conform* traceability, it is different in terms of level of strictness.
- *is referred by/refer*: This traceability indicates a relationship that a document $d2$ can be used as a rationale of a document $d1$. This traceability is presented as $d1$ refers to $d2$ and $d2$ is referred by $d1$. This traceability is very general in research because most of research is built upon previous researches or technologies. In addition to these, changes of social environments such as Law also motivates the needs of the R&D projects.

3.2 Traceability Between a Document and Document Items

A document is composed of many document items. In software engineering, traceability is defined only in between documents. However, document items should be carefully managed in a R&D project due to different coverage items. As an example, a research plan document may have several research goals, and later each research goal spawns diverse documents such as software requirements specification and papers. In this situation, most of the documents should define traceability with the entire research plan document, not the items in the document. Also, it is valuable for evaluators to monitor the progress of the project by checking linked traceability of items in the research plan.

Figure 5a presents traceability between *Document* and *Document Items*. As depicted in the figure, a *document includes* the *document items*, also a *document item is part of a document*. Although it is quite simple, it influences all documents because it is defined in the *Document* level, which is the ancestor class of other documents (see Fig. 2). According to this traceability, all documents and their

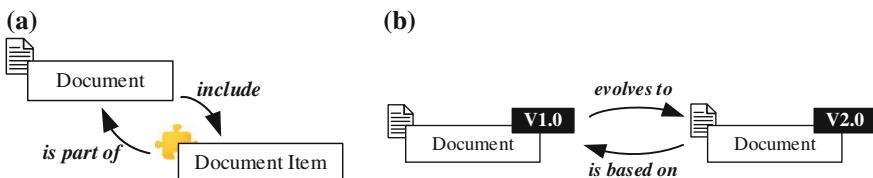


Fig. 5 Traceability between a document and document items/different versions. **a** Traceability between a document and document items. **b** Traceability between difference versions of a R&D document

document items are allowed to maintain traceability *include* and *is part of*. Furthermore, the *TR-BTW-Doc* traceability can also be applied in defining traceability between document items.

3.3 Traceability Between Difference Versions of Documents

While carrying out the R&D project, updating research outcomes frequently happens. Although all revisions cannot be recorded or controlled, major revisions of a document should be managed. This relationships can be captured as *evolves to* and *is based on* traceability. Figure 5b presents this relationship. When a document with *V1.0* (i.e., *d-v1.0*) is updated and it becomes the document with *V2.0* (i.e., *d-v2.0*), *d-v1.0* evolves to *d-v2.0*, and *d-v2.0* is based on *d-v1.0*. This traceability enables one to examine all revisions of a document as well as document items.

4 Conclusion

In this paper, we have presented new types of traceability for the software R&D project including traceability between documents, traceability between a document and its items, and traceability between different versions of documents. Based on the traceability, the stakeholders of the R&D project are able to recognize and monitor relationships between documents and outcomes in multiple-aspects. In this paper, we did not discuss how to establish the traceabilities in automatic/semi-automatic ways. As a future work, we are planning to build a system for supporting traceability management, and inferencing traceability in an automatic manner.

Acknowledgments This research was supported by the Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (NRF-2014M3C4A7030503).

References

1. Alexander, I.: Semiautomatic tracing of requirement versions to use cases experience and challenges. In: Proceedings of the 2nd International Workshop on Traceability in Emerging Forms of Software Engineering (TEFSE 2003), pp. 395–428. Canada (2003)
2. Cleland-Huang, J., Gotel, O., Hayes, J.H., Mader, P., Zisman, A.: Software traceability: trends and future directions. In: Proceeding of the on Future of Software Engineering (FOSE), pp. 55–69. New York, NY, USA (2014)
3. Hill, J., Tilley, S.: Creating safety requirements traceability for assuring and recertifying legacy safety-critical systems. In: Proceeding of 18th IEEE International Requirements Engineering Conference (RE), pp. 297–302 (2010)

4. Jain, R.K., Triandis, H.C., Weick, C.W.: *Managing Research Development and Innovation*, 3rd edn. Wiley, New York (2010)
5. Pressman, R., Maxim, B.: *Software Engineering: A Practitioner's Approach*, 8th edn. McGraw-Hill Science/Engineering/Math (2014)
6. Spanoudakis, G., Zisman, A.: Software traceability: a roadmap. In: *Handbook of Software Engineering and Knowledge Engineering*, pp. 395–428. World Scientific Publishing, Singapore (2004)
7. Wingate, L.M.: *Project Management for Research and Development: Guiding Innovation for Positive R&D Outcomes*. Auerbach Publications (2014)

Path Planning for Avoiding Obstacles for Unmanned Ground Vehicles

Gyoungeun Kim, Deok Gyu Lee and Byeongwoo Kim

Abstract An Unmanned Ground Vehicle (UGV) is an intelligent system that drives a vehicle safely to the defined destination without any driver support. The safe operation of a UGV requires studies on safe and efficient path generation. In this paper, we propose a path planning algorithm for avoiding obstacles using Vehicle to Vehicle (V2V) communication, which is a vehicle safety communication method, to overcome the limits of existing studies, such as blind zones, that recognize vehicles on the basis of the sensors attached to vehicles. The proposed path planning algorithm generates a safe path by using the cubic spline method and recognizes obstacles by the index values of various possible paths for auto-piloting the vehicle. This study is expected to aid in the application of V2V-based evasion path planning in various scenarios.

Keywords Path planning • Collision avoidance • Avoiding obstacle • Path planning • Cubic spline • Vehicle-to-Vehicle (V2V) communication • Path trajectory • Pre-safety system

G. Kim

Graduate School of Electrical Engineering, University of Ulsan, 93 Daehak-ro, Ulsan,
Republic of Korea
e-mail: gyg509@gmail.com

D.G. Lee

Department of Information Security, Seowon University, 377-3 Musimseoro, Heungdeok-gu,
Cheongju, Chungbuk, Republic of Korea
e-mail: deokgyulee@seowon.ac.kr

B. Kim (✉)

School of Electrical Engineering, University of Ulsan, 93 Daehak-ro, Ulsan, Republic of
Korea
e-mail: bywokim@ulsan.ac.kr

1 Introduction

Unmanned Ground Vehicles (UGVs) have attracted considerable attention across the world as a technology that can improve driving safety and convenience by recognizing the driving environment and aiding the driver or auto-piloting the vehicle. Many leading automobile companies and research institutes are pursuing research and development on this technology. Further, to encourage research in this field, the US Department of Defense has been hosting international autonomous vehicle competitions such as the Defense Advanced Research Projects Agency (DARPA) Grand Challenge since 2004 [1].

The UGV is an intelligent system that can drive the vehicle itself to a designated destination without the aid of a driver. For safe UGV operating, the vehicle must be able to recognize and assess the environment for dangers and drive the vehicle along the path trajectory. Therefore, studies on efficient UGV path planning considering safety are essential for the application of UGVs in the real world.

UGVs in existing studies generated path trajectories by recognizing the driving environment on the basis of various sensors attached to the vehicle, such as the GPS, cameras, lasers, and Lidar sensors. However, such path planning failed to consider vehicles existing outside the blind zones generated because of the limits of the installed sensors. Hence, studies on path planning that can overcome sensor limitations are required.

In this paper, we proposed the obstacle-avoiding path planning algorithm that uses the information of nearby vehicles provided by V2V communications rather than the sensors attached to the vehicles [2]. V2V communication, a vehicle safety communication method, was used for overcoming the limits of the attached sensors, such as blind zones, that appear in contemporary sensor installment-based vehicles. Further, in this study, we applied cubic spline interpolation for safer path planning. The PreScan tool and Matlab/Simulink, which provide an intelligent environment, were used for simulating and validating the developed algorithm.

2 Path Planning Algorithm

In this paper, we present a path planning algorithm for avoiding obstacles by employing V2V communication.

The steps of the proposed path planning algorithm include the generation of waypoints, the generation of candidate paths, and the selection of the paths and the path trajectory, as shown in Fig. 1.

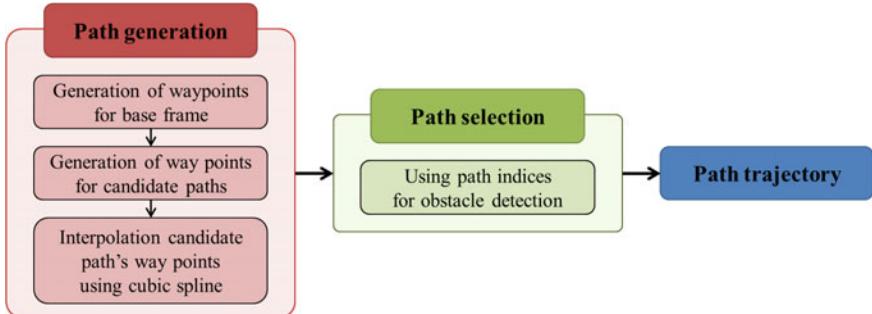


Fig. 1 Overview of the path planning algorithm

2.1 Path Generation

The path planning methods that generate possible paths to a destination in a driving environment are the global path planning method for generating the entire path and the local path planning method that plans a path by recognizing the surroundings. The path planning algorithm proposed in this paper combines the advantages of global and local path planning, as shown in Fig. 2a. A base frame path was initially generated by global path planning, and subsequently, the candidate paths that reflected the driving environment were generated by local path planning [3].

Global and local paths were based on the waypoints that were generated by waypoint generator. The global path that reflected the driving environment was set up to pass through the center of the road. The ego vehicle followed the same path as the base frame, that is, the global path.

To ensure that the proposed path planning algorithm is applicable under various road conditions, a curved path was considered. Therefore, the base frame and candidate paths were designed by using the cubic spline interpolation method.

This widely used interpolation method connects the given waypoints to form a smooth curve. Because the curve connecting the two points was represented by a third-degree polynomial, the proposed algorithm was called “cubic spline” [4].

The application of cubic spline interpolation for path generation is explained below. Assume there is a set of waypoints (x_i, y_i) with a total of $n + 1$ waypoints

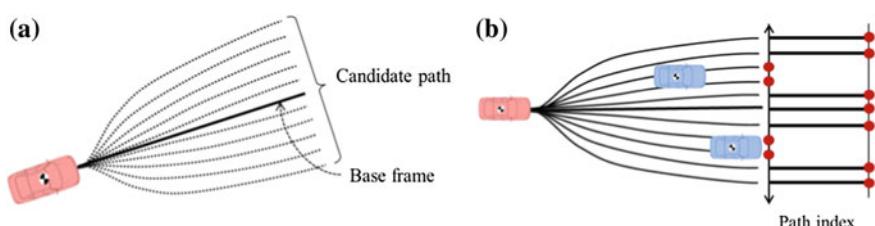


Fig. 2 Path generation and selection, **a** cubic spline interpolation, **b** candidate paths and indices

represented by x and y coordinates. In an interval $[x_i, x_{i+1}]$, the cubic spline $s(x)$ has the following form:

$$s(x) = y(x) = a_i(x - x_i)^3 + b_i(x - x_i)^2 + c_i(x - x_i) + d_i \quad i = 1, 2, 3, \dots, n \quad (1)$$

The length of the subinterval is expressed as. The following free boundary conditions are imposed.

$$x = x_i, \quad y = y_i \quad (2a)$$

$$x = x_{i+1}, \quad y = y_{i+1} \quad (2b)$$

$$S''(x_i) = S_i = 0 \quad (2c)$$

$$S''(x_{i+1}) = S_{i+1} = 0 \quad (2d)$$

Using Eqs. (2a–2d), we can write the coefficients a_i through d_i as shown below.

$$a_i = \frac{s_{i+1} - s_i}{6h_i} \quad (3a)$$

$$b_i = \frac{s_i}{2} \quad (3b)$$

$$c_i = \frac{y_{i+1} - y_i}{h_i} - \frac{2h_i s_i + h_i s_{i+1}}{6} \quad (3c)$$

$$d_i = y_i \quad (3d)$$

2.2 Path Selection

The candidate paths generated from the global path and ego vehicle information are presented as indices in Fig. 2b.

The number of indices corresponds to the number of candidate paths generated. An index value of 0 indicates the presence of an obstacle on the generated candidate path, and a value of 1 indicates that there are no obstacles on the path. Obstacles were detected on the basis of the information provided by V2V communications, and the most optimal candidate path was proposed for avoiding collision. Once a collision-free path was determined, the UGV was controlled such that the minimum distance between the front wheels and the path was maintained at 0, and the turning angle of the front wheels aligned with the tangent of the path; this allowed the UGV to stay on and follow the generated path [5, 6].

3 Simulation and Results

3.1 Simulation Scenario

The driving environments and scenarios were defined, as shown in Fig. 3, to analyze the V2V-communication-based algorithm for collision avoidance. The driving environments were set up such that the view of the driver of the ego vehicle was hindered by vehicle 1, which in turn made the detection of a possible threat, such as a stalled vehicle 2, difficult for the driver of the ego vehicle. By implementing the proposed algorithm, the ego vehicle was maintained at a constant speed. The driver of vehicle 1 detected the stalled vehicle 2 and changed the lane to avoid collision. The detailed scenarios of the simulation are presented in Table 1.

3.2 Simulation Results

Simulations of the collision avoidance algorithm for straight roads were conducted according to the scenarios defined above.

As a result of Fig. 4, ego vehicle trajectory the smooth and safety path for avoiding the stalled vehicle 2. The path with considering the driving environment was more smooth and effective than without considering the driving environment.

Because of utilizing the V2V communications, ego vehicle detected the stalled vehicle 2. In this context, the information of the vehicles was taken into account in the simulations. The presence of obstacle is checked with the information received through V2V communication.

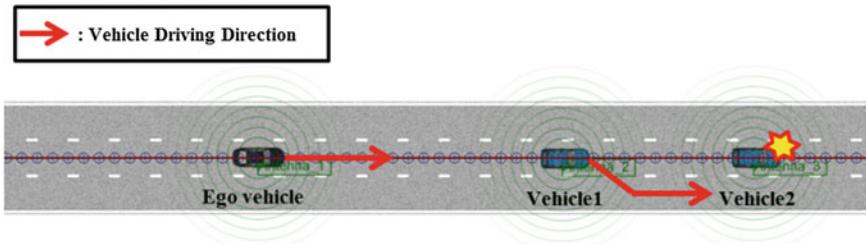


Fig. 3 The driving environment and initial scenario condition

Table 1 Simulation scenario

Vehicle	Condition	Vehicle velocity (km/h)
Ego vehicle	Path planning	80
Vehicle 1	Lane change	60
Vehicle 2	Stalled vehicle	0

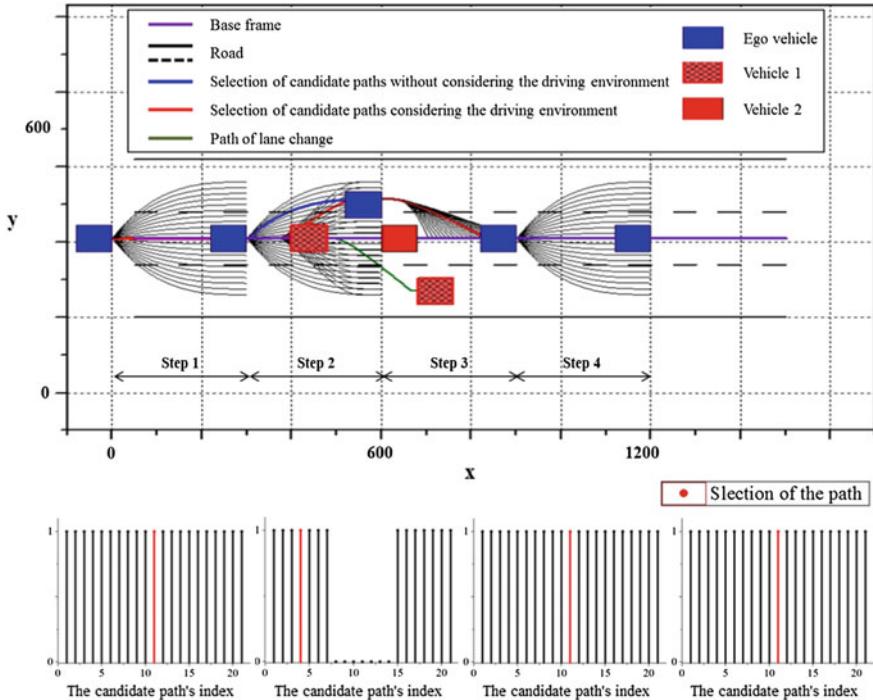


Fig. 4 Results of path planning algorithm simulation

The path planning process for obstacle collision avoidance is shown in Fig. 4. As shown in Fig. 4, Ego vehicle followed the base frame until vehicle 1 changed the lane to avoid collision. When vehicle 1 changed the lane to avoid the stalled vehicle 2, ego vehicle received the information about lane changing of vehicle 1 and stalled vehicle 2; thus the proposed path planning algorithm selected the path for collision avoidance. Candidate path has a fixed lateral offset based on base frame can be confirmed. The proposed path planning algorithm selects optimal path from candidate path index considering the presence of obstacle; and ego vehicle follows the selected path.

4 Conclusion

On-board sensors of automotive vehicles suffer from limitations such as blind zones. To overcome such limitations, an algorithm for collision avoidance based on V2V communications has been proposed in this paper.

In this paper, the proposed algorithm provided the smooth and safety path for avoiding the obstacles. The path considering the driving environment was more smooth and effective than without considering the driving environment.

Such as, verification and research on the stability of the algorithm will be pursued in the future by considering more diverse and complex scenarios. In addition, the research of the diverse standard for selecting the most appropriate path will proceed in the future.

Acknowledgments This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the C-ITRC (Convergence Information Technology Research Center) (IITP-2015-H8601-15-1005) supervised by the IITP (Institute for Information & communications Technology Promotion).

References

1. Woo, H.J., Park, S.B., Kim, J.H.: Research of the optimal path planning methods for unmanned ground vehicle in DARPA urban challenge. In: International Conference on Control, Automation and Systems, ICCAS, pp. 586–589 (2008)
2. Lee, S., Cho, J., Kim, S.: A 3-D Real-time simulation for autonomous driving with V2V communications. In: International Conference on Connected Vehicles and Expo, ICCVE, pp. 800–801 (2013)
3. Nelson, W.: Continuous-curvature paths for autonomous vehicles. In: IEEE Robotics and Automation pp. 1260–1264 (1989)
4. Bronshtein I.N., Semendyayev, K.A., Musiol, G., Muehlig, H.: Handbook of Mathematics. Springer, Berlin (2007)
5. Fausett, L.V.: Applied Numerical Analysis Using Matlab. Prentice-Hall, Upper Saddle River (2000)
6. Chu, K.Y., Lee, M.C., Sunwoo, M.H.: Local path planning for off-road autonomous driving with avoidance of static obstacles. IEEE Trans. Intell. Transp. Syst. **13**(4), 1599–1616 (2012)

Improved AEB Performance at Intersections with Diverse Road Surface Conditions Based on V2V Communication

Sangduck Jeon, Deok Gyu Lee and Byeongwoo Kim

Abstract This paper proposes a Vehicle to Vehicle (V2V) communication-based Autonomous Emergency Braking (AEB) system that considers the road surface condition at intersections. Because existing AEB systems reflect only specific road conditions, braking is ineffective in road conditions other than the ones specified. To rectify this problem, a new control logic, in which the braking time is considered in various road conditions, is introduced. Application of the proposed AEB system with the Time to Collision (TTC) method showed reduced collision risk at an intersection based on information from the Host Vehicle and a neighboring vehicle. The determination as to whether to collide or avoid was made using the relative angle between the vehicles. Further, the results of analysis of the proposed AEB control logic indicate that it offers better braking and anti-collision performance than conventional logic systems.

Keywords V2V (Vehicle to Vehicle) · AEB (Autonomous emergency Braking) · V2I (Vehicle to Infra) · TTC (Time to Collision) · Road condition

S. Jeon

Graduate School of Electrical Engineering, University of Ulsan, 93 Daehak-ro,
Ulsan, Republic of Korea
e-mail: jsd0831@gmail.com

D.G. Lee

Department of Information Security, Seowon University, 377-3 Musimseoro, Heungdeok-gu,
Cheongju, Chungbuk, Republic of Korea
e-mail: deokgyulee@seowon.ac.kr

B. Kim (✉)

School of Electrical Engineering, University of Ulsan, 93 Daehak-ro, Ulsan,
Republic of Korea
e-mail: bywokim@ulsan.ac.kr

1 Introduction

As concerns over vehicle safety increases annually, automobile makers have introduced various functions for driver convenience and safety. Among various kinds of functions, Autonomous Emergency Braking (AEB) is a system that analyzes the risk of collision with the vehicle ahead and prevents or avoids an accident through autonomous braking in emergencies. However, despite the development of this state-of-the-art system, domestic and foreign data show that one-half of all traffic accidents still occur at intersections [1, 2].

Blind spots existing in the corner of intersections are one of the major causes for this high intersection accident rate. In an effort to reduce the traffic accident rate at blind spots, various AEB systems, integrated with Vehicle to Vehicle (V2V), Vehicle to Infra (V2I), and other such technologies, have been introduced [3, 4]. However, because these existing AEB systems usually only reference the Euro NCAP AEB assessment procedure, they reflect only a uniform road condition, i.e., dry asphalt [5]. This limitation makes effective braking difficult in other road conditions. Some researchers have proposed AEB systems that consider various road conditions; however, the AEB systems proposed are only suitable for straight roads. Consequently, applying them to intersections is difficult [6, 7].

This paper proposes a control method that improves AEB systems by considering road conditions at intersections based on the V2V communication environment. Comparison of the proposed AEB system with a conventional system that only considers uniform road condition proves that it is effective.

2 AEB System Control Method

2.1 Determination of Collision Risk at an Intersection

To determine the collision risk at an intersection based on V2V communication, we considered the relative angle and collision risk—Time to Collision (TTC).

Figure 1 presents an image that shows the relative angle between a Host Vehicle and the point of collision with a Target Vehicle based on distance and speed. The relative angle was calculated with Eq. (1). The intersection collision risk, TTC_{Host} , is the time when a vehicle reaches the point of collision with another, and can be calculated using Eq. (2). The smaller the TTC is, the closer is the point of collision.

$$\theta = \tan^{-1} \left(\frac{\text{Intersection distance}_{Target}}{\text{Intersection distance}_{Host}} \right) \quad (1)$$

$$TTC_{Host} = \frac{\text{Intersection distance}_{Host}}{V_{Host}} \quad (2)$$

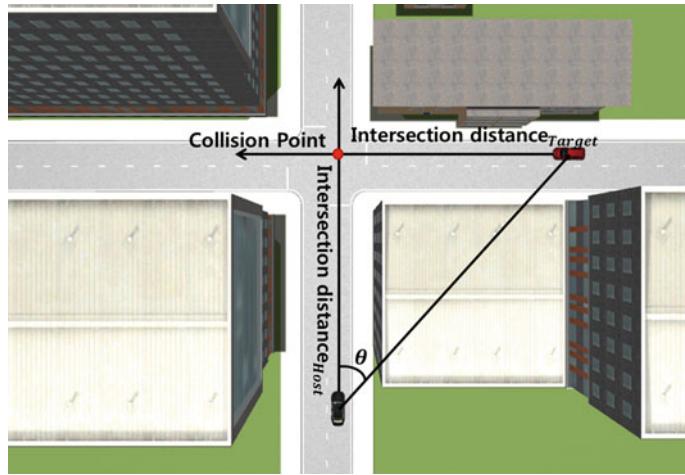


Fig. 1 Calculation of V2V communication-based relative angle

2.2 Calculation of the Braking Time According to Road Conditions

The time at which the proposed AEB system initiates is determined by the collision risk at an intersection, TTC_{Host} , the braking time for the road condition, TTC_μ , and the relative angle. The braking time for the road condition can be calculated with Eq. (3), with a maximum deceleration, a_μ , and vehicle speed, V_{Host} . The maximum deceleration can be calculated using Eq. (4).

$$TTC_\mu = \frac{V_{Host}}{2 \times a_\mu} \quad (3)$$

$$a_\mu = \mu \times g \quad (4)$$

Based on the relative angle calculated from Eq. (1), a collision risk can be determined according to the same conditions as Eq. (5). The proposed AEB system, upon determination of a possible collision, allows full braking as soon as the time to the point of collision, TTC_{Host} Eq. (6), becomes smaller than the braking time, TTC_μ , for the road condition.

$$\theta_{min} \leq \theta \leq \theta_{max} \quad (5)$$

$$TTC_{Host} < TTC_\mu \text{ (Full braking)} \quad (6)$$

3 Simulation and Results

We conducted simulations in which vehicle speeds were set by referencing the speed limit per road in accordance with the Road Traffic Act in Korea, to 40 km/h for city centers, 40 and 60 km/h for general roads, and 100 km/h for highways. The road conditions applied were dry asphalt ($\mu = 0.85$), wet asphalt ($\mu = 0.60$), and snow asphalt ($\mu = 0.30$).

Table 1 shows the resulting status of simulated collisions between vehicles according to speeds and road conditions. The conventional AEB system was confirmed to have a collision in all road conditions except for dry asphalt and wet asphalt at 40 km/h. By contrast, the proposed AEB system was confirmed to avoid collisions in all of the simulated speeds and road surface conditions. This negative result was earned by the conventional AEB system because it fails to consider diverse road conditions. This result proves that the proposed AEB system's control algorithm is effective.

As an exemplary condition further proving the effectiveness of the proposed AEB system, a snow asphalt road surface condition was selected at a speed of 60 km/h.

Figure 2a shows the collision status according to the relative angles deduced from Eqs. (1) and (5). The simulation in which the relative angles were varied between Host Vehicle and Target Vehicle showed that the vehicles collided at 42.5° – 47.5° .

Figure 2b shows different relative angles by time and shows whether the proposed AEB system caused a collision. The condition in which the relative angle was 45° at the start to 3.0 s, as Fig. 2a shows, satisfies the condition for a collision. Full braking was allowed at and after 3.0 s, and the relative angle at which a collision could be avoided was 42.5° (θ_{min}) and greater at 4.4 s.

Figure 3 shows the change in the TTC by time, with the results from the conventional and proposed AEB systems being compared and analyzed. Area I is the area prior to initiation of the AEB system, and the change in the TTC of the two systems is as follows: Area II is where braking is allowed. The existing AEB

Table 1 Collision or avoidance by speed and coefficient of friction

AEB system	Speed (km/h)	Road condition		
		Dry asphalt $\mu = 0.85$	Wet asphalt $\mu = 0.60$	Snow asphalt $\mu = 0.30$
Conventional	40	Avoidance	Avoidance	Collision
	60	Avoidance	Collision	Collision
	80	Avoidance	Collision	Collision
	100	Avoidance	Collision	Collision
Proposed	40	Avoidance	Avoidance	Avoidance
	60	Avoidance	Avoidance	Avoidance
	80	Avoidance	Avoidance	Avoidance
	100	Avoidance	Avoidance	Avoidance

system initiated braking at 4.3 s and full braking at and after 5.5 s. The proposed AEB system facilitated full braking at 3.0 s, which is 1.3 s earlier. In Area III, the conventional system had a collision, whereas the proposed system avoided one. This result was possible because the proposed AEB system's control logic varied the time of braking while considering the road conditions.

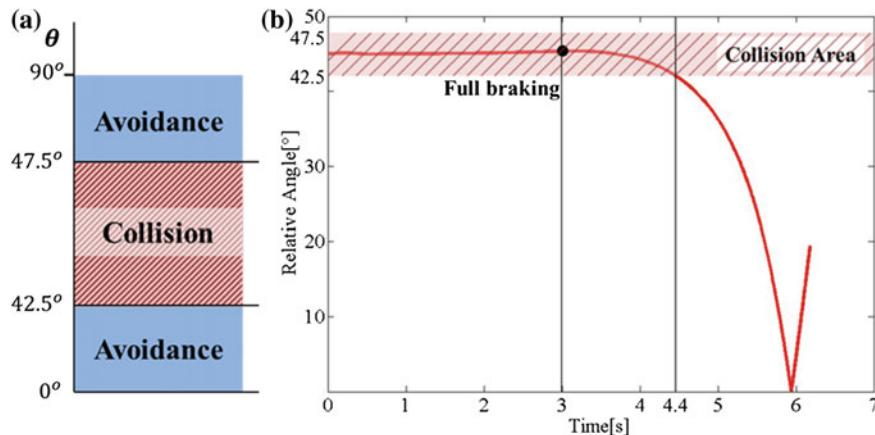


Fig. 2 a Ranges of θ_{min} and θ_{max} . b Change of relative angles in snow asphalt at 60 km/h

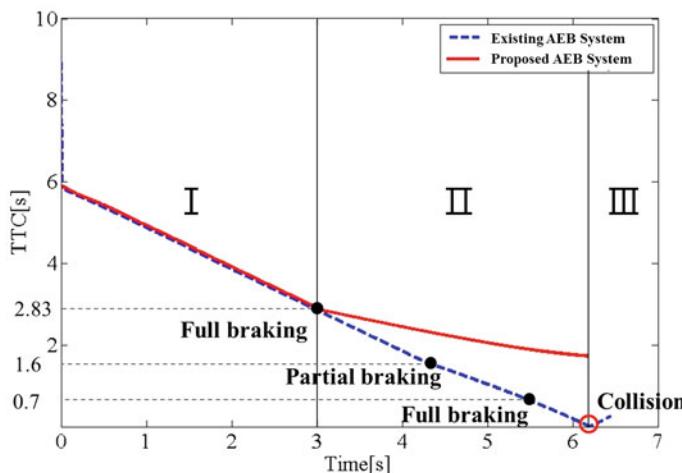


Fig. 3 Change of TTC in road and speed conditions (Snow asphalt, 60 km/h)

4 Conclusion

In this paper, we proposed a braking control method that considers various road conditions in a V2V communication environment and improves the performance of AEB. The proposed AEB system calculates the TTC and relative angle using the information on neighboring vehicles and the Host Vehicle in order to determine intersection collision risks. In addition, it varies the time of braking in accordance with various road conditions.

The results of simulations conducted show that the conventional AEB system had collisions in all road conditions, except for the dry asphalt, while the proposed AEB system was confirmed to have avoided collisions in all of the proposed scenarios. This result, in which no collision occurred, was possible through the use of a control logic that changes the braking time to accord with diverse road conditions. This result confirms that the proposed AEB system has improved braking and collision performance compared to that of the conventional system.

A method for estimating the coefficient of friction of road surfaces will be studied and an additional intersection scenario developed and simulated in the future.

Acknowledgments This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the C-ITRC(Convergence Information Technology Research Center) (IITP-2015-H8601-15-1005) supervised by the IITP (Institute for Information & communications Technology Promotion).

References

1. U.S. Department of Transportation—Federal Highway Administration, Institute of Transportation Engineers: Intersection safety Briefing Sheet (2004)
2. Road Traffic Authority Driver's License Examination Office. <http://www.koroad.or.kr/>
3. Basma, F., Tachwali, Y., Refai, H.H.: Intersection collision avoidance system using infrastructure communication. In: 14th International IEEE Conference on Intelligent Transportation Systems, pp. 422–427 (2011)
4. Ibanez-Guzman, J., Lefevre, S., Mokkadem, A., Rodhaim, S.: Vehicle to vehicle communications applied to road intersection safety, field results. 2010 13th International IEEE Conference on Intelligent Transportation Systems, Funchal, pp. 192–197 (2010)
5. Euro NCAP AEB Test Protocol. <http://www.eriunoncap.com>
6. Geamanu, M.S., Cela, A., LeSolliec, G., Mounier, H.: Road condition estimation and longitudinal control for electric vehicles. In: 11th International Conference on Control, Automation and Systems, pp. 599–604 (2011)
7. Han, I., Luan, B., Hsieh, F.: Development of autonomous emergency braking control system based on road friction. In: IEEE International Conference on Automation Science and Engineering (CASE), Taipei, pp. 933–937 (2014)

Integrated Information Retrieval for Distributed Heterogeneous Ontology Systems

**Sang-Won Hwang, Young-Kwang Nam, Lee-Nam Kwon,
Jae-Soo Kim and Byoung-Dai Lee**

Abstract In this paper, we propose a novel system that integrates heterogeneous semantic web systems based on schema mapping. The user can generate only one SPARQL query using the integrated schema without the necessity of checking the schemas of individual systems each time thereby reducing additional costs to generate queries for individual systems. Furthermore, the user is not required to collect individual query results manually after performing a query and additional costs for reestablishing systems can be reduced because no change in existing system structures is required. If currently established systems are expanded by adding the schema structures of other ontology systems, the cost to establish another integrated retrieval system can be saved. To evaluate our approach, we have implemented a prototype that integrates two national information retrieval systems.

1 Introduction

As the semantic search has been positioned as a killer service, many conventional information retrieval systems have been transformed into semantic web systems. In the early days, individual semantic web systems were built based on their own requirements and operated independently. As a result, users of a particular semantic web system were presented with the information managed only by the system. More recently, user demands for integrated searches over several independently operating semantic web systems have been increasing rapidly. This has occurred because integrated semantic searches enable more meaningful results to be

S.-W. Hwang · Y.-K. Nam (✉)

Department of Computer Science, Yonsei University, Seoul, Korea
e-mail: yknam@yonsei.ac.kr

L.-N. Kwon · J.-S. Kim
KISTI, Daejeon, Korea

B.-D. Lee
Department of CS, Kyonggi University, Suwon, Korea

generated, as information having similar meanings in diverse areas and domains is likely to be used for inference. However, it is not an easy task to integrate physically independent, distributed, and heterogeneous database systems to provide a single, integrated semantic web system to end-users. For physical integration, existing legacy data from participating systems must be transformed according to the integrated schema and whenever new data are accumulated in the participating systems, the transformation process must be repeated.

In this paper, we propose a novel system that integrates heterogeneous semantic web systems based on schema mapping. The proposed system works by first creating integrated schema that includes all the attributes of the ontology schemas of participating semantic web systems (e.g., local schema). In the process, it maintains schema-mapping information that indicates which attribute of the local schema corresponds to that of the integrated schema. Second, user queries are generated against the integrated schema. Third, for query execution, the system re-generates actual queries from the original user query in such a way that the attributes of the integrated schema are replaced with the corresponding attributes of the local ontology schema of the individual semantic web systems using the schema-mapping information.

To evaluate the effectiveness of our approach, we have implemented a prototype that integrates two national information retrieval systems, the National Discovery of Science Leader (NDSL) and the National Science and Technology Information Service (NTIS). The NDSL retains more than one hundred million pieces of information regarding research papers and reports, patents, standards, factual information, etc. Approximately, 3.5 million pieces of data about national projects have been accumulated in the NTIS, but it does not retain information on what research projects' outcome the papers contained in the journals are. On the other hand, the NTIS does contain information about papers and reports for the outcomes of R&D projects carried out since 2008. Therefore, if the NDSL and the NTIS were linked to perform integrated retrieval, the value and the reliability of information from the two systems could be enhanced to be complementary.

2 Related Work

Similar to ours, the ISENS [1] system provides a function to integrate and retrieve different real-world data sources having different ontologies. This system is less useful because queries in this system cannot be answered independently using a single system; instead, integrated queries can be made only to systems composed of mutually complementary data. In [1], mapping information for ontology schemas was not gathered in one place, but instead was made only for two ontologies with fields that can be mapped. Therefore, the existence of mapping information cannot be known without accessing the source system, which means that the mapping information can be accessed only through navigation. The largest difference between the present paper and [1] is that, instead of creating queries appropriate for

individual local sites using mapping information, performing the queries, and integrating the results, in the case of [1], the same query is performed using mapping information, the next system is accessed using the KAON2 reasoner to collect information, and results are presented but each database system cannot be accessed without using the system's source description and the source selection algorithm.

The DARQ system in [2] is also intended to perform integrated retrieval for distributed systems. However, this study is quite different from the study set forth in the present paper. In the DARQ system, heterogeneous data should be accessed using wrappers and the service description describes the kinds of data and access patterns that can be used for individual sites (endpoint) using sets of predicates. The DARQ system is different from the system in the present paper in that it focuses on query optimization using statistical information for integrated retrieval. The SECO system in [3] enables efficient collection of any RDF files existing on the Web and provides interfaces in the form of HTML so that users can easily identify integrated data. This system is composed of a collector, a wrapper, a transformer, a user interface, a remote query interface, and data storage. The data storage is composed of multiple different sets of RDF data. Among them, MetaModel has Metadata information collected from files. SourceModel stores original RDF triples collected from files existing on the web and triples created here are purified through the transformer and stored in the TargetModel thereafter. The TargetModel enables access to user interface for creating HTML and remote query interfaces for query processing. The MetaModel, the UsageModel, and the TargetModel are described as ontologies and the SourceModel is composed of diverse schemas without any particular form.

3 System Architecture

The primary components of the system are the Schema Manager, the Query Manager, and the Result Manager. In what follows, we present detailed information about each component.

3.1 *Schema Manager*

The NDSL and the NTIS define their own ontology schemas. Using the individual schemas, the administrator connects classes and attributes having the same meaning but different names with each other to create integrated schema that can link and retrieve two ontologies together. Therefore, the Schema Manager must maintain the Resource Description Framework (RDF) schema created by the administrator, as well as RDF schemas for individual ontologies of the NDSL and the NTIS. Figure 1 shows an example of how the classes and attributes of the NDSL and the NTIS

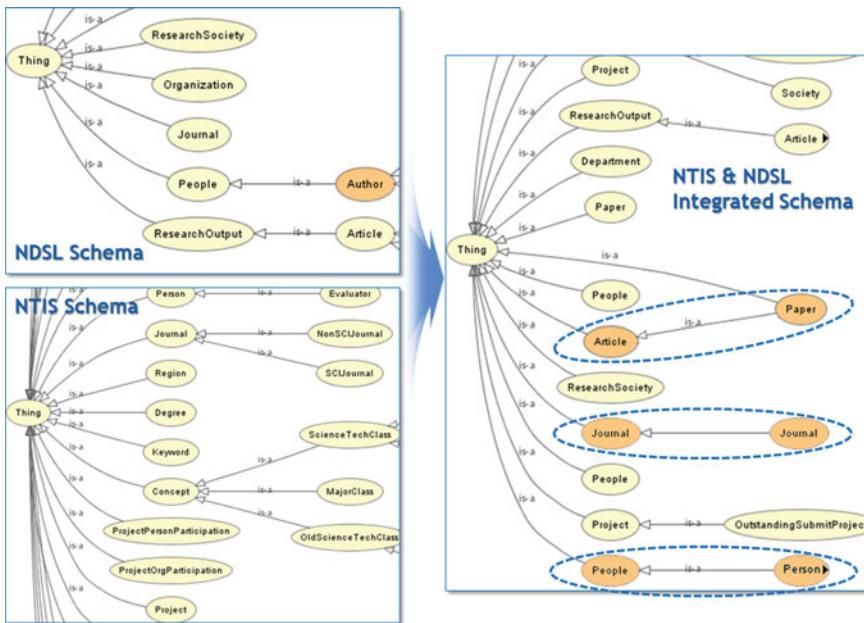


Fig. 1 Schema integration between the NDSL schema and the NTIS schema

schemas are linked to each other. Both the NDSL and the NTIS have information on papers and the authors of the papers in common but they represent the information in different ways. For example, the ‘People’ class in the NDSL schema and the ‘Person’ class in the NTIS schema both indicate a ‘human’, but they are expressed differently. In contrast, both the NDSL and the NTIS schemas have a ‘Journal’ class to represent a ‘journal’. Among the attributes defined in the classes, ‘korName’, ‘hasName’, ‘write’, and ‘hasFirstAuthor’ are the attributes with the same meanings in both schemas.

Therefore, the primary role of the Schema Manager is to maintain information about how individual classes and attributes defined in one schema are mapped to those defined in another schema. Note that classes and attributes in the integrated schema contain the level information, which is the core information needed by the sub-query generator. For example, classes such as ‘Paper’ and ‘Project’ are assigned to level 1, while ‘hasFirstAuthor’ and ‘korName’ attributes are assigned to level 2 and level 3, respectively. The level is related to the linkage of the class or the attribute. The Schema Analyzer analyses the schema information in the single SPARQL query submitted by the users. Then, it compares it with the RDF schemas of individual ontologies. Next, it delivers the information of the corresponding ontologies to the Query Manager.

3.2 *Query Manager*

The user creates queries based on integrated schema regardless of whether the ontologies managed by different information retrieval systems exist. The user query is first analyzed using individual ontology schemas and the mapping schema. Subsequently, sub-queries are generated that are suitable for the corresponding ontologies. The Query Manager consists of the query parser, the query analyzer, and the sub-query generator.

When the query is verified by the query parser, the type of the query is determined through the query analyzer. Using the integrated schema created by the administrator, the query parser checks the validity of the class and the attributes in the query. Then, using the schemas of individual ontologies, the query analyzer determines whether the user query can be transformed into sub-queries for individual ontologies. Finally, the sub-query generator re-generates sub-queries. The sub-query generator and the query type will be dealt with in detail in the next section.

3.3 *Result Manager*

The Result Manager manipulates the intermediate results obtained by performing sub-queries in individual ontology for further processing or preparation for the final results. For example, depending on the query type, the Results Manager uses the results from one sub-query as filter information for another sub-query or it combines the results of sub-queries for the final query result.

4 Sub-query Generation

The query types determined by the query analyzer are automatically generated into four types depending on the content of the information that the integrated query will retrieve.

Type-1 queries represent those queries using the schema that exist only in one ontology. Table 1 shows an integrated query for “all projects that include ‘DB’ in their title” and sub-queries for individual ontology schemas. The integrate query includes the ‘Project’ class, which exists only in the NTIS schema. Therefore, there is no corresponding sub-query for the NDSL.

Type-2 queries include classes and attributes that exist in every ontology. When retrieved using classes and attributes such as ‘Paper’ and ‘Author’ that the NDSL and the NTIS have in common, the query result is in the form of a combination of individual sub-query results. Table 2 shows an example of Type-2 queries, as well as how it is transformed into sub-queries for the NDSL and the NTIS, respectively.

Table 1 Sub-query generation for Type-1 queries

Integrated query	NDSL query	NTIS query
<pre>SELECT ?projName WHERE { ?project rdf:type ys:Project. ?project ys:hasName ?projName. FILTER regex(str(?projName), "DB") }</pre>	<pre>SELECT WHERE { }</pre>	<pre>SELECT ?projName WHERE { ?project rdf:type ntis:Project. ?project ntis:hasName ?projName. FILTER regex(str(?projName), "DB") }</pre>

Table 2 Sub-query generation for Type-2 queries

Integrated query	NDSL query	NTIS query
<pre>Select ?paperName ?personName where { ?paper rdf:type ys:Article. ?paper ys:korTitle ?paperName. ?paper ys:write ?author. ?author ys:korName ?personName. FILTER regex(str(?personName),"서 강") }</pre>	<pre>Select ?paperName ?personName where { ?paper rdf:type ndsl:Article. ?paper ndsl:korTitle ?paperName. ?author ndsl:write ?paper. ?author ndsl:korName ?personName. FILTER regex(str(?personName),"서 강") }</pre>	<pre>Select ?paperName ?personName where { ?paper rdf:type ntis:Paper. ?paper ntis:hasName ?paperName. ?paper ntis:hasFirstAuthor ?author. ?author ntis:hasName ?personName. FILTER regex(str(?personName), "서 강") }</pre>

Type-3 and Type-4 queries have similar basic preconditions but differ from internal processing. For Type-3 queries, intermediate query results are first obtained by using classes and attributes that exist exclusively in one ontology. Then the commonly existing classes and attributes are applied to the intermediate query results. On the other hand, Type-4 queries apply query results obtained using common classes and attributes to queries for exclusive classes and attributes in a certain ontology.

In Type-3 and Type-4 queries, some of the classes and attributes included in the queries exist in every ontologies, but the other classes and attributes are defined only in one ontology. For instance, the ‘Author’ attribute exists both in the NDSL and in the NTIS schemas, whereas the ‘Project’ attributes exists only in the NTIS schema. When queries are received of these types, the sub-query generator first separates the commonly existing classes and attributes from those that exist individually. For those classes and attributes that exist in only one ontology, they will be eliminated from the sub-queries for the ontologies that do not support them. Tables 3 and 4 show examples of these types of queries, as well as how they are transformed into appropriate sub-queries. The example for a Type-3 query is “papers written by those who participated in the project for establishment of a driving safety DB and development of operating technology”. The example for a Type-4 query is “projects in which the author of the paper ‘Diesel Engine’

Table 3 Sub-query generation for Type-3 queries

Integrated query	NDSL query	NTIS query
<pre> SELECT ?authorName ?paperName WHERE {?paper rdf:type ys:Article. ?paper ys:korTitle ?paperName. ?paper ys:hasFirstAuthor ?author. ?author ys:korName ?authorName. ?author ys:participateIn ?project. ?project ys:hasName ?projName. FILTER (regex(?projName,“주행안전 DB ”)) </pre>	<pre> SELECT ?authorName ?paperName WHERE {?paper rdf:type ntis:Paper. ?paper ntis:hasName ?paperName. ?paper ntis:hasFirstAuthor ?author. ?author ntis:hasName ?authorName. ?author ntis:participateIn ?project. ?project ntis:hasName ?projName. FILTER (regex(str(?projName), “주행안전 DB ”)) </pre>	<pre> SELECT ?authorName ?paperName WHERE {?paper rdf:type ndsl:Article. ?paper ndsl:korTitle ?paperName. ?paper ndsl:hasFirstAuthor ?author. ?author ndsl:korName ?authorName. FILTER (regex(str(?authorName), “정재우”) regex(str(?authorName), “유시복”)) </pre>

Table 4 Sub-query generation for Type-4 queries

Integrated query	NDSL query	NTIS query
<pre> SELECT ?authorName ?projName WHERE {?paper rdf:type ys:Article. ?paper ys:korTitle ?paperName. ?paper ys:hasFirstAuthor ?author. ?author ys:korName ?authorName. ?author ys:participateIn ?project. ?project ntis:hasName ?projName. FILTER regex(str(?paperName), “디젤엔진”) </pre>	<pre> SELECT ?authorName WHERE {?paper rdf:type ndsl:Article. ?paper ndsl:korTitle ?paperName. ?paper ndsl:hasFirstAuthor ?author. ?author ndsl:korName ?authorName. FILTER regex(str(?paperName), “디젤엔진”) </pre>	<pre> SELECT ?authorName ?projName WHERE {?paper rdf:type ntis:Paper. ?paper ntis:hasName ?paperName. ?paper ntis:hasFirstAuthor ?author. ?author ntis:hasName ?authorName. ?author ntis:participateIn ?project. ?project ntis:hasName ?projName. FILTER regex(str(?paperName), “디젤엔진”) </pre>

participated”. As shown in Table 3, ‘Project’ and ‘participateIn’ in the integrated query do not appear in the sub-query for the NDSL because the relevant classes and attributes do not exist in the NDSL schema. In the sub-query for the NTIS, the class ‘Article’ is transformed into ‘Paper’ and the attribute ‘korTitle’ is transformed into ‘hasName’.

5 Conclusion

In this paper, we proposed a novel system that will enable storing data from two different ontology systems in one physical system without converting the data to conduct integrated retrieval. Based on the individual schemas of the separated systems, the administrator can connect the schemas that have the same meanings but different forms of expression with each other between the two systems to generate integrated schemas. The user can then generate integrated SPARQL queries utilizing the schemas generated by the administrator to perform queries without recognizing the existence of individual ontologies. Furthermore, the user is not required to collect individual query results manually after performing a query and additional costs for reestablishing systems can be reduced because no change in existing system structures is required. If currently established systems are expanded by adding the schema structures of other ontology systems, the cost to establish another integrated retrieval system can be saved. Although the complexity of applications will increase in this case, it should be a trivial problem compared to the cost to integrate several millions or several dozen millions of triples manually.

Acknowledgment This research was supported by the Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2014M3C4A7030505).

References

1. Abir, Q., Dimitre, A., Jeff, H.: ISENS: A system for information integration, exploration, and querying of multi-ontology data sources. In: Proceedings of the 2009 IEEE International Conference on Semantic Computing, pp. 330–335 (2009)
2. Q. Bastian and L. Ulf, Querying Distributed RDF Data Sources with SPARQL (2008)
3. Andreas, H.: An integration site for semantic web metadata. In: Proceedings of World Wide Web Conference, vol. 1, pp. 229–234 (2004)

hFractal: A Cloud-Assisted Simulator of Virtual Plants for Digital Agriculture

Fei Hao, Doo-Soon Park, Young-Sik Jeong and Jong Hyuk Park

Abstract Virtual Reality (VR) techniques advances the development of research on virtual plants which is regarded as one of the most important parts of Digital Agriculture. At present, there exists a gap that the existing simulation techniques of virtual plants cannot achieve the purpose of real-time drawing by virtue of cloud computing according to captured picture of plants. To fill the this research gap, this paper presents a real-time simulator prototype, termed *hFractal*, of virtual plants in Digital Agriculture. The proposed prototype contains: (1) fast generation algorithm of topological structure of plants; (2) flexible simulation of virtual plants based on Fractals as well as L-systems. In summary, *hFractal* provides a convenient and flexible modeling and simulation environment that is particularly useful for those without an intensive programming background.

Keywords Virtual plant • Fractal • Digital agriculture • Simulator • Cloud computing

F. Hao · D.-S. Park (✉)

Department of Computer Software Engineering, Soonchunhyung University, Asan, Korea
e-mail: parkds@sch.ac.kr

F. Hao

e-mail: fhao@sch.ac.kr

Y.-S. Jeong

Department of Multimedia Engineering, Dongguk University, Seoul, Korea
e-mail: ysjeong@dongguk.edu

J.H. Park

Department of Computer Science and Engineering, Seoul National University of Science and Technology, Seoul, Korea
e-mail: jhpark1@seoultech.ac.kr

1 Introduction

The advancement of computer graphics and Virtual Reality (VR) technology facilities the progress of Digital Agriculture, especially the simulations on topological structure and growth of virtual plants [1, 2]. Virtual Plant, a virtual entity composed of topological structure and growth, is generated by using computer graphics. Therefore, the research on simulation of virtual plant has penetrated into broad fields, such as Digital Agriculture, Ecosystem, and Edaphology and so on.

In Digital Agriculture, the real-time simulation of crop attracts much attention from researchers [2–4]. In order to implement such simulation, the following technical parts are generally carried out: (1) a mathematical model of the virtual plant should be established according to the preliminary knowledge of botany; (2) a fractal process is fed to this mathematical model for generating the resulting virtual plant. Due to the various effects (e.g., environment, random factor) on plant growth, it is urgent to build a simulation platform of virtual plant for assisting farmers understanding and adjusting the growth of plants.

To summarize, our major contributions are twofold: (1) we construct a cloud-assisted architecture of simulator for virtual plant; (2) we develop a cloud-assisted simulator prototype of virtual plants in both desktop and mobile phones by using Delphi 7.0 and Java Andorid toolkit, respectively. Particularly, our client of simulator can capture the pictures of plants and upload them to the cloud in real-time for further fractal processing and simulation of plants. Finally, the virtual plant is sent back to users' clients.

The remainder of this paper is structured as follows. Section 2 presents the overall design architecture of our simulator. The relevant use cases of digital agriculture are presented in Sect. 3. Section 4 concludes this paper with the future work.

2 hFractal Architecture

Figure 1 presents the architecture of our simulation tool for virtual plant *hFractal* that is available for downloading [5]. Clearly, this software architecture is composed of three modules (i.e., image collection, skeleton extraction, and fractal processing modules) in which different functionalities are enabled. We elaborate each module and corresponding functionalities by a bottom-up view approach.

- **Image Collection:** As a bottom layer of the our software architecture, it is a fundamental image sensing module which is in charge of capturing the images of crop by mobile phones, Google Glass and other sensing devices. After images collection, the sensing devices will upload these collected images to the cloud, namely fractal cloud, for further simulation of virtual crop.
- **Skeleton Extraction:** Once the images are offloaded to the cloud, the skeleton extraction module calculates the topological branch structure of the images

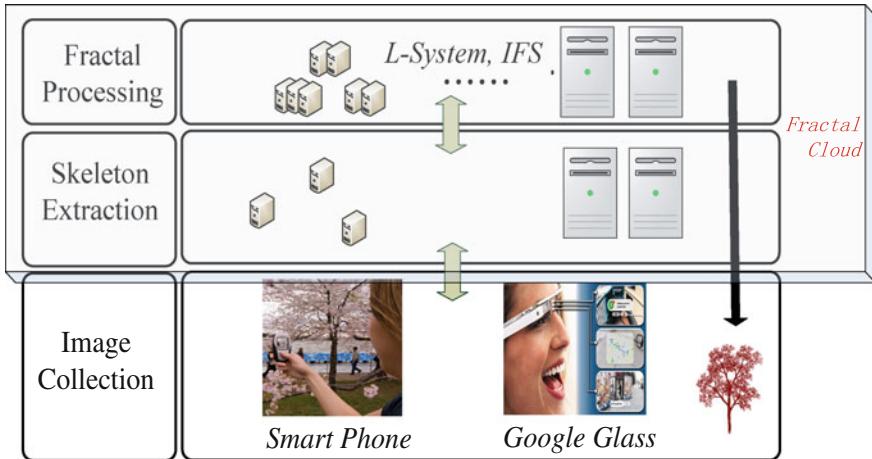


Fig. 1 The Architecture of *hFractal*

according to the graphics processing algorithms by virtual of power fractal cloud. Then, the associated parameters are quickly evaluated.

- **Fractal Processing:** Fractal processing module is the most critical and practical part in the proposed architecture of *hFractal*. It implements the simulation of virtual crop by multiple fractal iterations under *L-System* as well as *IFS* system [6, 7]. Generally, the more times iterations it has, the more accuracy of virtual crop is. Then, the generated virtual crop are sent back to graphic display terminals, such as Google Glasses and Mobile Phones.

3 hFractal: Use Cases

This section firstly shows the usage of *hFractal* and some simulation results of virtual plants like heterogeneous trees and fern leaf by using *L-System* and *IFS* system, respectively. Then, a random fractal scenic is creatively implemented with our simulator *hFractal* as well as the mobile device.

3.1 Heterogeneous Tree

We demonstrate a series of heterogeneous tree simulated by *hFractal*, along with the respective parameters of skeleton of the tree. As shown in Fig. 2, the image of heterogeneous tree is firstly taken by farmer or agronomists' sensing devices and

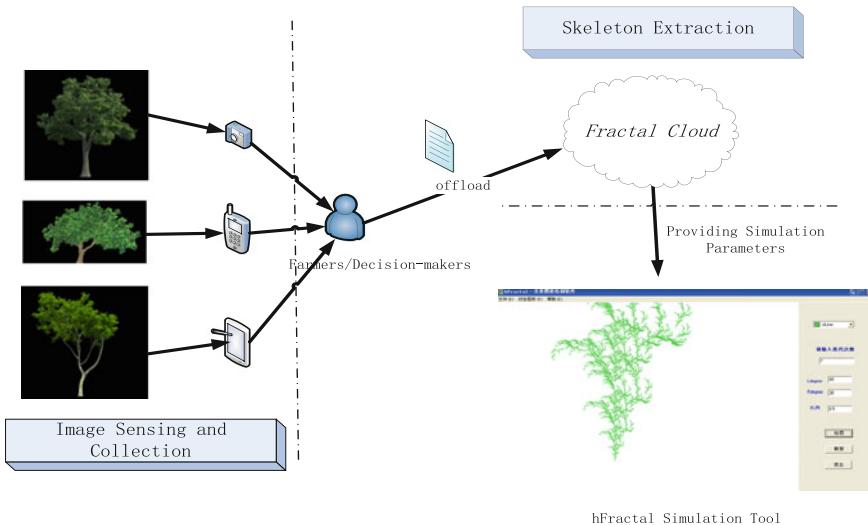


Fig. 2 Simulation of heterogeneous trees and Teng plants using *hFractal* simulator

then sent to the cloud for further processing and simulation. Clearly, skeleton extraction algorithm is applied to the image and the corresponding parameters including the fractal images are obtained. Then, the simulation results are demonstrated with different number of iterations.

Figure 3 presents the simulation results of heterogeneous trees with the different parameters. According to the shape of the simulated heterogeneous trees, it is easily to find that the tree inclines towards the right side if $R_{degree} > L_{degree}$ (shown in Tree2 and Tree4), otherwise, the tree inclines towards the left side (shown in Tree5).

3.2 Fern Leaf

To evaluate the IFS system module about feasibility and effectiveness of *hFractal*, this section also presents a group of fern leaf implemented by *hFractal* with the respective parameters of skeleton of the fern leaf by adopting the working process shown in Fig. 2. For the simulation of Teng plants, skeleton extraction algorithm is applied to the image and the corresponding IFS parameters including the fractal images are listed in Table 1. Then, Fig. 4 shows the results simulated with different number of iterations, i.e., $N = 1000, 10,000, 100,000, 1,000,000$. As can be seen from Fig. 4, the simulated fern leaf is getting closer to the real fern leaf as the iteration increases.

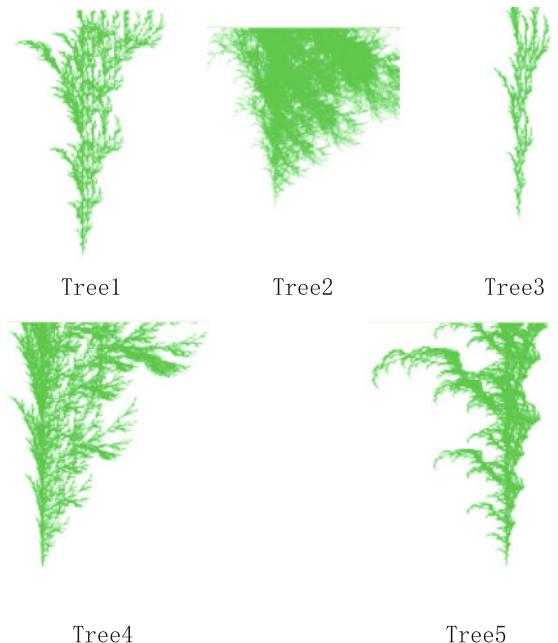


Fig. 3 Simulation results of heterogeneous trees with different parameters

Table 1 IFS code of fern leaf

	a	b	c	d	e	f	g
w_1	0	0	0	0.16	0	0	0.01
w_2	0.85	0.04	-0.04	0.85	0	1.6	0.85
w_3	0.2	-0.26	0.23	0.22	0	1.6	0.07
w_4	-0.15	0.28	0.46	0.24	0	0.44	0.07

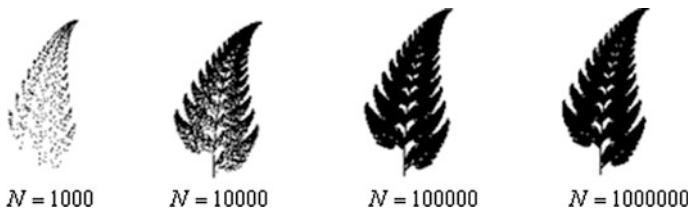


Fig. 4 Simulation results of fern leaf based on different iterations

The detailed simulation algorithm for fern leaf is as follows,

- 1) Define a two-dimensional array $d[i, j]$, and store the above IFS code in this array.
- 2) Set a pixel amplifier $(x, y) = (1000, 1000)$
- 3) Set a loop variable i from 0 to 18000;
- 4) Define a generator of random interval $k = \text{random}(i) + 1$

$$x = d[k,1]*\text{temp}x + d[k,2]*y + d[k,5]$$

$$y = d[k,3]*\text{temp}x + d[k,4]*y + d[k,6]$$
- 5) $x' = \text{round}(x * j) + \Delta x$
 $y' = \text{round}(y * j) + \Delta y$
- 6) Scan each pixel and draw the fern leaf eventually
 where $\text{temp}x$ denotes the temporal information of x ; Δx and Δy refer to the translational distance on the screen.

3.3 Random Fractal Scenic

The mountains, rivers, mainland and islands, are a product of nature itself, they cannot be described with classical Euclidean geometry. However, these natural objects own the self similarity statistically, so called fractal feature. This section attempts to extend the simulation of virtual plants for creating a piece of random fractal scenic. The basic solution idea is that all objects appear in the random fractal scenic are simulated based on IFS coding, then reset the layout of these simulated results.

This creative random fractal scenic contains mountain, various trees (e.g., palm, and gray tree) and sun. Based on requirements of IFS system, the IFS codes for each object are obtained and can be shared upon readers' request.

After obtaining the IFS codes from cloud, *hFractal* simulates and arranges the fractal images of each object. Finally, to support the better mobile experience of users, android-based *hFractal* application and random fractal scenic are implemented as shown in Fig. 5.

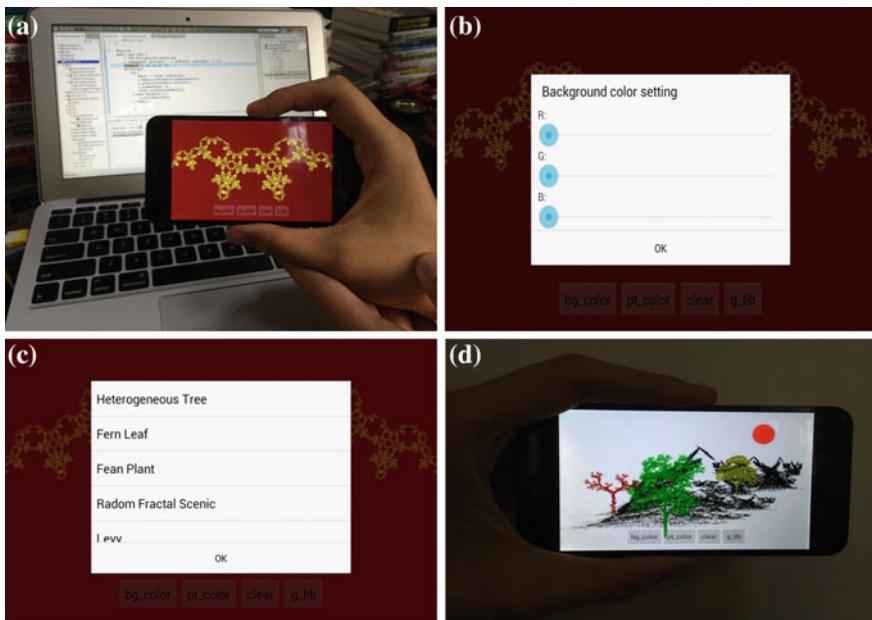


Fig. 5 The android application interfaces of *hFractal* and random scenic. **a** The crown fractal. **b** Background color setting. **c** Fractal image library. **d** Random fractal scenic

4 Conclusions

Aiming to simulate the virtual plants in Digital Agriculture, this paper presents a cloud-assisted simulator *hFractal* of virtual plants based on L-system and IFS system. This simulator *hFractal* firstly integrates the fractal arts and digital agriculture based on advanced cloud computing technology. It generates not only the fundamental fractal images, but also creates more complex objects in natural. Regarding to the heterogeneous trees and fern leaf, L-system and IFS system based simulation algorithms are provided, respectively. Further, *hFractal* extends the simulation of virtual plants for creating a piece of random fractal scenic by arranging and coloring the simulated objects implemented with IFS code. Simulated results show that our simulator *hFractal* has a better feasibility and extensibility compared with other previous works. In the future, we are going to conduct the skeleton algorithms of plants and integrate into our system.

Acknowledgments This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. NRF-2014R1A1A4A01007190).

References

1. Liu, Y., Pan, J., Yang, L., Zhu, X., Zhang, N.: Visualization of virtual plants growth based on open L-system. In: Computer and Computing Technologies in Agriculture V, pp. 90–96. Springer, Berlin (2012)
2. Tang, W., Li, J., and Hu, X.: Notice of retraction virtual simulation of plant with individual stem based on crop growth model. In: International Conference on Natural Computation, pp. 2377–2380 (2011)
3. Stella, T., Frasso, N., Negrini, G., Bregaglio, S., Cappelli, G., Acutis, M., Confalonieri, R.: Model simplification and development via reuse, sensitivity analysis and composition: a case study in crop modelling. *Environ. Model Softw.* **59**, 44–58 (2014)
4. Ding, D., Fang, K., Jing, S., Bo, L., Bo, Q., Yu, H.: Virtual medical plant modeling based on L-system. *Afr. Health Sci.* **14**(4), 1056–1062 (2015)
5. <https://kanbox.com/f/0I0VV> (hFractal Version 2.0)
6. Wang, H., Hao, F.: Adapted algorithm of virtual plants simulation based on stochastic L-system. *Int. J. Comput.* **9**(2), 165–174 (2010)
7. Ijiri, T., Owada, S., Igarashi, T.: The sketch l-system: global control of tree modeling using free-form strokes. In: Smart Graphics, pp. 138–146. Springer, Berlin (2006)

Proposal and Validation of AEB System Algorithm for Various Slope Environments

Ming Lin, Jaewoo Yoon and Byeongwoo Kim

Abstract This study was conducted to propose and validate an algorithm designed to improve the braking application timing of the autonomous emergency braking (AEB) system. Proposed algorithm using the time-to-collision (TTC) index based on a distance sensor to improve the performance of algorithm in a multi-slope environment. The current AEB system does not reflect changes in slope angle, which act as a limiting factor on the braking function. To address this problem, this study proposes an AEB algorithm that reflects the gradients and thus improves the braking performance in a slope environment. The minimum braking distance necessary for braking, and the braking application timing, were determined by analyzing the force exerted on the vehicle by a slope. The collision-avoiding efficiency of the proposed algorithm was validated in an algorithm performance test.

Keywords AEB (Autonomous emergency Braking) · Collision avoidance · TTC (Time to Collision) · Slope · ASS (Active safety Systems)

M. Lin · J. Yoon

Graduate School of Electrical Engineering, University of Ulsan, 93 Daehak-ro, Ulsan,
Republic of Korea
e-mail: flaaud159@gmail.com

J. Yoon

e-mail: jewos0127@gmail.com

B. Kim (✉)

School of Electrical Engineering, University of Ulsan, 93 Daehak-ro, Ulsan, Republic of
Korea
e-mail: bywokim@ulsan.ac.kr

1 Introduction

In recent years, extensive studies have been conducted on active safety systems (ASS) [1]. In particular, the autonomous emergency braking (AEB) system is a system in which a vehicle can autonomously sense an emergency situation and apply its brakes [2]. Euro NCAP analyzed real accident data and reported that the EAB system is expected to reduce the accident rate by 27 % [3].

In previous studies, the AEB system was tested under many different scenarios and in various traffic environments such as crossroads and curved roads [4–6]. The AEB performance tests were mostly performed on flat-surface driving environments. An AEB algorithm designed for such flat-surface driving environments encounters difficulties in slopes because the dynamic effect of a slope cannot be sufficiently reflected in the calculation. To overcome this problem, this paper presents an accurate braking-application timing algorithm to successfully avoid collisions in emergency situations by analyzing the dynamics of a vehicle under the effect of a downward slope. In addition, the efficacy and reliability of the proposed algorithm will be tested in simulations under various slope and velocity conditions, thus proving the advantage of the proposed algorithm over conventional algorithms.

2 Braking Application Time Configuration According to Slope

In this paper, a new TTC_{need} is provided as the ratio of slowing-down length to relative velocity. This determines the braking application time. In previous studies, the time-to-collision (TTC) was calculated as the ratio of relative distance to relative velocity. Full braking is applied at a $TTC \leq 0.9$ s [7], as Formula (1) where V_{rel} is the relative velocity and S_{rel} is the relative distance of the two vehicles involved.

With the risk index calculated by Formula (1). However, a collision cannot be avoided unless the friction force of the road in question is the same as that exerted by a flat road. In reality, the vehicle cannot achieve the ideal braking performance owing to environmental variables, as expressed by Formula (2). In this paper, only considering the slope of the road. Using Newton's Law of Motion to calculate the minimum brake distance.

Based on the maximum friction force, the limit maximum deceleration rate is calculated using Formula (3). As Formula (3) shows the road slope is considered into algorithm and effect braking application time. The minimum slowing-down length is calculated using Formula (4). The ratio of the minimum slowing-down length to the relative velocity, i.e. the new braking application time, is determined using Formula (5).

$$TTC = \frac{S_{rel}}{V_{rel}} \quad (1)$$

$$F_{deceleration} = -\mu \cos(\theta)mg + mg \sin(\theta) \quad (2)$$

$$\frac{dv}{dt} = -\mu \cos(\theta)g + g \sin(\theta) = a_{limit} \quad (3)$$

$$S_{need} = \frac{v^2 dt}{2dv} \quad (4)$$

$$TTC_{need} = \frac{S_{need}}{V_{rel}} \quad (5)$$

If the condition of $TTC \leq TTC_{need}$ is satisfied, then algorithm judge vehicle has risk to collision with forward obstacle. At this time full braking will apply to vehicle and the collision is avoided.

3 Simulation and Experimental Results

PreScan and MATLAB were used to construct simulation environments. The simulation configurations for vehicle type, sensor, and road environment were done in PreScan. The PreScan vehicle's braking was configured to be implemented via linkage established by the AEB logic provided in MATLAB.

The test scenario was configured as a situation in which the ego vehicle driving at a constant speed on a slope is approaching a stopped preceding vehicle. In order to verify the performance of the AEB system algorithm in various traffic environments, various velocity and slope conditions were applied. The velocities of the ego vehicle were set at 60, 70, 80, 90, and 100 km/h, referencing the ADAC test speeds [6]. Slopes selected were 0, 4, 8, 11, 14, and 17 % between the minimum possible slope 0 % and the maximum permissible slope of 17 %, as specified in the road design standards [8].

The graphs in Fig. 1 represent the simulation results obtained by applying the conventional AEB algorithm. Given that the applied algorithm defines the critical point for full braking as $TTC \leq 0.9$ s, the ego vehicle driving at 60 km/h can avoid a collision in slope conditions of 0 and 4 %, but a collision occurs in all other slope conditions (8–17 %) despite deceleration of the vehicle. This demonstrates the limitation of the conventional AEB algorithm in avoiding collisions.

Figure 2 shows the simulation results obtained by applying the new AEB algorithm proposed in this paper. The graphs therein demonstrate that the ego vehicle avoids collisions in all slope conditions by adjusting the braking application time to the slope. The ego vehicle stopped at a TTC range of 0.3–0.4 s at 0 % slope and in a TTC range of 0.4–0.6 s at 4–17 % slopes.

Fig. 1 Conventional AEB system simulation results

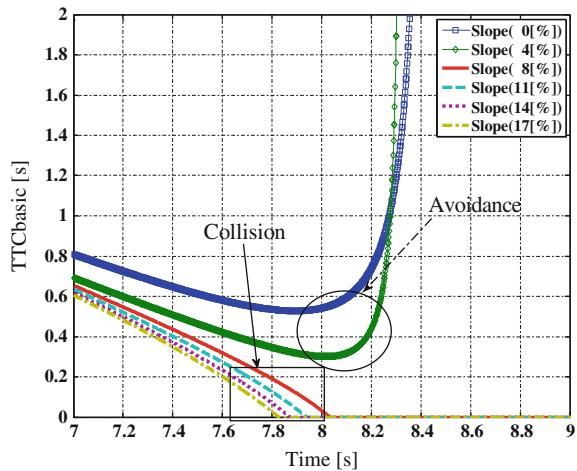


Fig. 2 Proposed AEB system simulation results

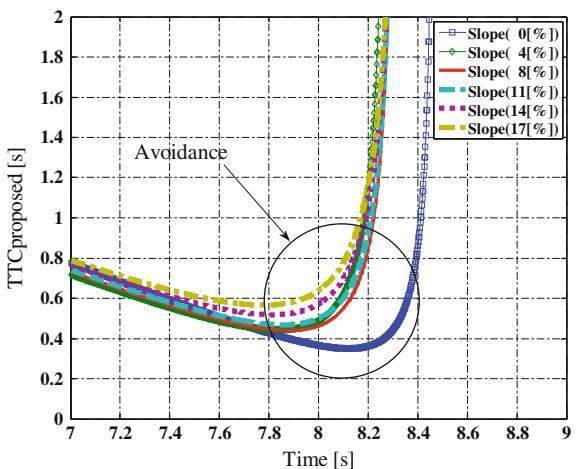


Table 1 shows the performances of the conventional and proposed AEB systems in cross-check tests under various velocity and slope conditions. In the conventional AEB system, a collision was avoided only at 0 and 4 % slopes at the 60 km/h velocity condition. In the AEB system proposed in this paper, collisions were avoided in all slope and velocity conditions, thus demonstrating the accuracy and reliability of the proposed algorithm.

Table 1 Simulation results of the conventional and proposed AEB systems

AEB system	Velocity (km/h)	Slope (%)					
		0	4	8	11	14	17
Conventional algorithm	60	A	A	C	C	C	C
	70	C	C	C	C	C	C
	80	C	C	C	C	C	C
	90	C	C	C	C	C	C
	100	C	C	C	C	C	C
Proposed algorithm	60	A	A	A	A	A	A
	70	A	A	A	A	A	A
	80	A	A	A	A	A	A
	90	A	A	A	A	A	A
	100	A	A	A	A	A	A

A avoidance; C collision

4 Conclusion

This paper presented a new AEB algorithm that takes slopes into account and validated its efficacy by comparing its performance with that of the conventional AEB algorithm in simulations under various slope and velocity conditions. The simulation results revealed that when the conventional algorithm was applied, the ego vehicle collided in all slope and velocity conditions except for the 0–4 % slope at 60 km/h, whereas collisions were avoided in all slope and velocity conditions in the proposed AEB system, owing to the new algorithm capable of adjusting the braking application time to the slope. But the proposed algorithm has a limitation, which do not considering the road friction factor and road lateral angle.

In the simulations, it was verified that the proposed AEB algorithm improved the performance of the conventional AEB algorithm. A further study is envisioned in order to develop an algorithm that takes into account road conditions and slope angles simultaneously.

Acknowledgments This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the C-ITRC (Convergence Information Technology Research Center) (IITP-2015-H8601-15-1005) supervised by the IITP (Institute for Information & communications Technology Promotion).

References

1. Jang, H., Cho, S., Yong, B.: The Safety Evaluation Method of Advanced Emergency Braking System, pp. 162–168. KSAE, Korea (2013)
2. Woo, J., Kim, M., Lee, S.: Study on the Test Method of AEB and FCW System, pp. 1160–1163. KSAE, Korea (2010)

3. Euro Ncap. <http://www.euroncap.com/en>
4. Wesley, H., Iain, K., Alix, E., Matthew, A., Colin, G.: Autonomous Emergency Braking Test Results. Thatcham Research, UK (2013)
5. Cho, H., Kim, B.: A Study on Cooperative Intersection Collision Detection System Based on Vehicle-to-Vehicle Communication. *Adv. Sci. Technol. Lett. Electr. Eng.* 121–124 (2014)
6. Cho, H., Kim, B.: A Study on the Improvement of Cooperative Collision Warning System on Curve Road Based on Vehicle-to-vehicle Communication, Korea (2013)
7. Comparative test of advanced emergency braking systems. Test report, ADAC Vehicle Testing (2013)
8. Korea Ministry of Government Legislation. <http://www.law.go.kr/main.html>

A Study on the V2V-Communication-Based AEB System for High-Speed Driving Under Various Road Surface Conditions

Hyeonggeun Mun and Byeongwoo Kim

Abstract In this study, the performance of braking intervention point of time achieved by autonomous emergency braking (AEB) of a vehicle is travelling at high speeds on a curved road under various road surface conditions was analyzed using vehicle to vehicle (V2V) communication with other vehicles ahead. To calculate a braking intervention point of time, the concept of time to collision (TTC) was applied. TTC was calculated by using the velocities and relative distances between the vehicle and other vehicles. Furthermore, braking intervention point of time was calculated by considering the friction factor corresponding to the road surface conditions. The velocity of the vehicle and the change in road curvature, which are required for the performance analysis, were incorporated in the simulation scenarios. Excellent collision avoidance performance was confirmed for high-speed driving when the braking intervention time point of the vehicle was changed according to the road surface conditions.

Keywords Vehicle to vehicle (V2V) · Autonomous emergency braking (AEB) · Time to collision (TTC) · Curve road · Road friction · High speed

H. Mun

Graduate School of Electrical Engineering, University of Ulsan, 93 Daehak-ro,
Ulsan, Republic of Korea
e-mail: mhg0005@gmail.com

B. Kim (✉)

School of Electrical Engineering, University of Ulsan, 93 Daehak-ro,
Ulsan, Republic of Korea
e-mail: bywokim@ulsan.ac.kr

1 Introduction

According to the 2007–2011 data of the Road Traffic Authority in South Korea, traffic accidents caused by the negligence of drivers accounted for about 55 % of the total traffic accidents [1]. Furthermore, the difference between the traffic accident rates under icy road and dry road conditions was 15.2 %, which was larger than the difference between the rates for straight and other roads [1]. Studies are being carried out on various advanced driver assistant systems (ADASs) with respect to ensuring the safety and prevention of accidents caused by driver negligence. Among the ADASs, the autonomous emergency braking (AEB) system is a typical collision avoidance system. In recent times, to overcome the limitation in the sensor detection range of sensor-based AEB systems, vehicle to vehicle (V2V) communication technology is being applied [2, 3]. Further, studies on various road driving environments are being carried out by using V2V communication technology [4, 5]. The conventional AEB systems reflect the Euro NCAP's (The European New Car Assessment Programme) AEB system evaluation standards [6]. However, they have the limitation that the driving speed of the vehicle, various shapes of road, and various road surface conditions are not taken into consideration [7]. To overcome this limitation, an AEB system that takes into account various road surface conditions was proposed [8]. However, because the vehicle is considered to be travelling in a low-speed environment only, it was difficult to verify the performance of the AEB system in a high-speed environment to determine its usefulness.

Therefore, it is necessary to study the performance of the AEB system for a vehicle travelling at a high speed under various road surface conditions. This study aims to analyze the usefulness of the system by verifying its performance; for this purpose, the braking intervention time point is changed on the basis of V2V communication during high-speed driving on a curved road under various road surface conditions.

2 The AEB System: Braking Intervention Time Point Control

In this study, an algorithm in which the braking intervention time points under various road surface conditions are different for avoiding collision with a vehicle ahead. Figure 1 shows a flowchart of the algorithm for the AEB system. For determining the braking intervention time point, the time to collision (TTC) with a vehicle ahead and the braking intervention time point (TTC_{brake}) for avoiding a collision with a vehicle ahead were compared. Full braking was applied at the instant that TTC became smaller than TTC_{brake} .

Thus, TTC and TTC_{brake} have to be known to determine the braking intervention time point of the AEB system. They can be determined using Eqs. (1) and (2). Equation (1) calculates TTC using the relative distance (S) and relative velocity

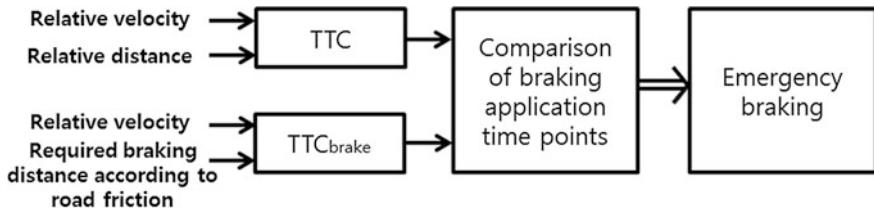


Fig. 1 Flowchart of AEB system

(V) of the vehicle and of a vehicle ahead. Through this, the time to collision (TTC) with a front vehicle can be derived.

$$TTC(s) = \frac{S}{V} \quad (1)$$

Equation (2) gives the braking intervention time point for a vehicle, which is calculated using the braking distance (S_{brake}) and relative velocity (V) according to the road surface condition. The maximum deceleration (a_{μ}) for deriving the braking intervention time point according to the road surface condition and the braking time according to the maximum deceleration need to be known.

The maximum decelerations required for braking under various road surface conditions are calculated using Eq. (5). Equation (4) calculates the time required to brake depending on the maximum deceleration. Equation (3) is an equivalent velocity motion equation. The distance required for braking can be obtained through Eqs. (4) and (5). TTC_{brake} can be derived through the relative velocity and the distance required for braking calculated through Eq. (3).

$$TTC_{brake}(s) = \frac{S_{brake}}{V} \quad (2)$$

$$S_{brake}(m) = Vt + \frac{1}{2}at^2 \quad (3)$$

$$t(s) = \frac{V}{a} \quad (4)$$

$$a_{\mu}(m/s^2) = -\mu \times g \quad (5)$$

3 Simulation and Results

3.1 Simulation Scenario

In vehicle driving scenarios, a vehicle driving on a curved road at a high speed recognizes a stationary vehicle ahead of it, and then the AEB system is operated according to the road surface condition using V2V communication. The speed of the vehicle was set to 100 km/h corresponding to the speed limit of high-speed driving as per the South Korean road traffic law. The road surface conditions of the curved road ranged from a snow-covered road ($\mu = 0.3$) to a dry road ($\mu = 1$). The curvature radius of the curved road was set to 300 m, which is the minimum radius to ensure that a vehicle does not skid. For a given vehicle speed, simulation was carried out by changing the road surface conditions.

3.2 Simulation Results

The simulation results of the AEB system for various driving scenarios were checked. Table 1 lists the braking intervention time points derived using Eq. (2) for various road surface conditions. As the road friction factor increases, the maximum deceleration increases. Because the required braking distance becomes shorter as the deceleration increases, it is confirmed that the required braking intervention time point is delayed.

Figure 2 presents the AEB system operation using the relative distances according to time for the given road surface conditions. It is confirmed that collision with a vehicle ahead was prevented under various road surface conditions because the braking intervention time point was changed depending on the road surface condition.

Table 2 lists the simulation results for the scenarios described in Sect. 3.1. It presents the relative distances from the vehicle ahead after the AEB system was operated on all road surface condition in the simulation. Since the braking starts as per the required braking distances for various road surface conditions, the vehicle stops without colliding by maintaining some distance from the vehicle ahead. Through Table 2, it is confirmed that collision with the vehicle ahead is prevented under all road surface conditions.

Table 1 Braking intervention time points (s) according to the road surface conditions during high-speed driving

Velocity (km/h)	Road friction (μ)							
	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
100	4.72	3.54	2.83	2.36	2.02	1.77	1.57	1.42

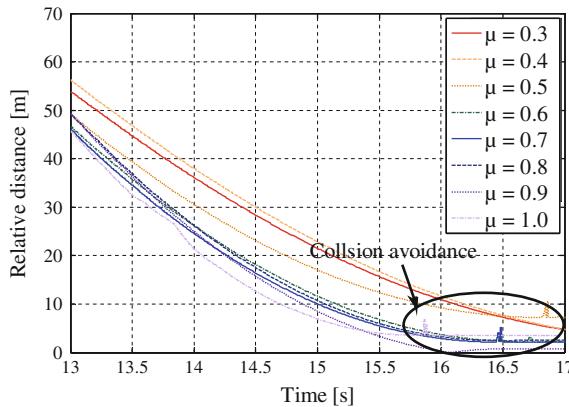


Fig. 2 Relative distance after AEB system operation according to time

Table 2 Relative distance (m) with the vehicle ahead after AEB system operation

Velocity (km/h)	Road friction (μ)							
100	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
	3.11	2.52	5.25	2.19	2.19	2.48	0.77	3.57

4 Conclusion

In this study, a performance analysis was carried out for the braking intervention time point of an AEB system based on V2V communication for a vehicle traveling at a high speed and under various road conditions. The braking intervention time points were derived using the relative velocity and distance required for braking according to the road surface conditions. The simulation results confirmed that collision with a stationary vehicle is prevented when the braking intervention time point was varied during high-speed driving depending on the road surface conditions. Thus, it was confirmed that the performance of the algorithm of the AEB system was useful for high-speed driving conditions.

Therefore, if this algorithm is applied to vehicles intended for high-speed driving, the instances of automobile accidents can be reduced.

In a follow-up study, the performance analysis of the AEB system will be carried out under different road scenarios by considering the gradients for various road surface conditions.

Acknowledgments This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the C-ITRC (Convergence Information Technology Research Center) (IITP-2015-H8601-15-1005) supervised by the IITP (Institute for Information & communications Technology Promotion).

References

1. Korea Road Traffic Authority. <http://www.koroad.or.kr>
2. Kim, N., Lee, J., Soh, M., Kwon, J., Hong, T., Park, K.: Improvement of Longitudinal Safety System's Performance on Near Cut-in Situation by Using the V2V, pp. 747–755. The Korean Society of Automotive Engineers (2013)
3. Ibanez-Guzman, J., Lefevre, S., Mokkadem, A., Rodhaim, S.: Vehicle to vehicle communications applied to road intersection safety, field results. In: 2010 13th International IEEE conference on Intelligent Transportation Systems, pp. 192–197 (2010)
4. Cho, H., Kim, B.: Study on cooperative intersection collision detection system based on vehicle-to-vehicle communication. *Adv. Sci. Technol. Lett. Electric. Eng.* 121–124 (2014)
5. Cho, H., Kim, B.: Performance improvement of collision warning system on curved road based on inter-vehicle communication. *Math. Probl. Eng.* (2014)
6. Euro NCAP AEB Test Protocol. <http://www.eriuncap.com>
7. Son M., Nam C.: Performance Analysis Simulation of Active Emergency Braking System under Various Road Condition, pp. 1828–1832. The Korean Society of Automotive Engineers (2012)
8. Kang, T., Yoo, W., Kim, N., Soh, M., Kwon, J., Hong, T., Park, K.: A Study on the Brake Time for AEB System Considering Road Condition, pp. 737–744. The Korean Society of Automotive Engineers (2014)

Implementation of Power off Recovery Scheme for Block Mapping FTL

Ji-Hwan Chung and Tae-Sun Chung

Abstract Flash memory is widely used in portable devices and storage system. But flash memory has its own characteristics compared to traditional hard disk. This feature is erase-before-write and different unit of read/write and erase. For these reasons, flash memory has FTL (Flash Translation Layer) system software. Up to now many FTL algorithms have been suggested, but power off recovery scheme has not been considered. In this paper, we suggest a power off recovery scheme for block-mapping FTL. And this recovery scheme can be applied to many FTL algorithms.

Keywords Flash memory · FTL · Power off recovery · File system

1 Introduction

Flash memory is widely used in portable devices and storage system. As flash memory has read, write and erase operations and its speed is faster than that of hard disk and it has characteristics of shock resistance, small size, and so on. Particularly, flash memory has erase-before-write architecture [1]. When file system issues the update operation at previously written location, erase operation should be performed before update data write operation. In this reason, flash memory has a FTL (Flash Translation Layer) [2] system software, this software is located in between file system and flash memory.

The general FTL algorithm is to translate of logical addresses to physical addresses, and then there exist three types of address mapping algorithms: page mapping scheme, block mapping scheme and hybrid mapping scheme [3]. In brief,

J.-H. Chung (✉) · T.-S. Chung

Department of Computer Engineering, Ajou University, Suwon, Korea
e-mail: batoski10@ajou.ac.kr

T.-S. Chung
e-mail: tschung@ajou.ac.kr

the page mapping scheme is to translate the logical page number to physical page number and the block mapping scheme is to translate the logical block number to physical block number and finally, the hybrid mapping scheme is mixture of page and block mapping scheme.

In this paper, we suggest a recovery scheme that can be applied to block mapping FTL. Up to now, researches of FTL algorithms are addressed but these algorithms do not consider the power off recovery [4].

This paper is organized as follows. Section 2 covers background and assumption. Section 3 shows our power off recovery scheme. Finally, Sect. 4 concludes the paper.

2 Background

2.1 Power off Recovery Problem

Traditional database recoveries are always maintain data consistency and atomicity. In view of this, flash memory also should maintain the properties. The FTL, only the write operation is related on these properties. Write operation can be divided into two cases as follows. First is the simple write operation without the erase operation. Second is the write operation with the erase operation [4]. We suggest a recovery scheme in these two cases.

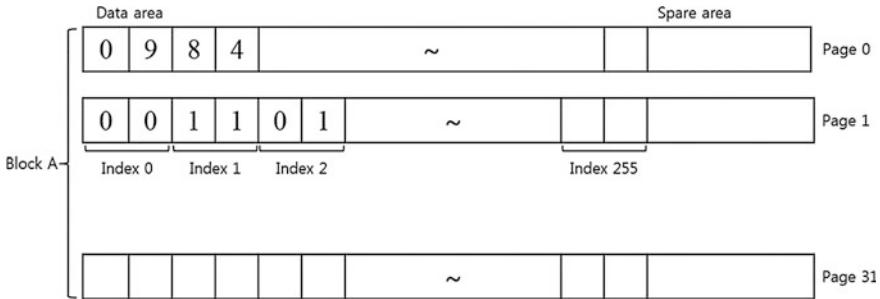
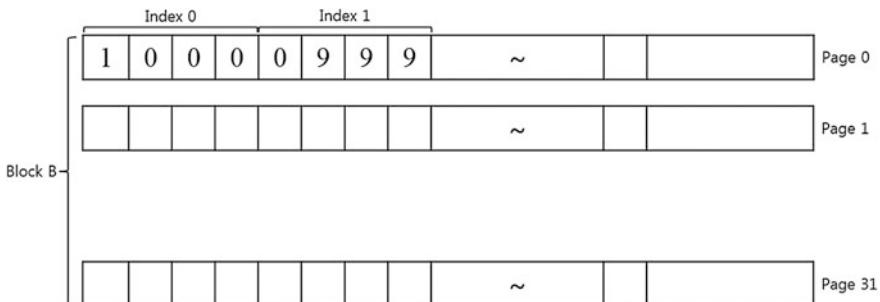
2.2 Assumptions

- Recovery algorithm is two dedicated blocks are used for store metadata and user data cannot be stored in the dedicated blocks.
- For convenience, two particular blocks called a Block A and Block B.
- Block A stores the location address of block mapping and valid information.
- Block B stores the block address mapping information.
- Write operations of flash memory are sequential order.

3 Power off Recovery Scheme

3.1 Design Structure of Block and Page

First of all, flash memory consists of a set of many blocks, and a block is composed of a set of 32 pages. One page (528 byte) is configured of data area (512 byte) and spare area (16 byte). Figure 1 shows a design structure of Block A. The first page of

**Fig. 1** Block A structure**Fig. 2** Block B structure

Block A is storing the block number of saved the block mapping information. In page 0, block number of saved mapping information is stored. From page 1 to page 31, the valid marks of all block (containing Block A and B) are stored. For example, if block pointed by index 0 is valid, “11” are stored. And its value is initial value. However, when block is invalid (case of a block space is full), set to “00”. Finally, when block is performed erase operation, set to “01”.

Figure 2 shows the design of Block B. As previously mentioned that the number of block address was stored in each index. With this information, we can create the mapping table. In this paper, we design the mapping table through the information of the Block B.

Figure 3 shows that index of Block B is become LBN (Logical Block Number) of mapping table. Also value of each index is become PBN (Physical Block Number). For example, in the PBN 1000 is mapped to LBN 0. So this information is represented by value at the Block B and Index 0. And value of Index 0 located in Block A, it is represented about PBN 1000 is valid or invalid.

LBN: Logical Block Number
PBN: Physical Block Number

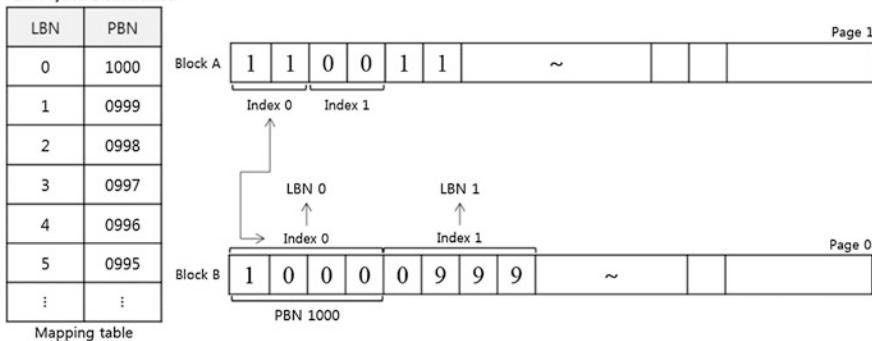


Fig. 3 Mapping table



Fig. 4 Page structure

3.2 Power off Recovery Scheme

Basic write operation. The basic write operation does not require the erase operation. Thus, recovery of this operation is very simple. In brief, when the write operation is performed, log record is written first and the write operation is performed. After the write operation succeeds, invalid mark is stored in the page spare area.

Figure 4 shows the spare area of the page, and S_0 is the valid mark of page. If $S_0 = 1$, it means valid page and the write operation is possible. When suddenly power off occurs while performing write operations, recovery is performed through the log record. After the write operation was success, S_0 is changed by 0.

Erase and write operation. Erase operation is performed by request of write operation. Erase operation algorithm is given by next page. When free page in the block was not existed (case of a block index is “00”), the erase operation is performed. And then, a free block to copy the previous block is searched. If algorithm finds a free block, it rewrites the valid mark of PBN to “01”. Then copy operation of the data to free block is performed. When copy operation is finished, the erase operation is performed. Finally, the rewrite operation of the valid mark is performed and valid mark of previous PBN is changed by “11”. And mapping table should be changed.

```

Algorithm Write operation request Erase operation
Input PBN(Physical Block Number)
Output None
Procedure recovery
If Search the free page then
  If There are no exist free page then
    Search the free block
    If exist free block then
      rewrite PBN valid mark 01
      Write operation(PBN)
    endif
  endif
endif
Erase operation(PBN)
rewrite PBN valid mark 11

```

When the sudden power off occurs while performing the erase and write operations, system recovery perform scheme is as follows: (1) when the system is booted, FTL will check the valid mark of all blocks. (2) If find an invalid mark “01”, this means that the erase operation was failed. (3) Thus FTL performs the erase and write operations again. (4) Finally, the rewrite operation of the valid mark is performed and recovery scheme was end.

4 Conclusion

In this paper, we suggest a power off recovery scheme, and this scheme is based on the block mapping. And two types of the write operations are applied. In case of basic write operation, our scheme applies the recovery scheme through the log record and valid mark. Other case of recovery operation, our scheme is required only valid mark of PBN (Physical Block Number). Therefore our recovery scheme is very easy. For this reason, this recovery scheme can be applied many FTL algorithms.

Acknowledgments This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2013R1A1A2A10012956, 2013R1A1A2061390).

References

1. Samsung Electronics: NAND Flash Memory & Smartmedia Data Book (2007)
2. Ban, A.: Flash file system. United States Patent No. 5,404,485 (1995)
3. Chung, T.-S., Park, D.-J., Park, S.-W., Lee, D.-H., Lee, S.-W., Song, H.-J.: A survey of flash translation layer. *J. Syst. Archit.* **55**, 332–343 (2009)
4. Chung, T.-S., Lee, M.-H., Ryu, Y.-S., Lee, K.-S.: PORCE: an efficient power off recovery scheme for flash memory. *J. Syst. Archit.* **54**, 935–943 (2008)

PLC Monitoring and Protection for SCADA Framework

Ahmed AlShemeili, Chan Yeob Yeun and Joonsang Baek

Abstract Supervisory Control and Data Acquisition (SCADA) is a set of systems which are used to monitor and control remote equipment that can be found in general power plants, nuclear power plants, telecommunication; and any other critical infrastructure entities. SCADA is used for data acquisition, data processing, alarm and event processing, sequential switching operation, etc. SCADA systems were previously used in a closed private network; meaning that it is isolated from being exposed to external world. As the technology changes the needs came to connect the SCADA into the internet; which means increasing the chances of exposing the critical infrastructure into cyber-attacks. In this paper, we present a new approach to monitor Programmable Logical Controllers (PLC) that is focused on analysing packets from PLC to understand the normal behaviour of the equipment and reflect it into analyst monitoring dashboard. As a proof of concept, we will build small PLC temperature sensor, and we capture and record traffic from PLC during normal operation and also while PLC is under attack. The data analysed and reflected into analyst monitoring dashboard.

1 Introduction

SCADA system is a control center developed to establish a centralized monitoring, status, and control of remote field devices; which are geographically distributed over long distances networks covering regions and countries. In cyber security, there is no one solution that can provide full secured systems to an organization. Security can be achieved by addressing operational needs and designing a secure platform that can meet the operational needs. In our paper we want to provide higher secure integrated system by addressing the need to aggregate PLC, open

A. AlShemeili (✉) · C.Y. Yeun · J. Baek

Electrical and Computer Engineering Department, Khalifa University, P.O. Box 127788, Abu Dhabi, United Arab Emirates
e-mail: 100037742@kustar.ac.ae

source vulnerabilities, and addressed NIST 800-82 vulnerability platforms into the Splunk (2014 SIEM Lead) [1] Security Information and Event Management (SIEM). This will provide the operators with insight of current security/health status of the system and also it will provide current or possible attacks outside of the operator network; that can target his/her system soon. Stuxnet was an inspiration to consider building an advanced proof of concept security protection mechanism to prevent PLC from malicious cyber-attacks. We believe the combination of:

- learning and capturing normal PLC behavior
- developing alerts once operating outside normal conditions
- 24/7 open source vulnerability monitoring

will enhance, protect and also increase prevention of cyber-attacks at least by 95 %. Figure 1 shows our approach in this paper; and due to time limitation our focus will be in the high level concept. The POC will maintain the monitoring of system and data integrity, authentication, and access control, which will help in protecting the PLC.

Siemens S7-1200 Vulnerabilities

Siemens has released on 20-03-2014 six vulnerabilities related to SIMATIC S7-1200 PLC [2].

Vulnerability 1 (CVE-2014-2249)

Cross-Site Request Forgery (CSRF) using malicious link could affect the PLC through the usage of TCP ports 80 and 443 of the integrated web server. The impact will be on the integrity and availability of the PLC.

Vulnerability 2 (CVE-2014-2258)

PLC will be defected once special crafted packets are received by TCP port 443. Restarting the system is a must to recover.

Fig. 1 PLC security enhancement model



Note that our PLC were defected after receiving single packet ping with 1500 MTU size. Restarting the PLC was a must to get back to normal operation. Our target from this POC is to discover and report such attacks within SCADA environment.

Section 2 covers related work readings. Section 3 discusses importance of PLC security. Section 4 discusses PLC building. Section 5 covers the analysis part of our paper. Finally, Sect. 6 concludes our work.

2 Related Works

As of today there are many articles discussing PLC security; especially after Stuxnet incident. We went through different articles [3–6, 7–10] which are focusing at discovering PLC vulnerabilities, analysis of PLC protocols, building secure DMZ layer, communication, encryption and implementing standard security policies. All of the articles agreed on the fact of the importance of securing PLC or the network element by understanding its operational activities. We are going show our new approach in monitoring and protecting PLC in Sects. 4 and 5.

3 Importance of PLC Security

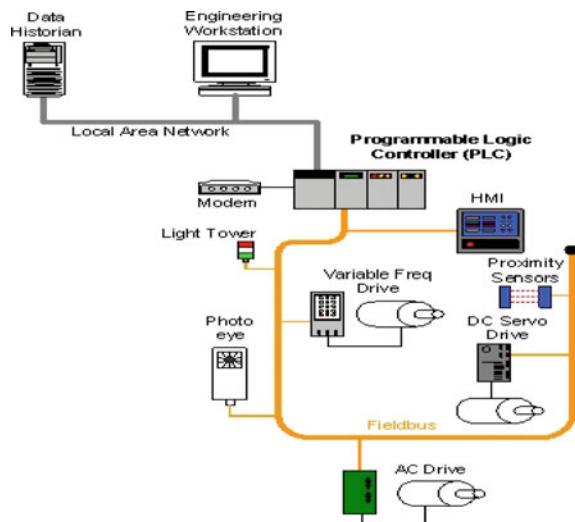
Compromising industrial critical PLC security will lead to catastrophic events, which not only harm the owner of the system, but it will also might impact environmental, political, and financial stand of a particular country. If the company running the nuclear plant in IRAN notices the sudden changes on their PLCs; they could probably stopped Stuxnet malicious attack.

Figure 2 [11] shows an example of PLC implementation used in manufacturing process. Utilizing the existed programmable memory inside the PLC allowed the operator to set specific controls such as I/O, timing, logic over the controlled and monitored field devices. For example it's possible now to collect alarms and data from field devices and also to send commands like opening and closing breakers and valves.

4 Building Siemens PLC for SCADA Framework

Building a secure monitoring and protection framework for PLC can be achieved by:

- Building proper physical infrastructure
- Building proper network infrastructure

Fig. 2 PLC implementation

- Maintain proper system and network configuration
- Maintain up-to-date SW release
- Building proper security layers
- Continuous monitoring

Figure 3 shows our built Siemens PLC S7-1200 temperature sensor.

**Fig. 3** Our built Siemens PLC S7-1200 temperature sensor

5 Analysis

Wireshark is used to capture the traffic from PLC. The captured data is uploaded into Splunk Security Information and Event Management (SIEM) that is used for aggregating and correlating incoming data. Also it is used to generate an automated alerts and displaying customized dashboards for operators to monitor. For big data analysis we can select to view many captured fields, such as, source IP, destination IP, protocols, and time stamp. SIEM solutions allow security engineers to automate fault/suspicious alerts generation. From search, you can type the keyword of interest (Fig. 4).

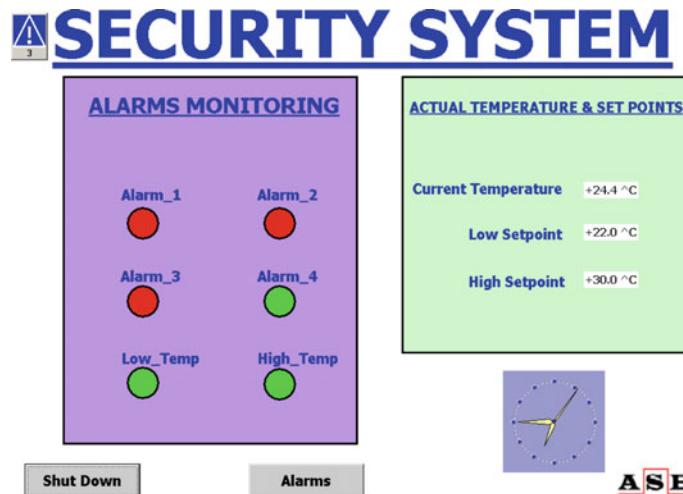


Fig. 4 PLC alarm panel display window

Security analyst can benefit from the feed of the captured data from PLC by knowing current sources, destinations, protocols. This information can be compared to the target hardened process. Figure 5 shows protocols captured from our PLC.

5.1 Open Source

It is of great importance for organizations to monitor threats and vulnerabilities which are related to their business. For example the recent announcement of sandworm CVE-2014-4114 [12] which was targeting Windows PCs (HMI). Spear phishing mail was used by the attacker, were by opening the attached file, a malware will be installed, enabling attacker to remotely access the system. Windows

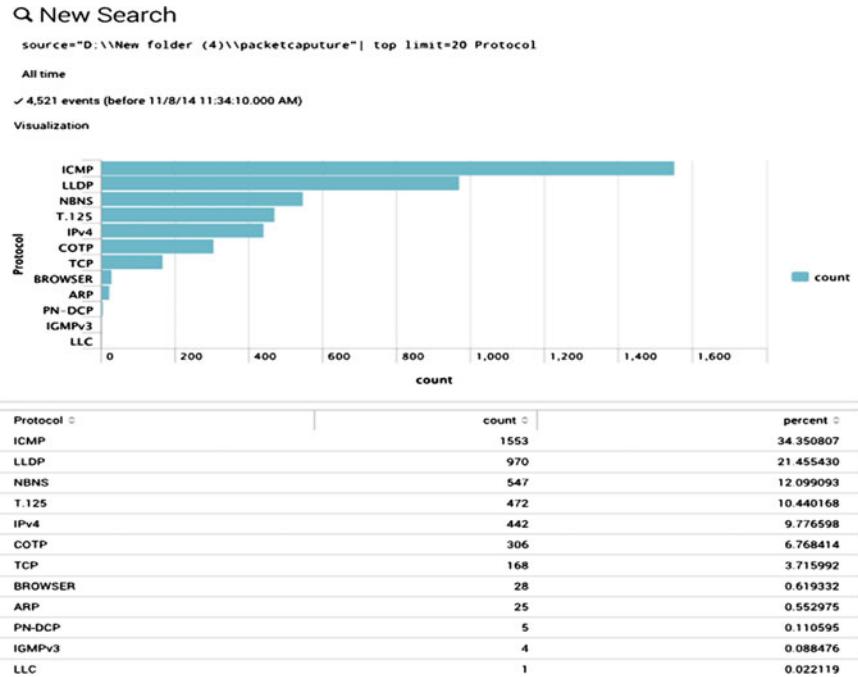


Fig. 5 Top protocols

released a patch to fix the vulnerability. Open source feeds can be obtained freely from US-CERT, or any other CERT after obtaining CERTs approval. And also it can be purchased from multiple companies like Anti-Virus companies. Another mechanism is to use Google alert and customize it to send alerts of interest to specific email address. Figure 6 show bad IP and users [13] feeds dashboard in Splunk. This feed can be used to compare it with existing source and destination IP and send alerts if there is any similarity discovered.

5.2 Vulnerability Test

Siemens released a number of vulnerabilities which are related to S7-1200 (similar to the one used in our paper) [14]. Figure 7 shows that Denial of Service and Overflow are rated as the highest PLC vulnerabilities for the year 2012/2013.

Now we will try to evaluate the security of the PLC. Our PLC does not support Password and hence were not going to capture the password on clear text. Also we don't have a complete SCADA architecture to conduct a full PLC analysis within an operating SCADA environment.

The following will be used to assess the security of Siemens PLC:

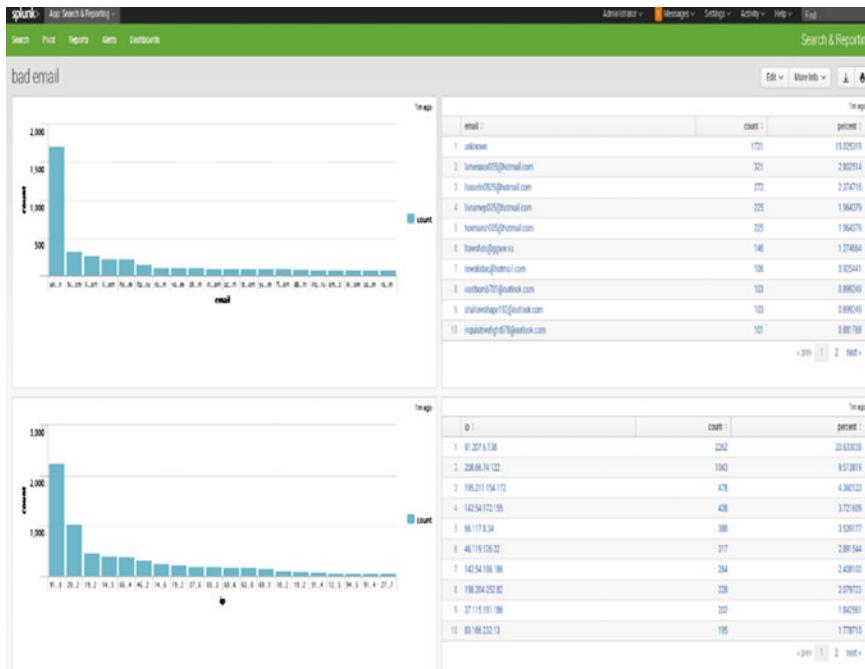


Fig. 6 Bad IP address data feed

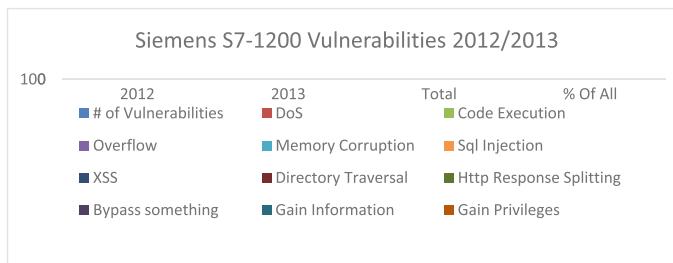


Fig. 7 SiemensS7-1200 vulnerabilities 2012/2013

- Wireshark
- Metasploit [15]
 - Open source penetration testing and security assessment tool

Figure 8 shows details of the discovered services

- Host IP 192.168.0.1 (PLC IP address)
- Open ports TCP 102 and UDP 161 (ISO-TSAP and SNMP)
- PLC HW details (Siemens SIMATIC S7, CPU 1200)

Fig. 8 Metasploit details of discovered services



Fig. 9 Captured protocols during Metasploit attack

Figure 9 shows captured protocols during Metasploit attack. Note: our PLC stopped responding after sending ping command to 192.168.0.1 and we had to restart the PLC to resume its functionality.

6 Conclusion

In security, it is very important to learn from previous mistakes/incidents. This is the reason why Stuxnet motivated many security specialists to study the field and come up with new ideas for protection. In this paper the motive was the same, and the objective was to provide the guideline along with small POC experiment on how to secure PLC from cyber-attacks using behaviour analysis. We have started

introducing SCADA systems and Stuxnet malware which was the motivation to work in this paper. Then we highlighted security concerns which are related to PLC. In our paper we managed to build a small Proof of Concept (PoC) Siemens PLC. We used SIEM solution (Splunk), which is used for the first time in SCADA network and uploaded captured PLC traffic from Wireshark. Also we used an open source vulnerability feed and integrate with the SIEM to provide unique dashboard for the SOC operator.

References

1. Splunk Named a Leader in the 2014 Gartner Magic Quadrant for Security Information and Event Management. (Online). Visited on 6 Dec 2014
2. Siemens Security Advisory by Siemens ProductCERT. SSA-654382: Vulnerabilities in SIMATIC S7-1200 CPU (Online). http://www.siemens.com/innovation/pool/de/forschungsfelder/siemens_security_advisory_ssa-654382.pdf. Visited on 6 Dec 2014
3. Milinkovic, S.A., Lazic, L.R.: Industrial PLC security issues. In: The 20th Telecommunications Forum (TELFOR), Serbia, Belgrade, pp. 1536–1539, 20–21 Nov 2012
4. Sandaruwan, G.P.H., Ranaweera, P.S., Oleshchuk, V.A.: PLC security and critical infrastructure protection. In: The 8th IEEE International Conference on Industrial and Information Systems (ICIIS), pp. 81–85, 18–20 Aug 2013
5. Queiroz, C., Mahmood, A., Hu, J., Tari, Z., Yu, X.: Building a SCADA security Testbed. In: Third International Conference on Network and System Security, pp. 357–364 (2009)
6. Pigglin, R.S.H.: Emerging good practice for cyber security of Industrial Control Systems and SCADA. In: The 7th IET International Conference on System Safety, incorporating the Cyber Security Conference (2012)
7. Baek, J., Vu, Q.H., Jones, A., Al Mulla, S., Yeun, C.Y.: Smart-frame: a flexible, scalable, and secure information management framework for smart grids. In: Proceeding of the International Conference for Internet Technology and Secured Transactions (ICITST'12), pp. 668–673, 10–12 Dec 2012, London, UK
8. Shemaili, M.A., Yeun, C.Y., Mubarak, K., Zemerly, M.J.: A new lightweight hybrid cryptographic algorithm for the internet of things. In: Proceeding of the International Conference for Internet Technology and Secured Transactions (ICITST'12), pp. 87–92, 10–12 Dec 2012, London, UK
9. Gajparia, A.S., Mitchell, C.J., Yeun, C.Y.: Supporting user privacy in location based services. In: The Special Issue on Mobile Multimedia Communications on IEICE Transactions on Communications, E88-B, pp. 2837–2847, July 2005
10. Konidala, D.M., Yeun, C.Y., Kim, K.J.: A secure and privacy enhanced protocol for location-based services in ubiquitous society. In: Proceeding of the IEEE Global Telecommunications Conference 2004, vol. 4, pp. 2164–2168, Dallas, Texas, USA, 29 Nov–3 Dec 2004
11. Guide to Supervisory Control and Data Acquisition (SCADA) and Industrial Control Systems Security
12. An Analysis of Windows Zero-day Vulnerability ‘CVE-2014-4114’ aka “Sandworm”
13. Abuse Reporting and Blacklisting (Online). <https://www.openbl.org/lists.html/>. Visited on 6 Dec 2014
14. Siemens Simatic S7-1200 PLC Vulnerability Statistics (Online). Visited on 6 Dec 2014
15. Metasploit (Online). <http://www.metasploit.com/>. Visited on 6 Dec 2014

Modeling of Smart Supply Chain for Sustainability

Kyeongrim Ahn, Sangwon Lim and Younggyo Lee

Abstract The advanced supply chain will transform traditional system into intelligent system to enhance work efficiency and productivity using advanced technology and modeling. Interoperability, visibility, and sustainability are critical objectives in smart supply chain. Especially, energy efficiency is a critical performance objective in sustainable supply chain. In order to address these objectives, the related data should be collected and quickly analyzed to provide a deeper and more complete understanding of the current operational status. This paper identifies requirements for smart supply chain, develops business process and information models using international standard modeling methodology, for the purpose of achieving the objectives of supply chain. These models can form a basis for new standards that enable parties in the supply chain to achieve objectives of smart supply chain. The work presented in this paper would help achieving critical objectives, interoperability, visibility, and sustainability, that will facilitate energy tracking and reporting across the supply chain between business entities of supply chain. Therefore, this enable parties in the supply chain to improve their work efficiency as referencing these models.

Keywords Smart logistics · Business process modeling · Information modeling · Supply chain · XML

K. Ahn (✉)

ADB Consultant, Korea, Seoul, South Korea
e-mail: ahn.kyeongrim@gmail.com

S. Lim

Logistics Management Programme, Chulalongkorn University, Seoul, South Korea
e-mail: sangwonlim@gmail.com

Y. Lee

Department of Internet Information, Seoil University, Seoul, South Korea
e-mail: younggyo@seoil.ac.kr

1 Introduction

Supply chain of today operates in a complex, highly competitive, environment affected by constantly changing market and technological forces. There are several issues to note, including various participating business entities, mass information, and complex network structure [1]. Supply chain users need to determine how new process and information technologies can help them in conducting activities better. They also need to consider how to effectively manage business process in their transactions, while minimizing usage of their resources.

Smart supply chain will transform traditional process and system into intelligent system using international standards and the advanced technology. Interoperability, visibility, and sustainability are critical objectives in smart supply chain. Under the circumstances that share of trade and logistics in a national economy is over 70 %, realizing an efficient supply chain management by achieving these objectives of a smart supply chain would help better control and monitor flow of physical consignments and information in a supply chain.

This paper defines information flows of a supply chain by modeling (1) the underlying logical network structure of the supply chain and (2) the business processes that are associated with the logistics. We model both structure and processes using the UN/CEFACT Modeling Methodology (UMM) specified by the United Nations Centre for Trade Facilitation and Electronic Business (UN/CEFACT) [2]. UMM is specifically designed to capture business process using commonly used business terms. From the resulting process models, we extract required flows and associated data elements, from which we then create information models. These information models, to be developed as XML (eXtensible Markup Language) schemas, are the principal research outputs in the paper. Instances of these models will be exchanged among the various nodes in a network.

These models can form a basis for new standards that enable parties in a supply chain to achieve objectives of a smart supply chain. The work presented in this paper would help achieving critical objectives of interoperability, visibility, and sustainability among business entities of supply chain. Furthermore, it would be able to support seamless, sustainable, and green supply chain.

The paper is organized as follows: Sect. 2 describes the relevant study used in the research; Sect. 3 uses the UMM to develop the necessary business process models and the derived information models; and, Sect. 4 concludes with a summary and describes opportunities for future work.

2 The Related Study

2.1 *The Needs of Modeling for Smart Supply Chain*

Since supply chains are now global and the work efficiency in supply chain is important, it is necessary to provide new traceable measurements and aggregation

methods for assessing energy efficiency of logistics to improve the efficiency (interoperability, visibility, and sustainability). As with all other aspects of the product realization process, Information and Communication Technology (ICT) will provide both (1) the foundation for making those measurements and their underlying logistics information and (2) the means for communicating them among global supply chain partners [1].

To date, little attention has been paid to the impacts of logistics on the overall energy efficiency of the supply chain. Therefore, in this paper, we focus on the information and material flows associated with the logistics into and out of supply chain. Inadequate data collection by, and the complex structural relationships among, supply chains make it difficult to track these flows. Capturing data related to these flows is also difficult because the necessary standards and information models do not exist. The aim of this paper is to define the information and material flows by modeling (1) the underlying physical and logical network structures of the supply chain and (2) the business processes that are associated with the logistics [2, 3].

2.2 *The Trend of International Standardization*

In order to accelerate realization of a smart supply chain in trade and logistics, international standard organizations and some countries have started to develop collaboration projects and standards. First, NEAL-NET (Northeast Asia Logistics Information Service Network) has been established to share logistics information among China, Japan, and Korea [4]. The goal of NEAL-NET is to support eco-friendly, reliable, and efficient logistics among those three countries, and to furthermore expand into European countries or United States. The three countries developed the specifications of exchanged information and interface used for data exchange. The exchanged information comes from each country's information systems such as China's LoginK, Japan's Collins, and Korea's Port-MIS. Another collaboration project, European Union (EU) e-Freight project, aims to achieve paperless business by applying information and communication technology [5]; The goal is to define common standard framework for interoperability, and develop singular e-document for all transport activities.

International standard organizations, in particular International Organization for Standardization (ISO) and UN/CEFACT, have started discussion on how to support interoperability and visibility. These organizations have launched standard projects to develop recommendations or specifications regarding common framework, interface, or standard messages. ISO Technical Committee (TC) 204 has developing framework and business process/information model for interoperability in supply chain [6]. UN/CEFACT and World Custom Organization (WCO) have been developing interoperability model based on single window standard.

3 Modeling for a Smart Supply Chain

3.1 *The Requirement of a Smart Supply Chain*

Future Logistics would cater green IT for a transition to a low-carbon society, taking into account environment issues. For eco-friendly industry, it should combine manufacturing, industrial complex management, logistics activities with green IT, and accelerate building intelligent transport logistics systems. Supply chain of today has many issues mentioned in Sect. 1. Instead of focusing on individual issues, it must find ways on how to support a smart supply chain. These issues not only affect supply chain users' ability to support and optimize their existing operations cost effectively, but also support the deployment and integration of new technologies with their current infrastructure. If the needs for creating a smart supply chain are to be met, better support for defining models for global supply chain must be provided. Additionally, the machines, devices, facilities, and information system for a smart supply chain must be developed. The objectives of supply chain should be able to improve safety and efficiency of supply chain operations while maintaining quality of their service.

3.2 *Business Process Modeling*

To address the objectives of smart supply chain, information should be collected and quickly analyzed to provide a deeper and more complete understanding of the current operational status. We must first analyze business requirements and model the associated business process (see Fig. 1) using the UN/CEFACT Modeling Methodology, Users Guide (UMM). Business Requirements Specification (BRS) and Requirements Specification Mapping (RSM), which are parts of UMM, provide documentation templates and conformity rules to guide the development of e-business requirements [7]. To build accurate models of the business processes, one must have a complete understanding of the information, energy usage, and material flows in a supply chain. Figure 1, which is a sequence diagram of the goods-delivery process, shows some of these flows. This sequence diagram of Fig. 1 involves a number of business processes and executed by different business partners and a number of business transactions.

Each Business transaction is realized by an exchange of Business documents (also called messages) among supply chain applications. Electronic exchanges are hampered by poor interconnectivity and multiple standards. Most important reason for conducting this research is that information flow problems in a supply chain still exist. To collect necessary data, extra processes or devices are needed to collect and send that data.

“Pickup Consignments” business process links the Consignor (Manufacturer, Supplier, etc.), and Carrier that are the participating entity at in this process.

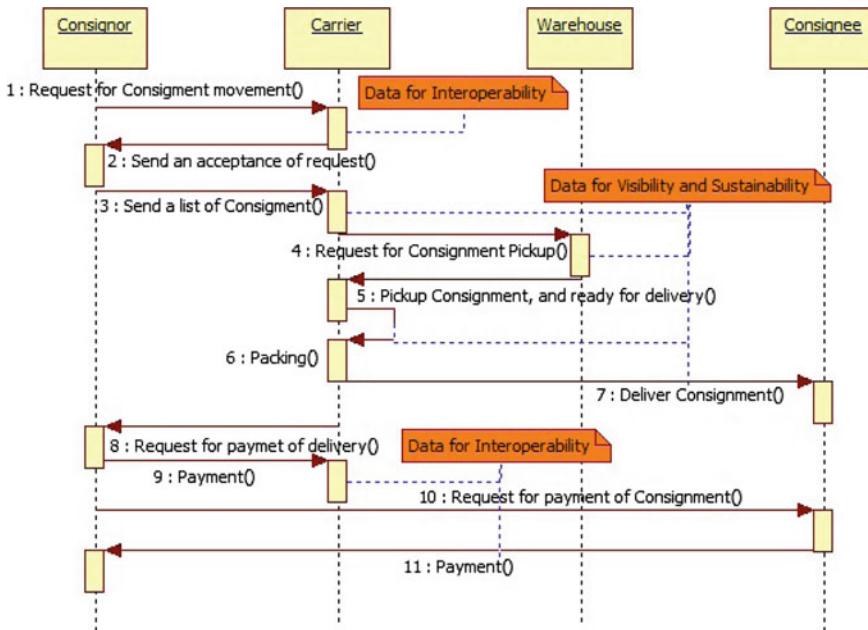


Fig. 1 Sequence diagram of a supply chain

Figure 2 shows an activity diagram of “**Pickup Consignments**” business process. This diagram shows the sequence of activities between the participating entities, which include Consignor (Manufacturer, Supplier, etc.), and Carrier. This business process starts when Manufacturer or Supplier send “Delivery Request” to Carrier. Carrier then picks up goods or parts from their current location and delivers them to their assigned destination. Energy data in logistics is related to picking or delivering activities.

3.3 Information Modeling

This section proposes information model to support sustainability among the objectives of a supply chain. This information model can facilitate the collection and representation of the required energy data. In developing model, we will follow the UMM [2, 8]. Assessing the level of energy efficiency requires the collection of energy data associated with every work process. This data provides the necessary inputs to computation that measures the actual energy consumption. First it puts Root element of XML assembly into “**EnergyData. Details**”. This XML assembly contains three sub elements; **Document. Details**, **Consignment. Details**, and **Product. Details**. These sub elements are defined according to UN/CEFACT Core

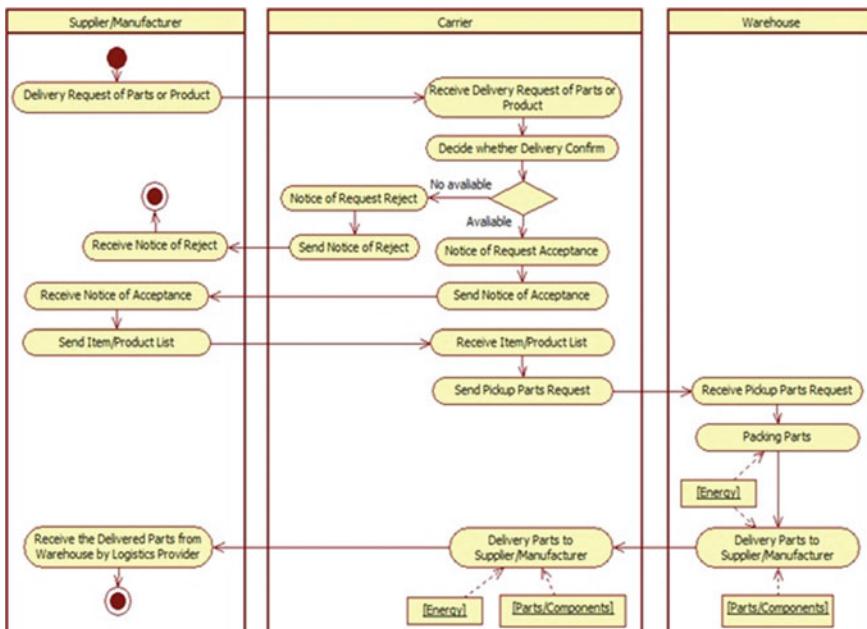


Fig. 2 Activity diagram of “Pickup Consignments” business process

Component Library (CCL) as ACC. Table 1 shows the detailed data elements of sub elements. **Energy_ Document. Details** ABIE describes the related management data associated with the energy data collecting process. It contains creation date and time, originator, and recipient, etc. **Of the Related Energy_Consignment. Details** ABIE describes goods or parts with embodied energy from manufacturing processes, logistics process, or storage processes. **Energy In_Product. Details** ABIE describes the measured energy, such as the measured amount, energy type, or process or party that generates energy. This ABIE contains BBIEs and ASBIEs as child elements. These ASBIEs, which comprises numerous sub elements of Aggregated BIE (ABIE), uses Aggregated Core Component (ACC) as a qualifier. In this paper, we develop ASBIE to add a qualifier to ACC as shown in Fig. 3.

According to the UN/CEFACT modularity model, XML schemas categorize modules into business information payloads and external schema modules. The business information payload consists of a root schema and internal schema modules that reside in the same namespace. Figure 3 shows a class diagram of the business information payload and Fig. 4 shows a root schema. The external schema modules consist of a set of reusable schema for ABIEs, unqualified data types, qualified data types, code lists and identifier lists. The reusable ABIE schema

Table 1 The mapped data elements of “EnergyData. Details”

ABIE Dictionary entry name	ACC Dictionary entry name	Included (BBIE)
Energy Data_ Document. Details	Document. Details	<ul style="list-style-type: none"> • Identification. Identifier • Description. Text • Issue. Date Time • Submission. Date Time • Sender. Service_ Party • Receiver. Service_ Party
Of the Related Energy_ Consignment. Details	Consignment. Details	<ul style="list-style-type: none"> • Identification. Identifier • Consignment Item. Quantity • Summary Description. Text • Name_ Information. Text • TypeSize_ Information. • Text • Status_Information. Text • Availability Due.Date Time • Include. SubConsignmentItem • Associated. Supplier_ Party
Energy In_ Product. Details	Product. Details	<ul style="list-style-type: none"> • Identification. Identifier • Name. Text • Energy_Type. Code • Energy_GrossVolume. Measure • Measurement. Code • Applicable. Characteristic • Energy_ Manufacturer. Party

module always imports the unqualified data type and qualified data type schema modules. The unqualified data type schema imports necessary code list schema modules and may import identifier list schema modules. The core component type schema module is provided as reference documentation and is used as the basis for the unqualified data type.

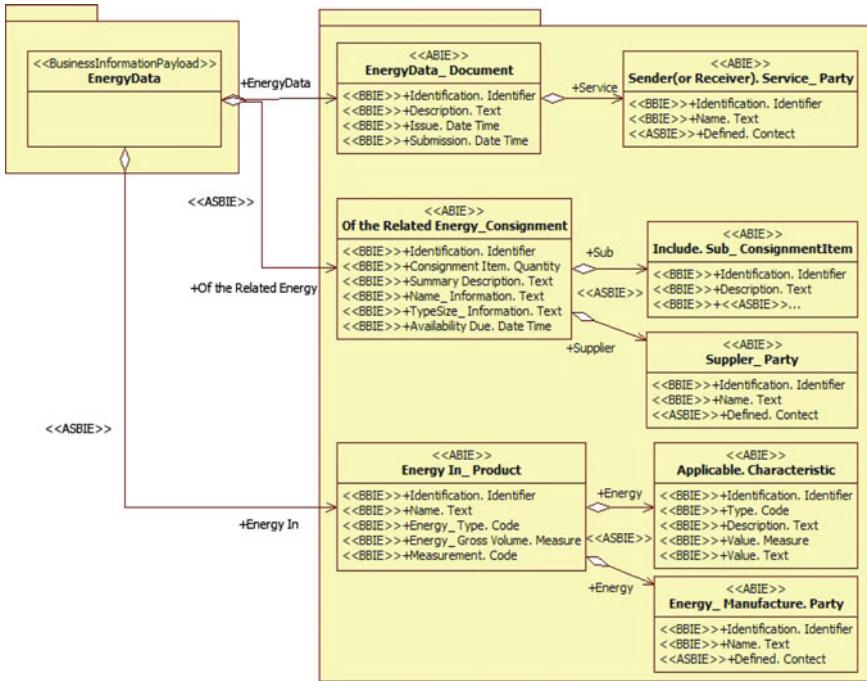


Fig. 3 Class diagram of the business information model

```

<?xml version="1.0" encoding="UTF-8"?><xs:schemaxmlns:xs="http://www.w3.org/2001/XMLSchema
a" elementFormDefault="qualified" attributeFormDefault="unqualified">
<xs:include schemaLocation="ReusableAggregateBusinessInformationEntitySchemaModule_1p0.xsd"/>
<xs:element name="EnergyData" type="EnergyDataType">
    <xs:annotation>
        <xs:documentation>Energy Data that is collecting from manufacturer, supplier, carrier, warehouse.</xs:documentation>
    </xs:annotation>
</xs:element>
<xs:complexType name="EnergyDataType">
    <xs:sequence>
        <xs:element ref="EnergyDataDocument"/>
        <xs:element ref="Of the Related EnergyConsignment" maxOccurs="unbounded"/>
        <xs:element ref="EnergyInProduct" maxOccurs="unbounded"/>
    </xs:sequence>
</xs:complexType>
</xs:schema>
  
```

Fig. 4 Root schema of XML assemble “EnergyData. Details”

3.4 How to Apply the Output

If the electronic documents contain energy data developed by Sect. 3.2, it is possible to collect energy data occurred at the activities of supply chain. With these

energy data, it is possible to define an energy-saving method at the supply chain processes. Additionally, if the status information of cargos or transportation in supply chain is collecting after transaction, user can grasp correct status of cargos and take follow-up action if some problems are occurred.

4 Conclusion

Supply chain operates in a complex, highly competitive, environment affected by constantly changing market and technological forces. There are several issues to note, including of various participating business entities, mass information, and complex network structure. Under the circumstances that share of trade and logistics in a national economy is over 70 %, realizing an efficient supply chain management by achieving these objectives of a smart supply chain would help better control and monitor flow of physical consignments and information in a supply chain.

This paper did so by modeling the business processes and information of a supply chain. These models can form a basis for new standards that enable parties in the supply chain to achieve objectives of a smart supply chain. The work presented in this paper would help achieving critical objectives of interoperability, visibility, and sustainability among business entities of a supply chain. Furthermore, it would be able to support seamless, sustainable, and green supply chain. It can also be a first step toward achieving the objectives between the participating business entities of the supply chain. Future work could include extending business process and information model based on the output of this paper, and applying it to real-world problems that need a smart-supply chain-based solution.

References

1. Jin, Yun-Jun, Lee, Yu-Bin, Bae, Ki-Hyung: A Study on the SCM integration & green growth strategy of logistic company in Korea. *Int Commer. Inf. Rev.* **15**, 3–23 (2013)
2. Ahn, K., Lim, S., Lee, Y.: modeling for smart supply chain: interoperability, visibility, and sustainability, In: The 2015 World Congress on Information Technology Applications and Services (World IT Congress 2015) (2015)
3. Feng, S., Kumaraguru, S., Law, K., Ahn, K.: Modeling energy information for integrating product assembly processes and supply chains, part I, International Federation for Information Processing (IFIP). In: *Advances in Information and Communication Technology (AICT) 414 Proceedings of the 2013 APMS 2013 International Conference Advances in Production Management Systems*, pp. 310–317 (2013)
4. United Nations Centre for Trade Facilitation and Electronic Business, The Cargo Tracing and Tracking Project (P108, Project Leader Kyeogrim Ahn) (2012)
5. The Ministry of Oceans and Fisheries, The 10th NEAL-NET Experts Meeting Report (2013)
6. Vennesland, A.: e-Freight: common framework and standardization. In: The 2nd Asia-EU Co-operation Technical Workshop (2013)

7. ISO/TC204/WG7, ISO/PDTS 17187.: Intelligent transport systems—electronic information exchange to facilitate the movement of freight and its intermodal transfer—governance rules to sustain electronic information exchange methods (2013)
8. United Nations Centre for Trade Facilitation and Electronic Business, UN/CEFACT Modeling Methodology, Users Guide (2003)

Human Action Classification Using Multidimensional Functional Data Analysis Method

Wanhyun Cho, Sangkyoon Kim and Soonyoung Park

Abstract In this paper, we describe a novel approach that can classify a human action by using a multidimensional functional data analysis (MFDA) and the Cartesian product of reproducing kernel Hilbert spaces (CPRKHSs). The main idea is to represent the human action video dataset into a multidimensional functional data framework, and then apply the mathematical properties of CPRKPHS to classify these datasets. First, we extract the feature vector that can properly describe the shape of the human action from each frame of a given video. Here, a set of features extracted from a given video can be expressed as a multivariate functional data format depending on an order of time. Since a multidimensional functional data belongs to the non-linear manifold, we embed a multidimensional functional data into the CPRKPHS by using the idea of kernel methods. Then, we have shown that the geodesic distance between two human actions on manifold can be approximate with the product Hilbert norm for a difference between two multidimensional functional datasets in RKPHS. Finally, we have applied common classification rules such as the k-NN method based on these distances in order to classify a human action.

Keywords Human action classification · Multi-dimensional functional data · Cartesian product of reproducing kernel Hilbert space · Generalization of Tikhonov regularization · Distance in CPRKHS · k-NN classification method

W. Cho (✉) · S. Kim · S. Park

Department of Statistics, Chonnam National University, Gwangju, South Korea
e-mail: whcho@chonnam.ac.kr

S. Kim
e-mail: narciss76@mokpo.ac.kr

S. Park
e-mail: sypark@mokpo.ac.kr

W. Cho · S. Kim · S. Park
School of Electronic Engineering, Mokpo National University, Mokpo, South Korea

1 Introduction

In this paper, we propose a novel approach that can execute the classification of human actions by using a theory of functional data analysis and mathematical properties of reproducing kernel Hilbert space. The main idea is to convert the human action video dataset into a functional data framework, and then apply the Cartesian product of reproducing kernel Hilbert space (CPRKHS) approach to analysis these functional datasets. First, we extract the feature vector describing the content of the human action from each frame of the given video. And then we express a set of features extracted from a given video as a multidimensional functional data structure according to an order of the time. Second, since a functional data could be thought as the non-linear manifold, we embed this data into a high dimensional RKHS by using the idea of kernel methods. And then we have shown that the geodesic distance between two feature vectors extracted from two human action videos in the non-linear manifold can be approximate with the Hilbert norm between two multidimensional functional data embedded in CPRKHS. Therefore, we can apply the k-NN method based on Hilbert norm in order to classify a human action.

2 Classification of Human Action Video

2.1 *Representing Human Action Video as Multidimensional Functional Data*

We have represented a video \mathbf{V} of duration T frames as a time series data $\mathbf{V} = (\mathbf{f}(t))_{t=t_1, \dots, t_T}$ where $\mathbf{f}(t) = (f^1(t), \dots, f^p(t))_{t=t_1, \dots, t_T}^T$ denotes a p -dimensional feature vector extracted from each frame in image sequence. In this case a path of \mathbf{V} is represented by a set of p smooth curves. Here, we have called this dataset as a multidimensional functional data. But, a type of this data is actually given by the following discrete form:

$$\mathbf{V} = (f^1(t), \dots, f^p(t))_{t=t_1, \dots, t_T}^T = \begin{pmatrix} f^1(t_1) & \dots & f^1(t_T) \\ f^2(t_1) & \dots & f^2(t_T) \\ \vdots & \ddots & \vdots \\ f^p(t_1) & \dots & f^p(t_T) \end{pmatrix} = \begin{pmatrix} y_1^1 & \dots & y_T^1 \\ y_1^2 & \dots & y_T^2 \\ \vdots & \ddots & \vdots \\ y_1^p & \dots & y_T^p \end{pmatrix}$$

where $y_j^i, i = 1, \dots, p, j = 1, \dots, T$ represents the actual observations for j th time of i th function.

Here, the first task in a multidimensional functional data analysis (MFDA) is to transform each datum $\mathbf{y}^i = (y_1^i, \dots, y_T^i), i = 1, \dots, p$ into a p -dimensional smooth function $\mathbf{f}^i = (f^1(t_1), \dots, f^i(t_T))$. Of course, the set of data points which can be observed is finite while an accurate description of the underlying function would

require an infinite number of observations. Therefore the choice of a particular type of function f^i will be done in general by selecting it from an infinite collection of alternative models. This is the typical context in which ill-posed problems arise.

2.2 *Multidimensional Version of Regularization Method*

Here, we will consider the generalization of regularization issue suitable for our multidimensional functional data. Then, the generalization formula of Tikhonov regularization method can be represented by:

$$E(\mathbf{f}) = \frac{1}{T} \sum_{j=1}^T \|\mathbf{f}(t_j) - \mathbf{y}_j\|^2 + \lambda \|\mathbf{f}\|_{\mathbf{H}}^2.$$

Here, the concept of kernels will basically provides us with a flexible, computationally feasible method for implementing this scheme. In this case, a particularly useful family of hypothesis spaces is the Cartesian product of reproducing kernel Hilbert space (CPRKHS), each component of which is associated with a particular kernel. Then, the extended general solution of Tikhonov regularization in CPRKHS, thought as the generalization of representer theorem, can be given as a compact representation as described in the following theorem.

Multidimensional Version of Representer Theorem: The minimizer $\hat{\mathbf{f}}_{K,\lambda,n}$ over the Cartesian product space $\mathbf{H} = \mathbf{H}_1 \times \dots \times \mathbf{H}_p$ of RKHSs $\mathbf{H}_k, k = 1, \dots, p$ for the regularized empirical functional

$$E(\mathbf{f}) = \frac{1}{T} \sum_{j=1}^T \|\mathbf{f}(t_j) - \mathbf{y}_j\|^2 + \lambda \|\mathbf{f}\|_{\mathbf{H}}^2$$

can be represented by the expression

$$\hat{\mathbf{f}}_{K,\lambda,n} = (\hat{\mathbf{f}}_{K,\lambda,n}^1, \dots, \hat{\mathbf{f}}_{K,\lambda,n}^p)^T, \hat{\mathbf{f}}_{K,\lambda,n}^k = \sum_{j=1}^T \alpha_j^k K(t_j, t), k = 1, \dots, p,$$

where the t_j points are the sample time data and the coefficients $\alpha_j^k \in \mathbb{R}$ are the solutions to the linear system:

$$(\lambda T \mathbf{I} + \mathbf{K}) \boldsymbol{\alpha} = \mathbf{y}, \quad k = 1, \dots, p,$$

where \mathbf{I}_T is the identity matrix of $(T \times T)$ dimension, $\mathbf{K}|_t$ is the matrix of kernel functions, $\boldsymbol{\alpha}^k$ is the coefficient vector, and \mathbf{y}^k is observation vector.

2.3 Distance Between Two Human Action Videos

Here, we consider two multidimensional functional datasets $\mathbf{f}_{T_1}(t)$ and $\mathbf{g}_{T_2}(t)$ given by time series of two videos V and W corresponding to two human actions [7]:

$$\mathbf{f}_{T_1}(t) = (f^1(t), \dots, f^p(t))^T, t = (t_1, \dots, t_{T_1}), \mathbf{g}_{T_2}(t) = (g^1(t), \dots, g^p(t)), \\ t = (t_1, \dots, t_{T_2}).$$

By applying the multidimensional version of Tikhonov regularization method with these functional datasets, we can have two minimizer functions $\hat{\mathbf{f}}_{K,\lambda,n}$ and $\hat{\mathbf{g}}_{K,\lambda,n}$ respectively given as the following forms:

$$\hat{\mathbf{f}}_{K,\lambda,n} = (\hat{f}_{K,\lambda,n}^1, \dots, \hat{f}_{K,\lambda,n}^p)^T, \hat{f}_{K,\lambda,n}^k = \sum_{j=1}^{T_1} \alpha_j^k K(t_j, t), k = 1, \dots, p,$$

and

$$\hat{\mathbf{g}}_{K,\lambda,n} = (\hat{g}_{K,\lambda,n}^1, \dots, \hat{g}_{K,\lambda,n}^p)^T, \hat{g}_{K,\lambda,n}^k = \sum_{j=1}^{T_2} \beta_j^k K(t_j, t), k = 1, \dots, p,$$

where the coefficient α_j^k and β_j^k are respectively the components of the coefficient vectors $\boldsymbol{\alpha}^k$ and $\boldsymbol{\beta}^k$ given as

$$\boldsymbol{\alpha}^k = (\lambda T \mathbf{I}_T + \mathbf{K}|_{\mathbf{t}=(t_1, \dots, t_T)})^{-1} \mathbf{y}^k, \boldsymbol{\beta}^k = (\lambda T \mathbf{I}_T + \mathbf{K}|_{\mathbf{t}=(t_1, \dots, t_T)})^{-1} \mathbf{z}^k.$$

Then, we can consider the distance of two multidimensional functional datasets $\mathbf{f}_{T_1}(t)$ and $\mathbf{g}_{T_2}(t)$ as the Hilbert norm of difference between two minimizer functions $\hat{\mathbf{f}}_{K,\lambda,n}$ and $\hat{\mathbf{g}}_{K,\lambda,n}$ belonging to Cartesian product of reproducing kernel Hilbert space (CPRKHS). Since \mathbf{H} is the CPRKHS induced by the reproducing kernel K , the square of this norm is given by inner product defined at CPRKHS. That is, it can be given by

$$d(\mathbf{f}_{T_1}(t) | \mathbf{g}_{T_2}(t)) = \|\hat{\mathbf{f}}_{K,\lambda,n} - \hat{\mathbf{g}}_{K,\lambda,n}\|_{\mathbf{H}}^2 \\ = \sum_{k=1}^p ((\boldsymbol{\alpha}^k)^T K(\mathbf{t}_{T_1}, \mathbf{t}_{T_1}) \boldsymbol{\alpha}^k + (\boldsymbol{\beta}^k)^T K(\mathbf{t}_{T_2}, \mathbf{t}_{T_2}) \boldsymbol{\beta}^k - 2(\boldsymbol{\alpha}^k)^T K(\mathbf{t}_{T_1}, \mathbf{t}_{T_2}) \boldsymbol{\beta}^k)_{\mathbf{H}}.$$

2.4 Classification Rule for Human Action Video

Here, we use the classification rule as the k-nearest neighbor (k-NN) procedure that is one of the most popular methods to perform unsupervised classification in multivariate settings. The k-NN method starts by computing the Hilbert distance between the new input data $\mathbf{f}_{new}(t)$ and the any functional datum $\mathbf{f}_{data}(t)$ in the observed data set. Next, the method finds the k functional observations in the sample closest in distance to new functional data $\mathbf{f}_{new}(t)$. Finally, the new observation $\mathbf{f}_{new}(t)$ is classified by using majority of votes among the k neighbors.

3 Experimental Results

There are existed several types of data sets that can be used to recognize the human behavior, and KTH data set and Weizman data set among them are the most typically used. Here, we used the KTH data set for our experiments. The KTH human action dataset consisted with totally 600 video sets which 25 peoples performed 6 action classes at 4 different scenarios. Human actions are namely walking, jogging, running, boxing, hand-waving, and hand-clapping. All videos were shot in black and white images with the 160 * 120 resolution. In this paper, we use the training data consisting of ten peoples performing six actions. It is a very difficult task to define the feature vector to exactly classify human behavior. In this paper, we used the HOG (histogram Oriented of Gradients) features that are mostly used at pedestrian detection in order to express the behavioral changes of each video sequence into a feature vector. In order to calculate the HOG feature vector, we divide each frame into 3 * 5 blocks, and used eight histogram bins from each block. Therefore, the total dimension of the feature vector for each frame is 120 dimension.

In order to evaluate the performance of the proposed method, we use a validation data as 150 video sequences consisting of 25 peoples performing 6 different human behavioral. Table 1 shows a classification rate of six human behavior with respect to the proposed method in a matrix form. From the results in Table 1, we note that

Table 1 Classification rate of human actions for proposed method

Classification rate	Boxing	Hand clapping	Hand waving	Jogging	Running	Walking
Boxing	1	0	0	0	0	0
Hand-clapping	0.12	0.88	0	0	0	0
Hand-waving	0.12	0.08	0.8	0	0	0
Jogging	0	0	0	0.8	0.2	0
Running	0	0	0	0.08	0.92	0
Walking	0	0	0	0.12	0.16	0.72

six human behavior can be mainly divided into two categories with similar behavior. The first category of similar actions includes boxing, hand-clapping, and hand-waving, and the second category of similar behavior includes jogging, running, and walking. In this case, we can see that a misclassification happen a lot at the actions belonging to the same category. Consequently, we note that the correct classification rate of the proposed method totally appears to 85.33 % on average.

4 Conclusion

The main idea is to convert the human action video dataset into a functional data framework, and then apply the Cartesian product of reproducing kernel Hilbert space (CPRKHS) approach to analysis these functional datasets. First, we extract the HOG feature vector describing the content of the human action from each frame of the given video. And then we express a set of features extracted from a given video as a functional data form according to the order of the time. Second, since a functional data could be thought the non-linear manifold, we embed this data into a high dimensional RKHS by using the idea of kernel methods. And then we have shown that the geodesic distance between two functional data in the non-linear manifold can be approximate with the Hilbert norm between two functional data embedded in CPRKHS. Therefore, we can apply the k-NN method based on Hilbert norm in order to classify a human action.

Experimental results show that our method performs generally very well on the KTH public video dataset. The proposed approach provides a mathematical computation method of natural metric for video datasets, and it is easy to combine the more advanced classifiers. Our future work will extend the proposed method to other video recognition problems such as 3D human action recognition, gesture recognition, and surveillances system.

Acknowledgments This work was jointly supported by the National Research Foundation of Korea Government (2014R1A1A4A0109398) and (2012R1A1A4A01009097).

Light-Weight Authentication Scheme for NFC mCoupon Service in IoT Environments

Sung-Wook Park and Im-Yeong Lee

Abstract Recently, the potential of m-commerce by the combination of mobile and NFC aroused great momentum on e-commerce market. In particular, the variety of mCoupon services based on NFC has been utilized in consumer marketing instruments. However, the status of security technology for NFC based coupon service is very low. Related providers were considering method for just-in NFC service marketing case. And limited resource of NFC tag is difficult to deal with occurring security threats in NFC mCoupon service. In this paper, we proposed a light-weight authentication scheme for practical NFC mCoupon service that provides efficient and secure signature scheme against illegal use in mCoupon environment based on low-cost NFC tag for using light-weight authentication method.

Keywords Component · NFC · Light-weight authentication · NTRU · mCoupon · Digital signature

1 Introduction

Recently, NFC based mCoupon service has led to the movement of variety mobile marketing services activities in global mobile market [1]. NFC based newspaper or smart poster is example of it. With a prime example of leveraging NFC, in 2011, Google began ‘Google Offers’ service. Thus, related providers were considered method for just in new NFC service marketing case. As a result, existing NFC based service threats will expect to be maintained. At this point, targets of cost damage by mCoupon security threats are mCoupon service provider. The service

S.-W. Park · I.-Y. Lee (✉)

Department of Computer Software Engineering, Soonchunhyang University, Asan-si
336-745, Republic of Korea
e-mail: imylee@sch.ac.kr

S.-W. Park
e-mail: swpark@sch.ac.kr

provider may be received to damage of cost through Coupons forgery and unauthorized generation and so on by malicious attacker. The remainder of this paper is organized as follows. Section 2 analyzes the related mCoupon research and NFC-based mCoupon scheme. Section 3 analyzes the security requirements of mCoupon service. Section 4 proposes a light-weight authentication for NFC mCoupon environments. Section 5 presents its security requirements and our analysis of the proposed scheme, and Sect. 6 concludes the paper.

2 Related Work

In this chapter, we describe the related authentication scheme based on NFC mCoupon and mobile coupon method.

2.1 Related mCoupon Research

In 1999, Jacobsson et al. [2] first proposed the concept of the “E-Coupon” in their paper. General eCoupons are electronically issued via Internet, users used to coupon using download type or printing type [3]. In comparison with paper coupons, eCoupons have advantages as follow: eCoupons can publish to very low cost, the speed of delivery is very faster, and the users can download eCoupon through internet. Disadvantage of eCoupons are that Coupons are issued by complex procedures. When using coupons, all merchants should require printing the coupon of user. In addition, eCoupon has variety security threats [4]. Recently, the combination of mobile and NFC aroused great momentum on m-commerce market [5]. As a result, marketing services such as eCoupon has evolved mCoupon [6–9].

2.2 mCoupon Classification

In this paper, we will focus on “Smart Poster method based on low-cost NFC”. mCoupons to date have been issued in various forms and through various channels, classified in terms of four categories [10]:

- mCoupons that are published by the mobile operator, similar to SMS (Short Message Service).
- mCoupons is downloaded by user from a website that sells mCoupons.
- mCoupons transmitted via location-based services.
- mCoupons generated using RFID (e.g., handbills, RFID based smart poster, newspapers and so on).

2.3 *NFC mCoupon*

General mCoupon process flow is as following. The user downloads coupons application through the provider and then Coupons is issued to the user through smart poster. Merchant is verified to receive coupons from the user. But, the existing mCoupon is possible to illegal copying or modified without cost by attacker. In 2007, Dominikus et al. and Aigner et al. described two types of security threat on mCoupons [6, 7]. Also, they proposed a new mCoupon scheme for solving related problem. They were referring to ISO/IEC 18092 Standard. In ISO/IEC 18092, NFC Standard provide to NFC-SEC technical for the protection of the wireless communication. However, they were not considering to computing power of NFC tag. Secondly, tag used in smart poster is ISO 14443 standard. It is not provide to cryptosystem. Also, computations ability is very low. It will be able to bit operations. Therefore, in ISO 14443 standard, their scheme is insecure in smart poster environments. In 2009, Hsiang et al. [8] proposed a hash based scheme for solving efficiency problem of Dominikus et al. scheme. But, this scheme could be collected to user ID and coupon data on public network by attacker. Also, attacker could generate to illegal coupon using groups of collected user IDs and coupon data. In this paper, Our proposed light-weight authentication scheme for practical NFC mCoupon service to using authentication scheme based on low-cost NFC for Considering the safety and efficiency [11–13].

3 Security Requirements

Our proposed scheme should be satisfied to basic coupon service security requirements for mCoupon service environments as follows. Also, the proposed scheme needs to provide efficiency and security in limited device environments [6, 14].

- **Multiple Cash-In:** An attacker should not be able to use the same mCoupon multiple times.
- **Manipulation:** mCoupons should not remain to validity after manipulation by attacker.
- **Preventing Unauthorized Generation:** Only Issuer should offer valid mCoupons. An attacker cannot issue his own mCoupons.
- **Confidentiality:** mCoupon offer data generated by mCoupon process should safely protect. In other words, an attacker does not know the secret value generated by the mCoupon Issuer.
- **Efficiency:** mCoupon should be providing computational efficiency in a limited device environment.

4 Proposed Schemes

In this paper, our proposed schemes provided light-weight signature scheme based on lattice and signature scheme based on Schnorr for low-cost NFC using rotate function and hash chain [15–17]. We were considering the practical implementation in NFC mCoupon service environments based on low-cost NFC tag to using limited resources. Our proposed scheme consists of a mCoupon signature generation phase and mCoupon verification phase as follows (Fig. 1).

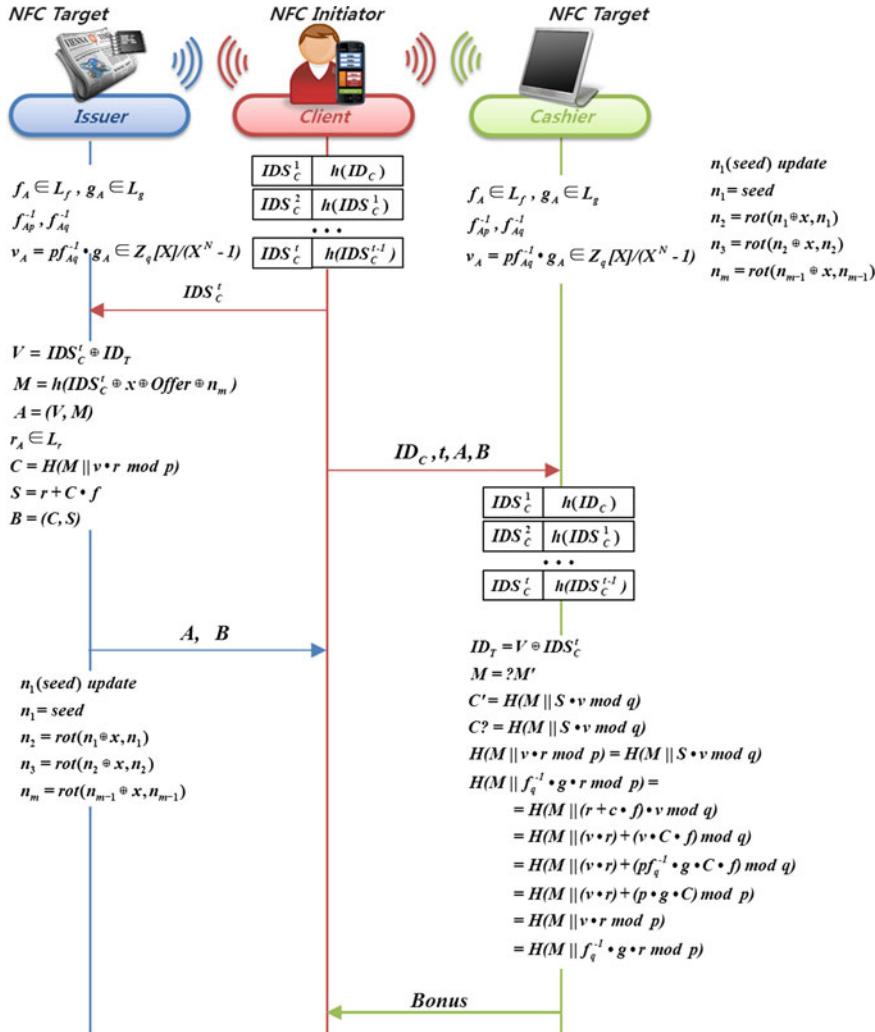


Fig. 1 Over all transactions of proposed mechanism

4.1 System Parameters

The system parameters in the proposed scheme are as follows.

- $*$: object (*A*: *Issuer*, *B*: *Cashier*, *C*: *Client*)
- ID_* : value to verifies the identity of $*$
- IDS^* : random *ID* value of $*$
- *Issuer*: *Issuer* is a passive NFC target attached to an advertising poster or an advertisement in a newspaper that allows issuing mobile coupons on a request from an NFC initiator device.
- *Cashier*: *Cashier* is a terminal with an NFC interface to validate the mobile coupons and provide the goods or service for *Client*.
- *Client*: a user who wants a coupon for a particular product or service and has a mobile device such as a PDA or mobile phone with an NFC initiator where the mobile coupons can be stored.
- t : count value of hash chain
- *Offer*: additional mCoupon data (e.g., validity range of the coupon, type, issuing time and etc.)
- $H()$: hash function
- $Rot()$: rotate function; $Rot(x, y)$ is defined to left rotate the value of $w(y)$ with x . $w(y)$ is hamming weight of y [15, 16].
- x : the secret key of the *Issuers* and *Cashiers*
- Z : set of integers
- L_f, L_g : subset of R (*truncated polynomial ring*)
- f, g : private key polynomial of $*$, $(f_* \in L_f, g_* \in L_g)$
- p, q : large prime number that satisfies $GCD(p, q) = 1$, $p > q$
- $g_{*p^{-1}}, g_{*q^{-1}}$: Inverse polynomial of g
- $f_{*p^{-1}}, f_{*q^{-1}}$: Inverse polynomial of f
- v_* : public key on truncated polynomial $R(v_* = pf_{*q^{-1}}g_* \in Z_q[X]/(X^N - 1))$
- p : prime (1024 bit)
- g : generator of group (primitive root)
- x : private key
- y : public key

4.2 Lightweight mCoupon Authentication Scheme 1 (Lattice Based)

4.2.1 Issuing Phase

For a user wishing to use the mCoupon service, the mCoupon is issued in the following manner. Calculated data (secret key $f_A, g_A, f_{Ap}^{-1}, f_{Aq}^{-1}, x$, coupon info *Offer* and public key v_A on the truncated polynomial ring) using CA (Certification Authority) are stored in advance in the *Issuer* memory.

$$\begin{aligned}
f_A &\in L_f, g_A \in L_g \\
f_{Ap^{-1}}, f_{Aq^{-1}} \\
v_A = pf_{Aq^{-1}} \cdot g_A &\in \mathbb{Z}_q[X]/(X^N - 1)
\end{aligned}$$

Step 1 The user installs the mCoupon software on his/she own mobile. Using this application, the user downloads the mobile identification values (*ID*) from the CA.

Step 2 The *Client* transmits the IDS_{C^t} (t times hashed ID value) to the *Issuer*.

$$\begin{aligned}
C : IDS_{c^1} &= h(ID_C), IDS_{c^2} = h(IDS_{c^1}), \dots, IDS_{c^t} = h(IDS_{c^{t-1}}) \\
C \rightarrow A : IDS_{c^t}
\end{aligned}$$

Step 3 The *Issuer* computes the following:

$$\begin{aligned}
A : V &= IDS_{c^t} \oplus ID_T \\
A : M &= IDS_{c^t} \oplus x \oplus Offer \oplus n_m \\
A : A &= (V, M)
\end{aligned}$$

Step 4 The *Issuer* selects a random polynomial r_A . The value of r_A is updated whenever a signature is generated.

$$A : r_A \in L_A$$

Step 5 The *Issuer* sends the mCoupon $A = \{V, M\}$ and A's signature $B = \{C, S\}$ to the *Client*, The *Client* saves A and B in memory. The *Issuer* updates the seed value n_1 using the *rot()* function. By updating the seed value, the mCoupon is safe from illegal mCoupon generation by attacker.

$$\begin{aligned}
A : C &= H(M || v \cdot r \bmod p) \\
A : S &= r + C \cdot f \\
A : B &= \{C, S\} \\
A \rightarrow C : A, B \\
A : n_1 &= seed \\
A : n_1 &= seed \\
A : n_2 &= rot(n_1 \oplus x, n_1) \\
A : n_3 &= rot(n_2 \oplus x, n_2) \\
A : n_m &= rot(n_z \oplus x, n_{m-1})
\end{aligned}$$

Algorithm number of message	SHA-1	AES	NTRU	RSA	AES	NTRU	RSA
	Encryption (s)				Decryption (s)		
1	0.001	0.023	0.003	0.014	0.017	0.009	0.042
100	0.022	0.164	0.056	0.186	0.121	0.102	0.783
500	0.047	0.204	0.143	0.366	0.247	0.549	3.305
1000	0.068	0.252	0.221	0.583	0.503	1.218	6.721
2000	0.092	0.332	0.408	0.964	0.784	2.312	13.27

4.2.2 Cashing and Authentication Phase

When the Client wishes to use the coupon service, he/she can tagging NFC reader (e.g., Cashier) to the his own NFC Mobile (e.g., mCoupon) and performs the following operations:

Step 1 The *Client*'s mobile device sends mCoupon $A = \{V, M\}$, signature value $B = \{C, S\}$, hash chain value t , and the identity of the *Client* ID_C to the *Cashier*.

Step 2 After the message ID_C, A, B and t are received, and the *Cashier* computes $IDS_{C'} = h(IDS_{C^{t-1}})$ and $ID_T = IDS_{C'} \oplus V$ to obtain $IDS_{C'}$ and IDT .

$$B : IDS_{C^1} = h(ID_C), IDS_{C^2} = h(IDS_{C^1}), \dots, IDS_{C'} = h(IDS_{C^{t-1}})$$

$$B : ID_T = IDS_{C'} \oplus V$$

Step 3 The *Cashier* verifies the integrity of the mCoupon value M using the secret value (x , seed table value, coupon info *Offer*) of the identity of the *Issuer*.

$$B : M = ?M$$

Step 4 The *Cashier* verifies the integrity of signature A , and authenticates the coupon.

$$B : C = ?C'$$

$$B : H(M||v \cdot r \bmod p) = ?H(M||S \cdot v \bmod q)$$

$$B : H(M||pf_{q^{-1}} \cdot g \cdot r \bmod p) = ?H(M||S \cdot v \bmod q)$$

$$B : H(M||pf_{q^{-1}} \cdot g \cdot r \bmod p) = ?H(M||(r + c \cdot f) \cdot v \bmod q)$$

$$B : H(M||pf_{q^{-1}} \cdot g \cdot r \bmod p) = ?H(M||(v \cdot r) + (v \cdot c \cdot f) \bmod q)$$

$$B : H(M||pf_{q^{-1}} \cdot g \cdot r \bmod p) = ?H(M||(v \cdot r) + (pf_{q^{-1}} \cdot g \cdot c \cdot f) \bmod q)$$

$$B : H(M||pf_{q^{-1}} \cdot g \cdot r \bmod p) = ?H(M||(v \cdot r) + (p \cdot g \cdot c) \bmod p)$$

$$B : H(M||pf_{q^{-1}} \cdot g \cdot r \bmod p) = ?H(M||(v \cdot r) \bmod p)$$

$$B : H(M||pf_{q^{-1}} \cdot g \cdot r \bmod p) = ?H(M||pf_{q^{-1}} \cdot g \cdot r \bmod p)$$

4.3 *Lightweight mCoupon Authentication Scheme 2 (Schnorr Based)*

4.3.1 Issuing Phase

For a user wishing to use the mCoupon service, the mCoupon is issued in the following manner. Calculated data (secret key x , random value w and coupon info $Offer$) using CA (Certification Authority) are stored in advance in the *Issuer* memory.

- Step 1 The user installs the mCoupon software on his/she own mobile. Using this application, the user downloads the mobile identification values (ID) from the CA .
- Step 2 The *Client* transmits the IDS_{C^t} to the *Issuer*.

$$C : IDS_{c^1} = h(ID_C), IDS_{c^2} = h(IDS_{c^1}), \dots, IDS_{c^t} = h(IDS_{c^{t-1}}) \\ C \rightarrow A : IDS_{c^t}$$

- Step 3 The *Issuer* computes the following:

$$A : V = IDS_{c^t} \oplus ID_T \\ A : M = IDS_{c^t} \oplus x \oplus Offer \oplus n_m \\ A : A = (V, M)$$

- Step 4 The *Issuer* generated coupon info A 's signature $B = \{C, S\}$.

$$A : C = H(A || g^w \bmod p) \\ A : S = w - Cx \bmod q \\ A : B = \{C, S\}$$

- Step 5 The *Issuer* sends the mCoupon $A = \{V, M\}$ and A 's signature $B = \{C, S\}$ to the *Client*, The *Client* saves A and B in memory. The *Issuer* updates the seed value w, n_1 using the $rot()$ function By updating the seed value. In this way, the coupon providing to security from unauthorized generation by an attacker.

$$\begin{aligned}
A &\rightarrow C : A, B \\
A &: n_1(\text{seed}) \text{ update} \\
A &: n_1 = \text{seed} \\
A &: n_2 = \text{rot}(n_1 \oplus x, n_1) \\
A &: n_3 = \text{rot}(n_2 \oplus x, n_2) \\
A &: n_m = \text{rot}(n_z \oplus x, n_{m-1}) \\
A &: \text{key}(w, g^w) \text{ update} \\
A &: w_i = w_{i-1} + 1 \\
A &: g^w = g^w_{i-1} + 1 = g^w \cdot w
\end{aligned}$$

Number of message					
Related work	1	100	500	1000	2000
Registration (s)					
Aigner	0.037	0.35	0.57	0.835	1.296
Dominikus	0.051	0.536	0.936	1.418	2.26
Hsiang	0.002	0.044	0.094	0.136	0.276
Schnorr	0.015	0.208	0.413	0.651	0.908
Proposed scheme 1	0.007	0.134	0.333	0.51	0.908
Proposed scheme 2	0.001	0.022	0.047	0.068	0.092
Authentication (s)					
Aigner	0.059	0.904	3.552	7.224	14.055
Dominikus	0.101	1.687	6.857	13.95	27.33
Hsiang	0.002	0.044	0.094	0.136	0.184
Schnorr	0.101	1.687	6.857	13.95	27.33
Proposed scheme 1	0.01	0.124	0.596	1.286	2.404
Proposed scheme 2	0.085	1.588	6.657	13.51	26.63

4.3.2 Cashing and Authentication Phase

When the Client wishes to use the coupon service, he/she can tagging NFC reader (e.g., Cashier) to the his own NFC Mobile (e.g., mCoupon) and performs the following operations:

- Step 1 The *Client*'s mobile device sends mCoupon $A = \{V, M\}$, signature value $B = \{C, S\}$, hash chain value t , and the identity of the *Client* ID_C to the *Cashier*.
- Step 2 After the message with ID_C , A , B and t is received, the *Cashier* computes $IDS_{C'} = h(IDS_{C'^{-1}})$ and $ID_T = IDS_{C'} \oplus V$ to obtain IDS_C' and ID_T' .

$$B : IDS_{c^1} = h(ID_C), IDS_{c^2} = h(IDS_{c^1}), \dots, IDS_{c^t} = h(IDS_{c^{t-1}})$$

$$B : ID_T = IDS_{c^t} \oplus V$$

Step 3 The *Cashier* verifies the integrity of the mCoupon value M using the secret value (x , seed table value, coupon info *Offer*) of identity of the *Issuer*.

$$B : M = ?M$$

Step 4 The next stage proceeds in a manner similar to that in the general Schnorr signature scheme.

$$B : C = ?C'$$

$$B : C = ?H(A \parallel g^s y^C \bmod p)$$

5 Analysis of Proposed Scheme

Our proposed scheme satisfies the following requirements. In particular, our proposed scheme provides higher efficiency. The first proposed scheme provides higher efficiency than the method based on the exponential because it is based on a lattice. The second scheme is storage-efficient because it uses only simple arithmetic and modulus operations.

- **Multiple Cash-In:** An attacker cannot be used to the same mCoupon multiple times because it is managed by the verification phase of database.
- **Manipulation:** mCoupons do not stay to validity after a manipulation by attacker because updated authentication data $M (=IDS_{c^t} \oplus x \oplus Offer \oplus n_m)$.
- **Preventing Unauthorized Generation:** An attacker cannot issue his own mCoupons. The proposed scheme protects the legal user using the signature based on lattice and signature based on Schnorr to the mCoupon of the *Client*.
- **Efficiency:** The proposed schemes are very efficient and secure because it uses only signature based on polynomial operations, arithmetic operations (e.g., +, xor, rotate function) and simple hash operations.
- **Non-repudiation:** Non-repudiation is ensured by the light-weight signature generated by the lattice method and proposed key update function ($n_I(seed)$ and $key(w, g^w)$ update methods).
- **Confidentiality:** mCoupon offer data generated by mCoupon process is safely protected because an attacker does not know the secret value x generated by the mCoupon issuer(legal object).

The Dominikus et al. scheme provides integrity for coupon data by using a digital signature based on PKI. However, their scheme is inefficient because it

increases computational complexity by using exponential computation and many exchanges of data. The Hsiang et al. scheme provides high efficiency using only operations based on the hash function and 2-round traffic. Their scheme, though, does not provide a digital signature function. Therefore, coupon information is transmitted in plaintext without signature on an open channel. Consequently, the attacker is able to collect the ID of the legal user for malicious purposes and can generate illicit coupons using the collected ID group. The proposed scheme prevents the re-use of the signature by using a random polynomial r, w updated whenever a signature is generated. Also, the user can minimize the damage using an updated signature each time a coupon is generated.

The proposed scheme can operate in low-cost NFC environments because, compared to existing methods, in terms of computational efficiency, it is very good. In existing studies, the designers of NTRU claim that when comparing a moderate NTRU security level to RSA with a 512-bit modulus, NTRU is approximately 5.9 times faster for encryption, 14.4 times faster for decryption, and 5.0 times faster during key creation [12, 18–20]. Also, In 2008, Atici et al. [21] designed an NTRU architecture for encryption/decryption that required only 10.8 kgate. This is similar to gate quantities for implementing hash functions. However, in this paper, we performed a source-code implementation to analyze the practical efficiency of each cryptographic algorithm. The parameters used in each cryptographic algorithm

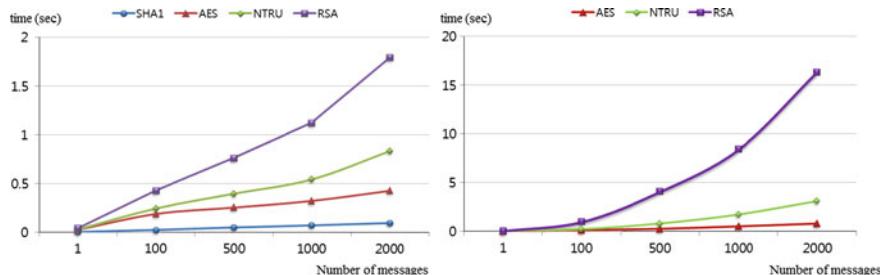


Fig. 2 Comparing computational of cryptographic algorithms

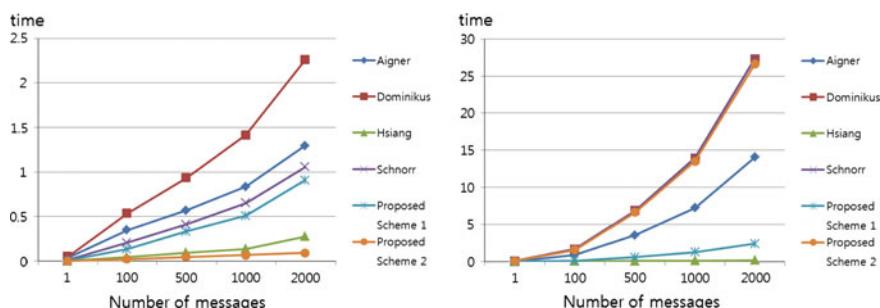


Fig. 3 Comparison of computational registration and authentication phases

Table 1 Analysis of the proposed schemes

			Efficiency		Computation quantities		Traffic	
			Reg. (tag)	Auth.	Reg. (tag)	Auth.	Reg. (tag)	Auth.
Aigner [6]	Δ	○	x	x	1U+1E	1U+1E	4rounds	5rounds
	ID based	DB based	PKI based (be able to attack)	Exponential based				
Dominikus [7]	Δ	○	○	x	2U+1E	2U+1E	4rounds	4rounds
	ID based	DB based	PKI based	Exponential based				
Hsiang [8]	Δ	○	x	Δ	2H+3 \oplus	2H+3 \oplus	2rounds	2rounds
	ID based	DB based	No have function	Hash based				
Schnorr [13]	Δ	○	○	x	1U+1H +1M	2U+1H	2rounds	2rounds
	ID based	DB based	PKI based	Exponential based				
Proposed scheme 1	○	○	○	○	2C+1H +5 \oplus	1C+1H +4 \oplus	2rounds	2rounds
	Hash ID based	DB based	PKI based (NTRU)	Polynomial based				
Proposed scheme 1	Δ	○	○	Δ	2M+1H +5 \oplus	2U+1H +4 \oplus	2rounds	2rounds
	Hash ID based	DB based	PKI based (Schnorr)	Exponential based				

× non-offer, insecure; Δ usually-offer; ○ offer, secure; E symmetric key; H hash algorithm; U public key; C convolution multiplication; M multiplication

(parameters: RSA-1024, NTRU-251, SHA-1-256, AES-256) were chosen to obtain a security level equivalent to 1024-bit RSA. As shown in Figs. 2 and 3, we obtained similar data to the existing studies. Figure 2 shows computation quantities of the encryption algorithm applied to the related schemes and proposed scheme. Also, in Fig. 3, computation quantities per message for each schemes are the same as computation quantities of the registration and authentication phases shown in Table 1 (per message—Aigner: 1U+1E, Dominikus: 2U+1E etc.).

In the registration phase, for the first case of the proposed scheme, according to the computational proof, the efficiency of the proposed method should have been higher than the efficiency in the existing methods. The actual implementation did not attain the calculated results. However, the computational complexity was found to be similar. The second proposed scheme, however, is very efficient because it uses only polynomial operations, arithmetic operations and simple hash operations.

6 Conclusions

In this paper, we proposed a practically secure mCoupon scheme for the protection of both the user and service provider in mCoupon service environments based on low-cost NFC tag. Our scheme satisfies the necessary requirements and therefore could be effectively applied in a mCoupon service environment. Furthermore, unlike existing schemes, our method can be used for practical NFC environments because it can be effectively applied in an ISO 14443 standard environment. We proved the efficiency of our method through a practical implementation of the cryptographic algorithm. However, since we have not applied our proposed method in a practical environment, directly computational times or other numerical measures are unknown. In a future work, we will compare our proposed model with previous mCoupon methods using a practical implementation of mCoupon service environment based on low-cost NFC tag.

In the authentication phase, in the second case of the proposed method, the computational complexity was comparable to the conventional methods. For the first case of the proposed method, compared with traditional methods, the proposed method provided a very high efficiency. Combining these results, the efficiency for the proposed method, for both the client and server sides, should be acceptable in smart poster environments.

Acknowledgment This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the MEST (Ministry of Education, Science and Technology) (2013R1A1A2012940) and the Soonchunhyang University.

References

1. Coskun, V., Ozdenizci, B., Ok, K.: A survey on near field communication (NFC) technology. *Wireless Pers. Commun.* **71**(3), 2259–2294 (2013)
2. Jakobsson, M., MacKenzie, P.D., Stern, J.P.: Secure and lightweight advertising on the web. *Comput. Netw.* **31**(11–16), 1101–1109 (1999)
3. Fortin, D.R.: Clipping coupons in cyberspace: a proposed model of behavior for deal-prone consumers. *Psychol. Mark.* **17**(6), 515–534 (2000)
4. Garg, R., Mittal, P., Agarwal, V., Modani, N.: An architecture for secure generation and verification of electronic coupons. In: Proceedings of the Usenix Annual Technical Conference, Boston, Mass, USA, 2001
5. Lee, M.-K., Kim, J.W., Song, J.E., Park, K.: Sliding window method for NTRU. In: Applied Cryptography and Network Security, pp. 432–442, Springer, Berlin (2007)
6. Aigner, M., Dominikus, S., Feldhofer, M.: A system of secure virtual coupons using NFC technology. In: Proceedings of the 5th Annual IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom '07), pp. 362–366, March 2007
7. Dominikus, S., Aigner, M.: mCoupons: an application for near field communication (NFC). In: Proceedings of the 21st International Conference on Advanced Information Networking and Applications Workshops (AINAW '07), pp. 421–428, May 2007
8. Hsiang, H.-C., Kuo, H.-C., Shih, W.-K.: A secure mCoupon scheme using near field communication. *Int. J. Innov. Comput. Inf. Control* **5**(11), 3901–3909 (2009)
9. Chang, C.-C., Sun, C.-Y.: A secure and efficient authentication scheme for E-coupon systems. *Wireless Pers. Commun.* **77**(4), 2981–2996 (2014)
10. Hsueh, S.-C., Chen, J.-M.: Sharing secure m-coupons for peer-generated targeting via eWOM communications. *Electron. Commer. Res. Appl.* **9**(4), 283–293 (2010)
11. Park, S.-W., Lee, I.-Y.: Efficient mCoupon authentication scheme for smart poster environments based on low-cost NFC. *Int. J. Secur. Appl.* **7**(5), 131–138 (2013)
12. Hoffstein, J., Pipher, J., Silverman, J.H.: NTRU: a ring-based public key cryptosystem. In: Algorithmic Number Theory (1998)
13. Schnorr, C.P.: Efficient signature generation by smart cards. *J. Cryptology* **4**(3), 161–174 (1991)
14. Chang, C.C., Wu, C.C., Lin, I.C.: A secure E-coupon system for mobile users. *IJCSNS Int. J. Comput. Sci. Netw. Secur.* **6**(1) (2006)
15. Chien, H.-Y.: SASI: a newultralightweight RFID authentication protocol providing strong authentication and strong integrity. *IEEE Trans. Dependable Secure Comput.* **4**(4), 337–340 (2007)
16. Phan, R.C.-W.: Cryptanalysis of a new ultralightweight RFID authentication protocol—SASI. *IEEE Trans. Dependable Secure Comput.* **6**(4), 316–320 (2009)
17. Patil, V., Shyamasundar, R.K.: E-coupons: an efficient, secure and delegable micro-payment system. *Inf. Syst. Front.* **7**(4–5), 371–389 (2005)
18. A Study on the Development of Cryptosystems for the Next Generation. National Security Research Institute (2006)
19. Stamp, M., Richard, M.L.: Applied Cryptanalysis: Breaking Ciphers in the RealWorld. Wiley, Hoboken (2007)
20. Noh, S.-K., Lee, S.-R., Choi, D.: Proposed M-payment system using near-field communication and based on WSNenabled location-based services for M-commerce. *Int. J. Distrib. Sens. Netw.* **2014**(865172): 8 pages (2014)
21. Atici, A.C., Batina, L., Junfeng, F., Verbauwhede, I., Yalcin, S.B.O.: Low-cost implementations of NTRU for pervasive security. In: Proceedings of the International Conference on Application-Specific Systems, Architectures and Processors (ASAP'08), pp. 79–84 (2008)

Author Biographies



Sung-Wook Park received the B.S. and M.S. degrees in Department of Computer Software Engineering from Soonchunhyang University, Korea, in 2011 and 2013, respectively. He is now a Ph.D. candidate in Department of Computer Software Engineering from Soonchunhyang University, Korea. His research interests include NFC Security, NTRU Cryptography, Ultra Lightweight Cryptography, etc.



Im-Yeong Lee is corresponding author. He received the B.S. degrees in Department of Electronic Engineering from Hongik University, Korea, in 1981 and the M.S. and Ph.D. degrees in Department of Communication Engineering from Osaka University, Japan, in 1986 and 1989, respectively. From 1989 to 1994, he had been a senior researcher at ETRI (Electronics and Telecommunications Research Institute), Korea. Now he is a professor in Department of Computer Software Engineering from Soonchunhyang University, Korea. His research interests include Cryptography, Information theory, Computer and Network security.

Model of CPU-Intensive Applications in Cloud Computing

Junjie Peng, Yongchuan Dai, Yi Rao and Xiaofei Zhi

Abstract CPU-intensive application is one of the most commonly used application type in cloud computing. In order to effectively deal with CPU-intensive applications and improve the application efficiency of cloud computing, we have studied a lot on CPU-intensive application. Through the studies, we draw out some common features and characteristics of this kind of applications. Based on the features found, we establish a mathematical model for the CPU-intensive application which can be used to predict and analyze whether an unknown application is CPU intensive one or not. To verify the correctness of the model, we have done extensive experiments. The experimental results show that the model is correct and reasonable. It can effectively distinguish CPU intensive application from other kinds of applications. This is very helpful as it can serve as the basis for study of special process strategies for CPU intensive applications which can much benefit the application improvement.

Keywords CPU-intensive application · Cloud computing · Efficiently · Model

J. Peng (✉) · Y. Dai · Y. Rao · X. Zhi

School of Computer Engineering and Science, Shanghai University, Shanghai, China
e-mail: jjie.peng@shu.edu.cn

Y. Dai
e-mail: dai.yongchuan@163.com

Y. Rao
e-mail: roydibo@qq.com

X. Zhi
e-mail: zhiyunbin@qq.com

1 Introduction

Cloud computing is considered as the revolution of IT area [1]. It is bringing as well as will lead to many changes to the society. Actually, it has much affected the customs how people use IT resources as well as its daily social life. The rapid development of cloud computing is led by its advantages over traditional IT services. Firstly, virtualization technology guarantees cloud computing provide much more powerful and flexible service capability compared with traditional PC era with the same resources. Secondly, it offers on-demand service according to the requirements of customers.

Though cloud computing offers many convenient services for people, it is far more perfect and there still some aspects need to be optimized. One of the key technologies of cloud computing is virtualization technology [2]. It can create a pool of resources—many physical nodes or servers are integrated into a whole or one physical server is splitted as many virtual servers, to offer sufficient resources to create virtual machines to clients. However, the performance of servers may be quite different from one to one. For example, one node may be in excellent performance on disk, however poor performance on CPU. But others are not or vice versa, even this is transparent to the clients. This difference may lead some issues. Such as different applications in cloud may have different resource preference. So there is the incompatibility between specification of physical machine and user requests in cloud, and it may lead towards some problems like poor load balancing, energy-performance trade-off and large power consumption et al. Thus, the efficiency of cloud computing will be greatly affected. For instance, if an application may require a better performance on CPU, but on the contrary it is assigned to a physical node with poor performance on CPU. Probably, this leads to one case that the node with the resources of CPU cannot accommodate any virtual machines except that one. The consequence of the result is to eliminate the advantage of the virtualization technology and reduce the processing efficiency.

In this paper, we study the rules of CPU intensive applications in cloud considering the performance loss caused by unreasonable match between the cloud resources and application requests. Based on the rules of CPU intensive application, we find some common characteristics and features which we can use to establish the mathematical model of the CPU intensive applications. What's more, we use the model to forecast and analysis the type—CPU-intensive application, and conduct a special processing strategy research.

2 Related Work

Cloud computing relies on virtualization technology to consolidate a strong pool of resources, but resources of cloud computing center, after all, are relatively limited. How to maximize the service capability, reasonably allocate and use cloud

resources on the premise of guaranteeing the quality of service becomes one of the hottest topics among scholars. To solve this problem, many research efforts have set focus on it. And lots of interesting results have been obtained. For example, Lee et al. [3] presented a dynamic voltage scaling method which can automatically adjust the resources by inspecting CPU utilization in cloud computing. The voltage of the idle or light loaded servers can be reduced and heavy loaded applications can be migrated to the machines with lighter load for the purpose of energy saving. Maurya and Sinha [4] tried to solve the issue, the incompatibility between the specification of physical machine and user requests in cloud computing, and put forward an energy conscious, power-aware load balancing strategy based on adaptive migration of virtual machines (VMs). This strategy tries to implement load balance in cloud by considering the high and low thresholds for migration of virtual machines. However, they do not consider the relationship the type of applications and features of physical machines. Other scholars put forward virtual machine deployment mechanisms with good performance by fully considering the competition of resources among virtual machines [5]. They proposed a model to classify the incoming workloads based on the heterogeneity of cloud resources and the resource usage patterns. In this model, they considered the features of CPU and Memory for applications before the migration of virtual machines. Besides, some other researchers presented different VM allocation mechanisms based on some load balancing strategies [6]. These mechanisms are also load based solution without considering the types of applications.

Different the research thread mentioned above, other scholars studied the efficiency of some specific cloud application based on known cloud architecture. For example, on the hadoop platform, Fadika et al. presented the design decisions and implementation tradeoffs for DELMA (Dynamically Elastic MapReduce), a framework that follows the MapReduce paradigm, just like Hadoop MapReduce. But it permits automatically scaling its cluster size as jobs are in operation [6]. Hiroaki et al. [7] analyzed the scheduling and processing strategy of the multi-thread memory-intensive applications on hadoop. To reduce the data delay between nodes on hadoop, Kuo et al. [8] puts forward a new algorithm on the allocation of virtual machines. In addition, there is a study on novel resource estimation model and scheduling algorithm about scientific application to optimally allocate the resources in cloud computing [9].

Besides the study mentioned above, some researchers studied performance of cloud computing partly considered the characteristics of applications. For example, to optimize the multimedia processing in cloud computing environment, Alasaad et al. [10] proposed a requirement-prediction, resource-reservation based algorithm which can be used to minimize the cost of using cloud resources. Wu et al. [11] has studied the characteristics of the scientific application, and designed a new framework to improve the efficiency of compute-intensive application processing.

All the research mentioned is interesting. They can improve cloud application performance at least from some aspects [12, 13]. However, they fail to consider the resource requirement for different cloud application and could not judge the type of an unknown cloud application. Especially, for many unknown applications in cloud

center, there is no solution for current research. Actually, the performance and efficiency of servers in cloud center are much related to the types of applications. As we know, different types of applications have different preference on cloud resources. The same type of applications running on a server may cause some resources are overloaded while others are little used. This will lead to low usage efficiency of the servers. To fully exert the efficiency of physical resources, we should clearly know the resource preferences for different kind of applications in cloud. So modeling the applications in cloud environment is very important. Based on this consideration, this paper attempts to study the CPU-intensive applications to find some common features through a lot of research and experiments. And using the features obtained we try to build a mathematic model of the CPU intensive applications so that this kind of applications in cloud computing environment can be predicted and analyzed. This further can be used as the theoretical basis for future research on special services strategy for CPU intensive applications.

3 Modeling for CPU-Intensive Applications

3.1 *Classification of Resources in Applications*

Generally, one virtual machine should include the following different types of resources. These are CPU frequency, number of CPUs, memory size, volume of disk and network bandwidth traffic etc. If resources of a virtual machine are initialized, computing power (depending on the CPU frequency and the number of the CPUs), memory capacity, storage capacity and bandwidth capacity are accordingly configured.

Each application has its own characteristics. And it has different demands for resources on the node. For example, requirements for a web server and a file server are not the same. Some scientific computing application requires a lot of CPU resources, and some media streaming application requires a lot of memory resources, some database application needs a lot of read and write operations of disks and so on. According to the different demands of different applications, we can classify the type of application. This paper especially discusses CPU-intensive applications and tries to establish the model of CPU-intensive applications.

3.2 *Modeling for CPU-Intensive Applications*

Our extensive research shows the CPU-intensive applications mainly consume the CPU resources, but also a small amount of memory resources (mainly occupied by the system of the virtual machine), etc. CPU resources are divided into four main parts: namely cpu-usr, cpu-sys, cpu-wai and cpu-idle. The cpu-usr represents the

user space consumption that accounts for the ratio of the CPU, the cpu-sys represents the kernel and interrupt that accounts for the ratio of the CPU, the cpu-wai represents the I/O wait that accounts for the ratio of the CPU, the cpu-idle represents the CPU idle time that accounts for the ratio of the CPU. When the CPU-intensive applications are on a physical node, the cpu/usr generally has maximum values, such as $\text{cpu-usr} \gg \text{cpu-sys}$. The values of sys is both less, but the cpu-wai may reflect whether the application has intense I/O operations.

Although the CPU usage can illustrate the current CPU usage, it can't reflect the rate of CPU load. For even the utilization of CPU is at full, it also cannot identify CPU's ability reach the maximum limit. So how to measure the utilization of the CPU load is prerequisite, we take an important parameter named loadavg—the average number of processes in the run queue within a specific time interval. For a sampling, when physical server only handle an application, the use of CPU(occupied by applications) utilization rate is 100 %, but its loadavg is very small due to that there is no other applications competing for the CPU. But when a CPU is occupied by multiple applications, the CPU will distribute the corresponding time slice in accordance with the proportion of frequency distribution to each application, the CPU utilization rate is 100 %, with a higher loadavg.

By observing CPU usage of the different applications, we find that the change of the load on CPU, its performance can be obviously divided into three different stages: In the first stage, the number of active processes on CPU increases which has no effect on each application, namely the execution time of the applications does not prolong. We call this stage sound stage. The second stage, with the increase of CPU activity processes, the execution time of the current application will prolong accordingly. But the prolonged increases slowly, not dramatically. We call it accepted stage. The third stage, with the increase of CPU activity processes, the current application execution time will increase sharply. In general, the increase of time will increase exponentially. We call this stage—unfavorable stage. In order to better distinguish among three different stages and make good use of CPU processing efficiency, we define two thresholds M and K ($M < K$) to distinguish between the working state of the CPU's stages. When the number of active processes on the CPU is less than M , then it is in the sound stage. And when the number of the active process on the CPU is greater than the number M and less than K , then it is acceptable performance stage; when the number of the active process on the CPU is greater than K , then it is in unfavorable stage. Obviously, in order to guarantee the CPU good processing efficiency, we should try to avoid its work on the third stage.

When CPU-intensive applications are running in physical nodes, there is little network load. It is that CPU-intensive applications don't need to much exchange data with the outside devices. In the actual monitoring, we find that there is still a very small number of data exchange. It is because the cloud management platform used to monitor the running status of the virtual machines. At the same time, the storage node and computing node are separated. So there is some network traffic between the storage node and computing node. But the communication volume is

little and the effect of bandwidth is very small. Then this can be used as a feature of the CPU-intensive.

By observing the disk I/O operations of CPU-intensive applications on the physical nodes, we find that this kind of application is very little demand for operating disk I/O. For disk rotational speed at 7200 r/s, read/write rate is generally greater than 60 MB/s of mechanical disk. Disk I/O read/write rate generally keeps at around 100 KB/s (including the influence of some monitoring programs) for CPU-intensive applications. Therefore, disk read/write rate, can be used as a measurement that the value of cpu-wai factors of the CPU. By observing the different disk I/O applications, we find that the variation of the cpu-wai mainly divided into two cases: In the first case, when its disk read/write rate reaches to 1 MB/s or even more, the disk I/O operations become the main mission in computer, so cpu-wai may far outweigh cpu-usr. The second case, when the disk read/write rate goes around 100 KB/s, the impact of disk I/O operations on the system is very small. So the cpu-wai will be significantly less than the cpu-usr. Obviously, read/write disk I/O rate and wai go hand in hand. So if cpu-wai is too high and more than cpu-usr, we can infer that this application is not CPU-intensive application.

CPU-intensive applications do not take up a lot of memory resources, but the cost of every virtual machine takes up part of physical nodes memory resources. In addition, the CPU-intensive applications do not have more demand for memory resources in physical nodes.

A standard Linux kernel can support the operation of 50–50,000 processes running. For the average CPU, the kernel will be scheduling and execution of these processes. Each process can have a time slice to run, but when a process is run out of time or taken by a process with higher priority, it will be hung up to the CPU running queue, and other process will take up the CPU resources. The switch of CPU from one process to another is called the context switch (csw). Too much context switch will cause much overhead. IO-intensive applications don't need a lot of time slice, because the system is mainly IO operations. So it will cause much context switches. However CPU-intensive applications need much time slice to maintain the effectiveness of the cache. So it causes little context switches. When CPU-intensive applications is in operation, the change of the number of context switches per second is very small.

To summarize, we can establish a quintuple model with the important features of CPU-intensive applications which can be used to judge whether an application is CPU-intensive application or not.

$$T = \{cpu\text{-}usr, otherCpu, Net, IO\} \quad (1)$$

The cpu-usr represents the ratio of user space consumption to CPU. other={cpu-sys}, cpu-sys represents the ratio of the kernel and interrupt to CPU. Net={rec, send}, rec represents the receiving rate of network, and send means the transmission speed of network. IO={read, write}, read represents the rate of read operation of disk I/O, and write is the rate of write operation of disk I/O. Through

Table 1 CPU-intensive application model

Parameter	General rules	Specific standards
cpu-usr	cpu-sr increases a lot	cpu-usr's growth is more than 20 %
csw	csw compared with CN, CN is a threshold	csw < 5000, CN is 5000 per second
cpu-sys	cpu-sys increases a little	cpu-sys is less than 8 %
rec+send	rec+send is very small	rec+send < 300 KB/s
read+write	read or write speed is almost zero, system needs a small amount of write operations	read+write < 100 KB/s

extensive analysis and experiments, we obtain some rules and paradigms about CPU-intensive applications as shown in Table 1.

3.3 Analysis of the Model

Model definition and features table of CPU-intensive applications is given by the previous section, you can refer to this model to distinguish whether an unknown application is the CPU intensive application or not. The specific steps are as the follows.

1. cpu-usr, monitor the cpu-usr and judge whether the cpu-usr's growth reaches to the threshold or not. In our paper, if its growth says more than 20 %, we go to the step (2), or go to the step (6).
2. csw, monitor the csw and judge whether the value of context switches per second is below a certain threshold, the current context switching threshold is 5000. If it is less than 5000, we go to the step (3), or go to the step (6).
3. cpu-sys, monitor the cpu-sys, and observe whether value of its growth is less than a threshold (the current environment value is 8 %) or not. If its value is less than the threshold, go to the step (4), or go to the step (6).
4. rec+send, monitor the rec and the send, and observe the traffic of the network exchange. If the traffic is less than one threshold (in our paper, this threshold is 300 KB/s), go to step 5, otherwise go to step (6).
5. read+write, analyze the speed of I/O read and write. Check if the read and write speed is very high and compare with the threshold. If the read/write speed is bigger than the threshold, go to next step, otherwise, go to step (6).
6. If all conditions are met, we can determine unknown application is one of the CPU-intensive applications. Otherwise, if the application cannot meet any one condition among these conditions, the application is not the CPU intensive applications.

4 Experimental Results and Analysis

4.1 Experiment Environment

This experiment is based on CLOUDSTACK version 4.2 in which virtualization technique is KVM 0.12.1, database is MySQL 5.1.61 and file system is NFS 4.0. The experimental environment is supported by four PCs connected with 1.0 Gbps network cards. Two PCs are in the same configurations which include I5 3470 CPU with frequency 3.6 GHz, DDR3 memory with 8 GB volume, 7200 r/m disk with 1 TB volume. The other two PCs have different configurations. One owns I3 2130 CPU with 3.4 GHz, DDR3 memory with 4 GB volume, 7200 r/m disk with 1 TB volume. The forth has AMD Phenom II CPU with 3.2 GHz, DDR3 memory with 2 GB volume, 7200 r/m disk with 500 GB volume.

In the experiment, PC with 4 GB memory serves as the management server for managing a plurality of regions (usually refers to the data center including a large number of physical hosts). The two PCs with the same configurations are used as the computing node. They run a large number of virtual machines. The PC with only two GB memory serves as the secondary storage server to store templates, snapshots, volumes. Besides, the management server also serves as the main storage server to store the VM disk images.

All of Physical hosts run 64-bit redhat6.3 server operating systems. And operating system on virtual machines is ubuntu12.04 64-bit server version. To conveniently observe physical host system parameters, we adopt dstat as monitoring tools through in the experiments. It is a versatile system information statistical tool that can conveniently set the sampling frequency and output the monitored information to files for the performance analysis.

4.2 Experiment

In this paper, we have done a lot of experiments on some typical CPU intensive applications, such as matrix calculation, PI calculation and computation of prime Numbers and etc. Matrix calculation is a $700 * 700$ application running with the MPI. PI calculation is a standard c++ program trying to compute the result with 10 million bits precision. Prime number calculation is an application to find out 100 million prime numbers from the start of one with standard java program.

We run the applications independently on different number of VMs with the same configurations which include a CPU with two cores and 1.0 GHz frequency, 1.0 GB memory 30 GB disk and unlimited network bandwidth. In order to verify the characteristics of resource consumption of the CPU-intensive applications, we run the same application with different number of VMs on one physical node simultaneously.

Fig. 1 The relations between cpu-usr and VMs

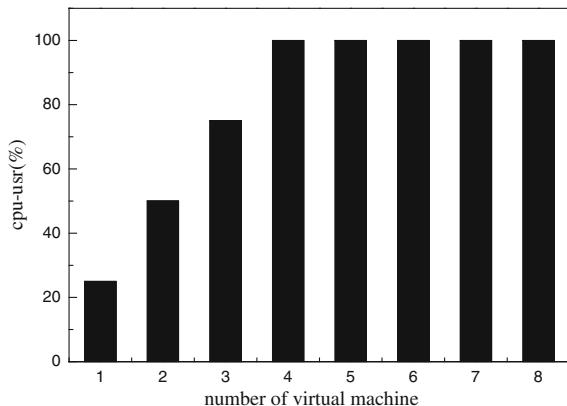


Figure 1 depicts the relations between the number of VMs and cpu-usr. Figure 2 presents the relations between the number of VMs and processing time of application. From Fig. 1, it is easy to find that when the number of VM is less than 4, cpu-usr increases with the increase of the number of VM. When the number of VM is greater than 4, cpu-usr does not change. Actually, when increase one VM cpu-usr increase about 25 %. And when the number of VM running on the physical node reaches to 4, cpu-usr gets to its maximum 100 %. Similarly, it is easy to find out from Fig. 2 that when the number of VM is less than 4, the execution time of the application does not affect by the number of VM. It is about 730 s. However, when the number of VM is 5, the execution time of the application increases almost 50 %, reaches to about 1075 s. And if the number of VM increases to 8, the execution time of the application is as high as 1800 s. Combine Figs. 1 and 2, we can conclude that the performance of the CPU-intensive application on physical sever node may has three different states according to the number of VMs running on the servers. These states are sound state, acceptable state and unfavorable state. Take

Fig. 2 The relations between processing time and VM number

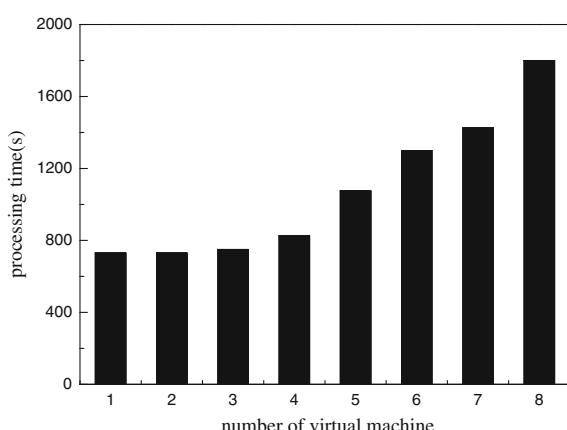
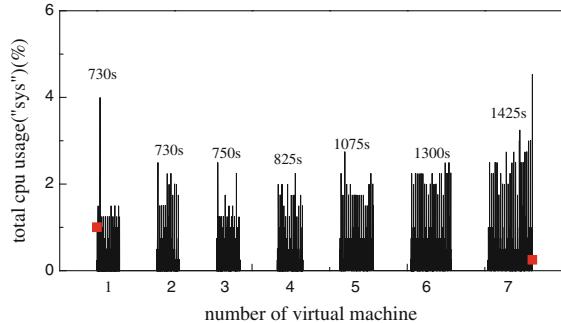


Fig. 3 The relations between cpu-sys and VM number



our experiments as examples, when the number of VM is in $[1, 4]$, it is the sound state. When the number of VM is in $(4, 5]$, it is the acceptable state. And when the number of VM is bigger than 5 or in $(5, \infty)$, it is the unfavorable state. Though the thresholds of the three states may be different when the configurations of server and CPU-intensive applications are different, the three states do exist. To a CPU-intensive application, it should be avoided to execute in unfavorable state.

From Fig. 3, we can easily get that the reflection time is always equal to the monitor time. So the increase of the number of VMs doesn't impact the performance of cpu-sys. And the consumption (sys) of the CPU resources under any circumstances is very small, and its value is generally less than 4 %. Obviously the sys consumed CPU resources very little for CPU-intensive applications. Therefore, the CPU-intensive applications are mainly consumed the cpu-usr resources, and other cpu-wai and cpu-sys, etc. will not be a significant change, and its values are very small.

Figure 4 presents the relation between recv traffic (flow rate in second) and number of VMs. Figure 5 shows the relation between number of VMs and send traffic (flow rate in second). From Figs. 4 and 5, it is easy to find out that the average receive traffic speed within the physical node is about 100 KB/s, and the flow speed of sending is less than 50 KB/s. As we use vpn to manage and control

Fig. 4 The relation between recv and VM number

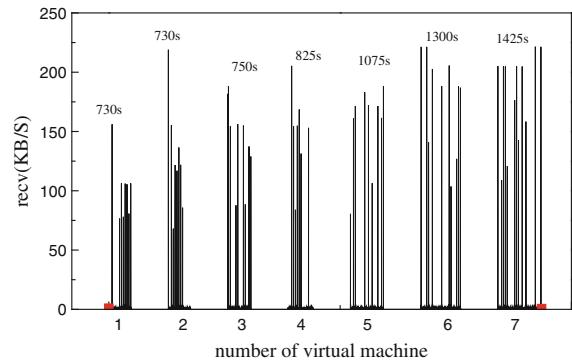
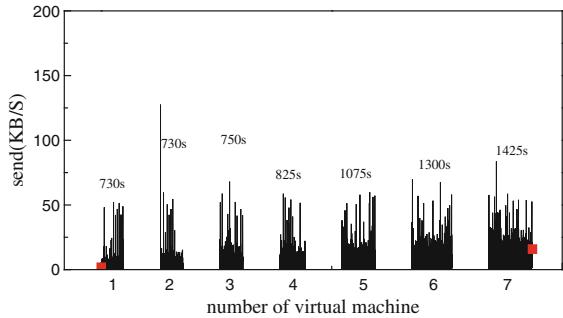


Fig. 5 The relations between send and VMs



virtual machines, it needs to send them some information to virtual machine or management server. At the same time, the virtual machine also needs some timing information to the system feedback, it will cause a small amount of network traffic. But compared with the speed of gigabit network bandwidth, cpu-intensive applications generate some neglected network traffic.

Figure 6 shows the relations between the number of VMs and disk IO write operations. It is easy to figure out from Fig. 7 that there is some disk writing speed. This is because that the monitor program should write some data to disk. But its write speed is generally not more than 150 KB/s, and its load is very light for disk. In the whole experiment process, CPU-intensive applications will not read on the

Fig. 6 The relation between IO write and VMs

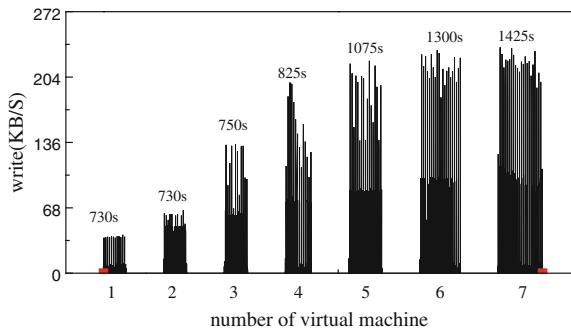
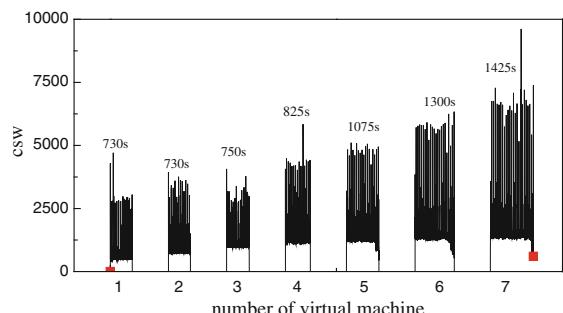


Fig. 7 The relations between the csw and VMs



disk, the disk IO read rate is mostly 0. Seen, CPU-intensive applications will not produce a lot of disk IO read and write operations.

Figure 7 shows the relations between change of context switches and VM numbers. The times of context switches per second will increase with the increase of the number of virtual machine. But the change is very small even it slightly with the increasing number of virtual machines. Because the CPU-intensive applications need a long time slice, so the change of context switches per second is relative little. And with the increase of CPU intensive applications, the process of CPU number increases, times of context switch per second increases accordingly. In the good performance stage, the value of context switch csw is always less than 5000 per second.

Figure 8 shows the relations between system load average of the physical node (loadavg) and the numbers of virtual machine. The time marked in the figure is the execution time used to process CPU-intensive applications. In the current system, when the number of virtual machines hosted on a physical host is less than 4 and $\text{loadavg} < 4$, processing time of the CPU-intensive applications does not change much, and it is about 730 s. At this case, the system is in the sound stage; When the V_m is in $[4, 5]$ and $\text{loadavg} > 4$, the processing time on the physical host increases less than 50 %. At this moment, the system is in acceptable performance stage. When the number of V_m is in $(5, +\infty)$ and $\text{loadavg} > 7$, the processing time is doubled to 1400 s. When $V_m > 8$, the processing time increases heavily. This time the performance of the system is in sharp decline stage.

Figure 9 shows the relations between the Memory used and VM numbers. It is easy to find out from Fig. 9 that when the number of VM is 0, the Memory used is about 1700 MB. With increasing a virtual machine, Memory used of system increases around 300 MB. Among them, the memory of each virtual machine is 1 GB. Visibly, each virtual machine only uses memory about one-third of its own, and the CPU-intensive applications have little demand on memory consumption. Therefore, memory will not become one of the bottlenecks on processing efficiency for CPU-intensive applications.

Fig. 8 The relation between the loadavg and VM number and the processing time

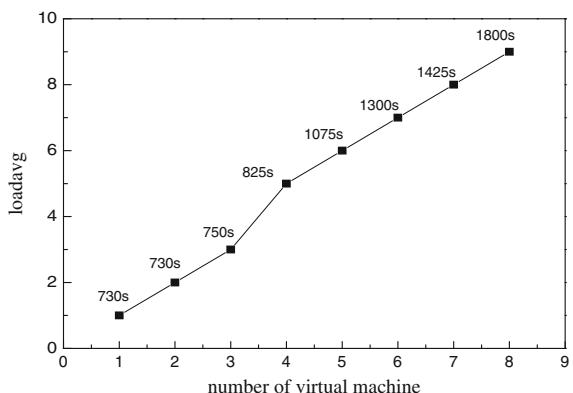
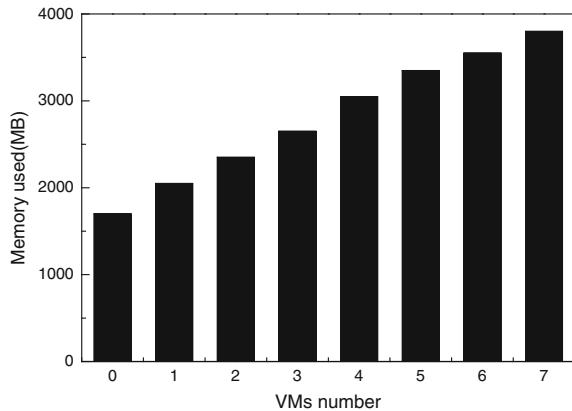


Fig. 9 The relation between the Memory used and VMs number



4.3 Model Validation

In order to prove the correctness and effectiveness of the model, we analyze the characteristics of the behaviors of different applications. And using the obtained results and the model put forward to judge whether the application is the CPU intensive application or not. The results are as shown in Table 2.

The crawler algorithm uses keywords and adopts depth first search strategy to search on the sina website. The streaming media server experiment is trying to test the video services provided by a streaming media server in a virtual machine. Database server experiment is a python scripts simulating the random operations of the database.

Considering too short-time application is not easy to obtain enough monitor data for the application prediction, and too long-time monitor for long-time application is needless. We adopt 10 min monitor for the application in each experiment. And in each of the experiment, we monitor collect parameters, state data of the system with different applications running on the cloud experimental platform. As the resource allocation and launching the virtual machine need some time, some interference may cause.

5 Conclusion

To be able to distinguish the different characteristics of the application and to be divided on the application type, we study the characteristics of the CPU-intensive applications, and have some qualitative and quantitative analysis on the CPU-intensive applications to establish a CPU-intensive application model. Referring to this model, we can discriminate the CPU-intensive applications. At the

Table 2 The results of model validation

Application type	cpu-usr (%)	cpu-sys (%)	csw	send+recv	read+write	CPU-intensive application or not	Judgement
Prime numbers	20	7	<5000	<300 KB/s	<100 KB/s	Yes	Correct
PI calculation	25	4.8	<5000	<300 KB/s	<100 KB/s	Yes	Correct
Matrix operations	21	3	<5000	<300 KB/s	<100 KB/s	Yes	Correct
Crawler algorithm	21	4	<5000	<300 KB/s	<100 KB/s	Yes	Correct
FTP server to download files	3	40	>10,000	>1 MB/s	>1 MB/s	No	Correct
FTP server to upload 2 GB files	5	42	>10,000	>1 MB/s	>1 MB/s	No	Correct
Streaming media server	7	4	<5000	>1 MB/s	>1 MB/s	No	Correct
Database server	15	36	>10,000	<300 KB/s	>100KB/s and <1MB/s	No	Correct

same time we can provide the standard of the optimization process and reduce energy consumption for CPU-intensive applications.

However, this model only considers the applications with relative long processing time. As how to model and predict the type of the applications with short processing time is the work that we will focus on in the next step.

Acknowledgments This work is supported by National Natural Science Foundation of China, (No. 61103054), Natural Science Foundation of Guangxi (No. 2013GXNSFAA019349), Foundation of Baoshan science and Technology Committee at Shanghai (12-B-16).

References

1. Dave, M., Dave, M., Shishodia, Y.S.: Cloud economics: vital force in structuring the future of cloud computing. In: 2014 International Conference on Computing for Sustainable Global Development (INDIACom), vol. 3, pp. 61–66 (2014)
2. Soares Boaventura, R., Yamanaka, K., Prado Oliveira, G.: Performance analysis of algorithms for virtualized environments on cloud computing. *Lat. Am. Trans. IEEE (Revista IEEE America Latina)* **12**, 792–797 (2014)
3. Lee, L.T., et al.: A dynamic resource management with energy saving mechanism for supporting cloud computing. *Int. J. Grid Distrib. Comp.* **6**(1), 67–76 (2013)
4. Maurya, K., Sinha, R.: Energy conscious dynamic provisioning of virtual machines using adaptive migration thresholds in cloud data center. *Int. J. Comput. Sci. Mob. Comput.* **3**(2), 74–82 (2013)
5. Moreno, I.S., et al.: Improved energy-efficiency in cloud datacenters with interference-aware virtual machine placement. In: 2013 IEEE Eleventh International Symposium on Autonomous Decentralized Systems (ISADS), vol. 3, pp. 1–8 (2013)
6. Fadika, Z., Govindaraju, M.: Delma: dynamically elastic mapreduce framework for cpu-intensive applications. In: 2011 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), vol. 5, pp. 454–463 (2011)
7. Takasaki, H., Mostafa, S.M., Kusakabe, S.: Applying eco-threading framework to memory-intensive hadoop applications. In: 2014 International Conference on Information Science and Applications (ICISA), IEEE, vol. 5, pp. 1–4 (2014)
8. Kuo, J.J., Yang, H.H., Tsai, M.J.: Optimal approximation algorithm of virtual machine placement for data latency minimization in cloud systems. In: International Conference on Computer Communications. IEEE (2014)
9. Pumma, S., Achalakul, T., Xiaorong, L.: Automatic VM allocation for scientific application. In: Proceedings of the 2012 IEEE 18th International Conference on Parallel and Distributed Systems. IEEE Computer Society (2012)
10. Alasaad, A., et al.: Innovative schemes for resource allocation in the cloud for media streaming applications. *IEEE Trans. Parallel Distrib. Syst.* **PP**(99), 1–1 (2014)
11. Wu, Y., et al.: NO²: speeding up parallel processing of massive compute-intensive tasks. *IEEE Trans. Comput.* **10**(63), 2487–2499 (2013)
12. Neumann, R., et al.: Caching highly compute-intensive cloud applications: an approach to balancing cost with performance. In: Software Measurement, 2011 Joint Conference of the 21st Int'l Workshop on and 6th Int'l Conference on Software Process and Product Measurement (IWSM-MENSURA). IEEE (2011)
13. Tan, Y.M., Zeng, G.S., Wang, W.: Policy of energy optimal management for cloud computing platform with stochastic tasks. *J. Softw.* **23**(2), 266–278 (2012)

A Study of Effects of UTAUT-Based Factors on Acceptance of Smart Health Care Services

Yoo-Jin Moon and Young-Ho Hwang

Abstract The paper analyzes factors influencing acceptance of smart health care services based on UTAUT. The result of T-test indicates that users with experiences of smart health care services have a higher degree of effort expectancy and intention to use the services than those without their experiences. And, the study shows that social influence of the services positively affects user intention to use the services, that performance expectancy is positively correlated with user intention to use the services, and that perceived enjoyment positively affects potential intention to use the services. According to the results, companies need to increase performance expectancy, to intensify word-of-mouth marketing, and to improve enjoyment and attractiveness of the services.

Keywords Smart health care • User intention to use • Effort expectancy • Social influence • Performance expectancy • Perceived enjoyment

1 Introduction

A smart health care system can be connected with a collection of health care applications within a smartphone. This health care system uses advanced technology of terabyte memories by synchronizing with cloud systems, to intellectually monitor and manage patients' conditions anytime anywhere, and to provide real-time cus-

This paper has revised and extended a version of paper published in World IT Congress 2015.

Y.-J. Moon (✉)

Hankuk University of Foreign Studies, Yongin-gu, Gyeonggi-do, South Korea
e-mail: yjmoon@hufs.ac.kr

Y.-H. Hwang

Kunsan National University, Gunsan, South Korea
e-mail: yhwang@kunsan.ac.kr

tomized services by analyzing patients' personal and disease information. This study analyzes factors influencing acceptance of smart health care services based on UTAUT (Unified Theory of Acceptance and Use of Technology) [1].

2 Theoretical Background

Various smart health care services usually include diagnosing and managing one's health condition by checking the amount of exercise, the blood sugar level, the electrocardiogram(ECG), and the heart rate using mobile devices connected with health measurement devices or smartphone applications. Also, various health bands can share and save data with smartphone applications about the amount of calories burnt, the user psychological state, and distance traveled etc. Other examples include a smart pace counter, a smart drunkometer, a smart diet coach, a smart toothbrush, a smart fork, a smart pill, smart contact lenses, a smart ECG measurement, and wearable smart skin patches. In addition, by blood pressure gauges, glucose meters and so on, smart home care services monitor high blood pressure, diabetes, and other conditions through gateway connected to internet. Using home care services is gradually on the rise [2].

The smart health care service is severely related with the advanced technology, which is the core subject of the information technology acceptance theory [3]. The information technology acceptance theory is related to decision making of human's willingness to accept new technology. So studies for the technology advancement include variables related to human attitude or intention. This study sets the research model using UTAUT which integrates the existing theories.

UTAUT proposed by Venkatesh et al. includes three variables (performance expectancy, effort expectancy, and social influence) that affect intention to use, one variable (facilitating conditions) that affects usage behavior, and four controlled variables (sex, age, experience, and voluntariness) [4]. This study includes additional exogenous variables such as personal innovativeness and perceived enjoyment, which become crucial when applying the smart health care service technology to UTAUT.

3 The Research Model and Hypotheses

3.1 Samples

To define the demographics of respondents, frequency analysis and descriptive statistics analysis were performed based on a total of 126 samples. The samples chosen for this study were two groups of university students at A University in Seoul and B University in the western part of Korea. We used a non-probability

sampling. Those questioned completed self-reported questionnaires and voluntarily participated in responding the questionnaires. The demographic statistics of the participants showed that the samples consisted of 67.5 % male and 32.5 % female, and 23.8 % experienced users and 76.2 % non-experienced.

3.2 Research Model

This study extracted variables such as effort expectancy, social influence, performance expectancy and facilitating conditions from the model proposed by Venkatesh et al., as important factors affecting intention to use smart health care services. Also, this study added additional variables—personal innovativeness and perceived enjoyment proposed by Agarwal and Karahanna using Davis' TAM (Technology Acceptance Model) [5]. The research model is illustrated in Fig. 1.

3.3 Hypotheses Setting

1. Demographic Variables

Hypothesis 1-1: Gender differences in personal innovativeness, effort expectancy, social influence, performance expectancy, perceived enjoyment, facilitating conditions, and intention to use would exist.

Hypothesis 1-2: Differences in service experience would show different levels of personal innovativeness, effort expectancy, social influence, performance expectancy, perceived enjoyment, facilitating conditions and intention to use.

2. UTAUT Variables

Hypothesis 2-1: Effort expectancy would have a positive impact on user intention to use smart health care services.

Hypothesis 2-2: Social influence would have a positive impact on user intention to use smart health care services.

Hypothesis 2-3: Performance expectancy would have a positive impact on user intention to use smart health care services.

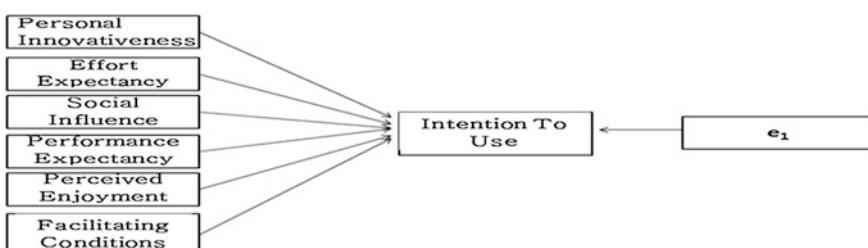


Fig. 1 The research model

Hypothesis 2-4: Facilitating conditions would have a positive impact on user intention to use smart health care services.

3. Additional Variables

Hypothesis 3-1: Personal innovativeness would have a positive impact on user intention to use smart health care services.

Hypothesis 3-2: Perceived enjoyment would have a positive impact on user intention to use smart health care services.

4 Hypotheses Verification and Empirical Analysis

4.1 Verification of the Research Model

This study verified reliability and validity of the model using collected data ($n = 126$).

1. Reliability Analysis

This study tested and analyzed reliability between multi-item scales on 22 measurement variables using SPSS 18 program. The reliability analysis showed that all of Cronbach α coefficients were above .7 and the reliability was secured.

2. Validity Analysis

The study performed the exploratory factor analysis about items of the questionnaire measuring constructs of the research model. The factor extraction method was based on principal component analysis and Varimax rotation with Kaiser-normalization [6].

Results of the exploratory factor analysis showed that all seven initially intended factors including the dependent variable were extracted; factor 1 'personal innovativeness', factor 2 'perceived enjoyment', factor 3 'social influence', factor 4 'performance expectancy', factor 5 'intention to use', factor 6 'facilitating conditions', and factor 7 'effort expectancy'. Each factor showed that the Eigen value was above 1 and the rate of cumulative variance showed 82.8 % of total variance. The study found that multi-collinearity did not exist in the items of the questionnaire.

4.2 Hypotheses Verification

1. T-test Verification regarding gender and user experience

Result of the T-test verification on samples showed that no statistically significant difference between genders existed, so we rejected Hypothesis 1-1. And we partially accepted Hypothesis 1-2 because among six independent variables only two variables of effort expectancy and intention to use showed statistically

Table 1 Results of T-test between genders

	Levine's equal variance test		T-test on identity of mean	
	F	α	t	α (two-tail)
Personal innovativeness	4.508	.036	1.043	.299
Effort expectancy	.330	.567	.780	.437
Social influence	6.772	.011	-1.317	.190
Performance expectancy	2.457	.120	-.292	.771
Perceived enjoyment	8.592	.004	.401	.689
Facilitating conditions	1.972	.163	-.020	.984
Intention to use	5.935	.016	-1.015	.312

significant differences between users and non-users with $t = 2.224$, $\alpha = .028$ and $t = 3.472$, $\alpha = .001$ respectively. Table 1 and 2 illustrates results of the T-test. This study used multiple regression analysis by setting intention to use smart health care services as a dependent variable and other six variables (personal innovativeness, performance expectancy, social influence, performance expectancy, perceived enjoyment, facilitating conditions) as independent variables. Results of the multiple regression analysis showed that three of the six hypotheses suggested turned out to be statistically significant. Table 3 illustrates results of the multiple-regression analysis.

2. Hypotheses Verification Using Multiple Regression Analysis

This study used multiple regression analysis by setting intention to use smart health care services as a dependent variable and other six variables (personal

Table 2 Results of T-test between users and non-users

	Levine's equal variance test		T-test on identity of mean	
	F	α	t	α (two-tail)
Personal innovativeness	.134	.715	.930	.354
Effort expectancy	.809	.370	2.224	.028
Social influence	.232	.631	.816	.416
Performance expectancy	.742	.391	1.927	.056
Perceived enjoyment	.283	.596	1.924	.057
Facilitating conditions	.112	.738	.826	.410
Intention to use	.012	.912	3.472	.001

Table 3 Results of multiple-regression analysis

Dependent variable	Independent variables	B	Standard error	β	t	α	Accept/reject
Intention to use	Constant	-.81	.397		-2.042	.043	
	Personal innovativeness	.089	.071	.084	1.253	.213	Reject
	Effort expectancy	.042	.101	.036	.417	.677	Reject
	Social influence	.172	.091	.167	1.911	.050	Accept
	Performance expectancy	.343	.101	.289	3.412	.001	Accept
	Perceived enjoyment	.319	.098	.288	3.266	.001	Accept
	Facilitating conditions	.098	.077	.094	1.268	.207	Reject
R^2 .614							
F-value 31.579							

innovativeness, performance expectancy, social influence, performance expectancy, perceived enjoyment, facilitating conditions) as independent variables. Results of the multiple regression analysis showed that three of the six hypotheses suggested turned out to be statistically significant. Table 3 illustrates results of the multiple-regression analysis.

Hypothesis 2-1 was dismissed because it indicated that effort expectancy did not have a statistical significance on user intention to use smart health care services at $\alpha = .05$ and $\beta = .036$, $t = .417$. This result did not support the preexisting studies [4, 7]; it indicated that ease of use, usefulness and ease of acquiring results had nothing to do with user intention to use.

Hypothesis 2-2 was accepted because it showed that social influence did have a statistical significance on user intention to use smart health care services at the $\alpha = .05$ and $\beta = .167$, $t = 1.911$. This result supported the preexisting studies [4]; it indicated that user intention to use smart health care services was positively affected by recognition of surrounding people, influential people, and important people who believed that I should use smart health care services.

Hypothesis 2-3 was accepted because it showed that performance expectancy had a statistical significance on user intention to use smart health care services at $\alpha = .05$, $\beta = .289$, and $t = 3.412$. This result supported the preexisting studies [4, 7, 8]; it meant that user intention to use smart health care services was positively affected by helping health enhancement, time saving and health condition improvement through the smart health care services.

Hypothesis 2-4 was dismissed because it showed that facilitating conditions did not have a statistical significance on user intention to use smart health care services at $\alpha = .05$, $\beta = .094$, and $t = 1.268$. This result did not support the preexisting studies [4]; it indicated that quick after-service support,

compatibility with smart-phones, and advice from experts had nothing to do with service usage intention.

Hypothesis 2-5 was dismissed because it showed that personal innovativeness did not have a statistical significance on user intention to use smart health care services at $\alpha = .05$, $\beta = .084$, and $t = 1.253$. This result did not support the preexisting studies [6]; it indicated that tendency of personal innovativeness had nothing to do with user intention to use.

Hypothesis 2-6 was accepted because it showed that perceived enjoyment had a statistical significance on user intention to use smart health care services at $\alpha = .05$, $\beta = .288$, and $t = 3.266$. This result supported the preexisting studies [6]; it indicated that enjoyment, attractiveness and interest of smart health care services had a positive effect on user intention to use.

5 Conclusions

This study was conducted to analyze and verify which factors affect user intention to use smart health care services among personal innovativeness, effort expectancy, social influence, performance expectancy, perceived enjoyment and facilitating conditions, by surveying university students who may potentially use the services.

Results of the T-test verification and the multiple regression analysis are as follows. First, T-test verification between gender differences did not show a statistically significant difference in seven variables. Second, T-test verification between the experienced and the non-experienced with smart health care services showed a statistically significant difference in effort expectancy and user intention to use. Third, effort expectancy did not show a statistically significant difference in user intention to use smart health care services. Fourth, social influence showed a statistically significant difference in user intention to use smart health care services. Fifth, performance expectancy showed a statistically significant difference in user intention to use smart health care services. Sixth, facilitating conditions did not show a statistically significant difference in user intention to use smart health care services. Seventh, personal innovativeness did not show a statistically significant difference in user intention to use smart health care services. Eighth, perceived enjoyment showed a statistically significant difference in user intention to use smart health care services.

These results imply that use of smart health care services are influenced by health enhancement, recommendation from family and colleagues, and perceived enjoyment more than by ease of use, compatibility and innovative tendency. Consequently, the results indicate that in order to dominate the smart health care service market, companies need to increase performance expectancy, to intensify word-of-mouth marketing, and to improve enjoyment and attractiveness of smart health care services.

Acknowledgment This work was supported by Hankuk University of Foreign Studies Research Fund of 2014.

References

1. Lescevica, M., Ginters, E., Mazza, R.: Unified theory of acceptance and use of technology (UTAUT) for market analysis FP7 CHOReOS products. *Procedia Comp. Sci.* **26**, 51–68 (2013)
2. Peck, J.L., Stanton, M., Reynolds, G.E.: Smartphone preventive health care: parental use of immunization reminder system. *J. Pediatr. Health Care* **28**(1), 35–42 (2014)
3. Sternad, S., Bobek, S.: Impacts of TAM-based external factors on ERP acceptance. *Procedia Technol.* **9**, 33–43 (2013)
4. Jung, C., Namn, S.: Cloud computing acceptance at individual level based on extended UTAUT. *J. Digit. Convergence* **12**(1), 287–294 (2014)
5. Agarwal, R., Karahanna, E.: Time flies when you're having fun: cognitive absorption and beliefs about information technology usage. *MIS Q.* **24**(4), 665–694 (2000)
6. Hwang, Y., Hwang, W., Moon, Y.: A study of factors influencing intra-organizational potential users' intention to use N-screen service. *Korean Comp. Gov. Rev.* **16**(1), 89–114 (2012)
7. Escobar-Rodríguez, T., Carvajal-Trujillo, E.: Online purchasing tickets for low cost carriers: an application of the unified theory of acceptance and use of technology (UTAUT) model. *Tour. Manag.* **43**, 70–88 (2014)
8. Oliveira, T., Faria, M., Thomas, M.A., Popovič, A.: Extending the understanding of mobile banking adoption: when UTAUT meets TTF and ITM. *Int. J. Inf. Manage.* **34**(5), 689–703 (2014)

Detection and Recognition of Road Markings for Advanced Driver Assistance System

JongBae Kim

Abstract This paper proposes a method for detecting direction indicators marked on road surfaces for safe driving support. In the proposed method, images are received from a vehicle's black box, and a method for template matching is used on such direction indicators to detect the indicator area. By detecting the Maximally Stable Extremal Regions (MSER), the matching method is used to detect the road indicator area after the areas where road indicator candidate regions and binary code result images overlap are detected through the multi-level threshold template. The results of the experiment conducted in an actual vehicle driving environment show that, from the total of 270 frames that include indicators, each frame requires approximately 0.34 s, and a minimum of 83 % detection rate is provided.

Keywords ADAS · MSER · Template matching · Car black-box

1 Introduction

As we have recently entered a rapidly aging society, the number of elderly drivers has increased. Furthermore, according to a 2013 traffic accident statistics analysis [1], there is an increasing trend of accidents between vehicles, and between vehicles and pedestrians. Many techniques are being developed to prevent these types of traffic accidents, but such techniques are mostly aimed at avoiding inter-vehicle collisions, or at warning about lane departures [2, 3]. Safe-driving support technologies for vehicles are continuously being developed [4], and driver awareness is also being improved greatly. Under these conditions, car black boxes are being installed to store information on the surrounding circumstances of a vehicle in real time. The United States has exhibited a positive attitude to passing a bill for mandatory black box installation. In domestic major cities, company taxies are

J. Kim (✉)

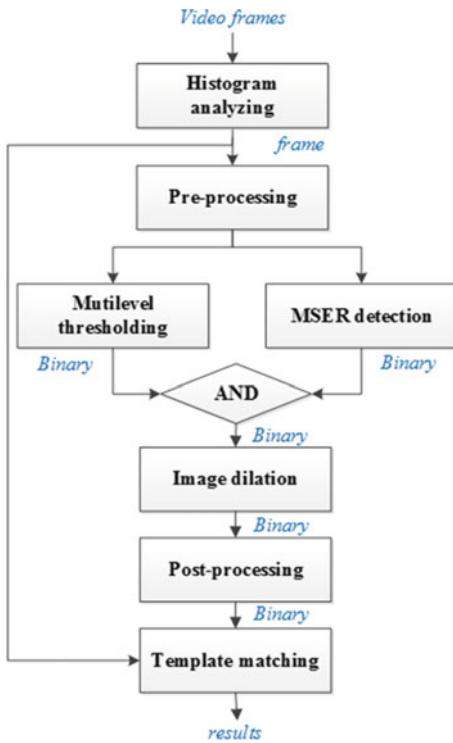
Department of Computer Engineering, Seoul Digital University, Seoul, South Korea
e-mail: jbkим@sdu.ac.kr

obligated to install black boxes. In public transportation and passenger ships, black box installation is strongly encouraged. The fundamental reason so many black boxes are being installed on means of transportation is that such boxes are used to determine accurately the causes of traffic accidents. In addition, black box installation has the advantage of discounts on car insurance, and of securing legal evidence. Presently, however, black boxes are used for situational assessments post-accident. In fact, black boxes have been categorized as vehicle IT equipment not helpful for safe-driving support. In other words, black boxes are not perceived as devices with built-in pre-processing functions for safe-driving support that can help drivers. Of the five senses, sight is the most used when driving. Thus, drivers operate vehicles using recognition and judgment based on visual information. However, because of the side effects of physical aging, farsightedness and the inability to react quickly are some of the causes of traffic accidents among elderly drivers. As technology aimed at counteracting such side effects, black boxes installed in most domestic vehicles can be more relevant for safe-driving support as a third-party visual information. To summarize, with the installation of image processors in black boxes, it is possible to provide drivers with preprocessed information.

2 Proposed Method

This study proposes a method for detecting the direction indicator region printed on road surfaces by inputting images from vehicle black boxes. Given that small memory, sound power, and real-time processes are required for small black boxes, a relatively simple, but effective, method was designed. The results from analyzing the characteristics of direction indicators marked on road surfaces indicate that corresponding indicators are made from the mono-color white, and they are located on a relatively black background. Thus, road indicators have the characteristic of appearing as a single color in a connected single range. Detecting connected pixels formed in bright mono-color in an input image is required. In order to do this, for the proposed method, the Maximally Stable Extremal Regions (MSER) are detected, and template matching is used to detect the surface indicator area. Figure 1 depicts the flowchart for the proposed method. Using video sequence input, the change in brightness of the road surface area is analyzed in each frame to determine whether the surface indicator is included. In addition, after undergoing pre-processing, such as noise removal and emphasizing pixel brightness in pertinent frames, a binary image is output using the MSER detection process. Subsequently, areas adjacent to the ROI region borderlines are removed through post-processing, and small areas are removed through connected component analysis. After matching the final binary rectangular areas with the templates of indicators marked on road surfaces, the most similar indicator area of the road surface is detected.

Fig. 1 Flowchart for proposed method



2.1 Histogram Analyzing

For the area that excludes the upper and lower section of an input image, change in pixel brightness at a specific area (240×810 pixel) is used to determine whether the indicator region is included in the pertinent frame. If various methods are used for pre-processing, better performance could be obtained, but difficulties might arise because of running time complexity. However, there is a relatively simple method: based on road surfaces with surface indicators as the background, intensity distribution can be used with the most differentiable colors. When brightness and histogram change exceed a given threshold, further processing is performed. If the brightness histogram change rate and average surface floor brightness change is greater than 30 %, we consider information similar to the road surface indicator to be included in the corresponding frame. The road surface indicator candidate ROI region is detected in the input image (Fig. 2).

Fig. 2 Input and candidate ROI image



2.2 Multilevel Thresholding

The most basic method for detecting the indicator area from the road surface indicator candidate ROI is binarization through thresholding. However, because of various changes in road environments, using the Otsu adaptive thresholding method [5] could cause problems of image over-segmentation. Thus, for the proposed method, multi-thresholding is used to produce a binary image that can differentiate the road surface from the indicator area. Figure 3 is the intensity histogram of the candidate ROI. The dotted line is the threshold value 85 calculated using adaptive thresholding, and the solid lines are calculated two-level thresholding: the lower limit is 78 and the upper limit is 141. In the proposed method, instead of using adaptive thresholding, the two-level thresholding method is applied, and the upper

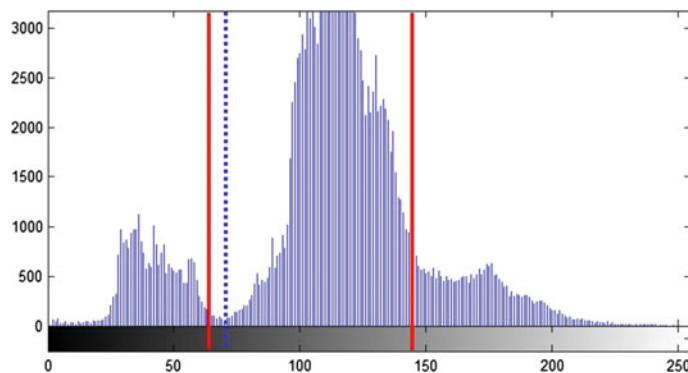


Fig. 3 Intensity histogram of candidate ROI

limit is used as the threshold value for binarization. The results of our experiment indicate that, for road images, changes in the threshold values are required based on surrounding illumination. Thus, during daytime, road surface brightness is high overall because of sunlight reflection. In cloudy weather, road surface brightness is mostly low. Therefore, for the proposed method, the input frame average overall brightness is calculated, and level one and two threshold values are used selectively.

2.3 MSER Detection

For the MSER method, pixels with similar color information form groups based on color information differentiated from adjacent pixels, and each group region that satisfies the critical condition is detected [6]. At this stage, the size of the group region is set to minimum 300 and maximum 7000; the degree of change in brightness is selected at 0.6; and the change in threshold is selected at 2. Figure 4 is the result of detecting MSER in the road surface indicator candidate ROI image of Fig. 2, and each region is expressed in a different color. The brightness change in the specific MSER has a value between zero and one. A lower value indicates a pixel region with lower brightness change.

2.4 Post-processing and Image Dilation

Through post-processing, the detected multi-level threshold common region and MSER binary images are selected. From the selected common binary images, the indicator candidate regions are detected through the connected component analysis of each pixel. In addition, using pixel connected component analysis, areas with fewer than 100 pixels and where a traffic lane extends over the ROI region boundary line are removed. Moreover, the connection component of regions labeled as single character regions using morphology, and of regions that are discontinued because of brightness change, is reinforced. The morphology operation uses an arithmetic operation for lines with 15 pixels and 90-degree rotation. Binary images and connected areas derived from the morphology operation are shown in Fig. 5d.

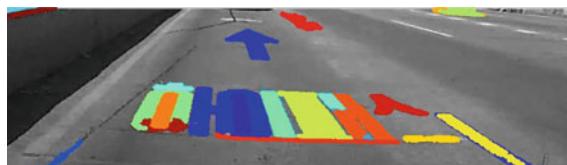


Fig. 4 MSER detection results

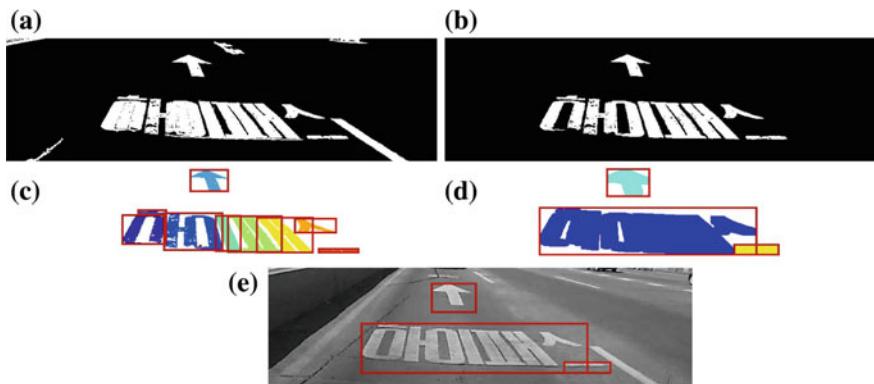


Fig. 5 Results of post-processing and image dilation. **a** MSER detection. **b** Overlap region of (a) and thresholding image. **c** Labeling result. **d** Labeling result after image dilation. **e** Detected candidate regions

Figure 5 shows the post-processing results and the results of detecting the candidate region from the road surface indicator through the image dilation stage.

2.5 Template Matching

For the final road surface indicator detection, the template with the Euclidean minimum distance is detected as the corresponding indicator using template matching for each indicator. Samples of road surface direction indicators are displayed in Fig. 6a, and the average image of the direction indicators is shown in Fig. 6b. For the indicator-matching template, the image in Fig. 6b is used to identify the region with the minimum Euclidean value as the corresponding indicator region.



Fig. 6 Sample images for template matching. **a** Sample images of road surface indicators. **b** Average images of each road surface indicators

3 Results and Future Works

In order to test the accuracy of the proposed method, the results of manually detecting the road surface indicators were compared to those obtained from the proposed method. From a total of 270 experiment images, 83 % showed detection accuracy, and the average processing time per frame was approximately 0.342 s. MSER detection and the template matching process required the longest processing time; for the remaining image processing, only the connected pixel areas in the 240×810 -pixel binary images were processed, thus minimizing processing time. The results indicate that most detection errors occur because of low brightness levels or road surface contamination. In addition, there were cases where the characters marked on the road surface were misrecognized as direction indicators. By detecting the most appropriate indicator region through template matching, in the case where two or more direction indicators are the same, only one is recognized. The results of testing the proposed method with daytime road images showed relatively good performance, but in order to apply the method to various road environments (at night or in the rain), a more detailed experiment is required. However, for small memory and minimization of power consumption and heat generation in small car black boxes, research on the most effective processing is required. For research on the recognition of characters that exist on road surfaces, a study will be performed on deduction-based robust detection for environment changes in the future.

Acknowledgments This research is supported by Basic Science Research Program through the NRF of Korea funded by the Ministry of Education, Science and Technology (2010-0021071) in 2014, and this paper is an extended version of work published in [7].

References

1. 2014 traffic accident statistics report, National Police Agency of Korea, <http://taas.koroad.or.kr> (2014)
2. Fujimura, K., Konoma, T., Kamijo, S.: Vehicle infrastructure integration system using vision sensors to prevent accidents in traffic flow. *Intell. Transp. Syst.* **5**(1), 11–20 (2011)
3. Ai, M., Falcone, P., Olsson, C., Shoberg, J.: Predictive prevention of loss of vehicle control for roadway departure avoidance. *IEEE Trans. ITS* **14**(1), 56–68 (2013)
4. Marfia, G., Roccati, M., Amorose, A., Pau, G.: Safe driving in LA: report from the greatest intervehicular accident detection test ever. *IEEE Trans. Veh. Technol.* **62**(2), 522–535 (2013)
5. Otsu, N.: A threshold selection method from gray-level histograms, *IEEE Trans. Sys. Man Cyber.* **9**(1), 62–66 (1979)
6. Donoser, M., Bischof, H.: Efficient maximally stable extremal regions (MSER) tracking. In: *IEEE Conference on CVPR*, pp. 553–560 (2006)
7. Kim, J.B.: Detection of road direction indicators using maximally stable extremal regions and template matching for vehicle black box system. In: *Proceedings of world IT congress*, p. 34 (2015)

Color and Depth Image Correspondence for Kinect v2

Changhee Kim, Seokmin Yun, Seung-Won Jung and Chee Sun Won

Abstract Kinect v2, a new version of Kinect sensor, provides RGB, IR (Infrared) and depth images like its predecessor Kinect v1. However, the depth measurement mechanism and the image resolutions of the Kinect v2 are different from those of Kinect v1, which requires a new transformation matrix for the camera calibration of Kinect v2. In this paper, we correct the radial distortion of the RGB camera and find the transformation matrix for the correspondence between the RGB and depth image of the Kinect v2. Experimental results show that our method yields accurate correspondence between the RGB and depth images.

Keywords Kinect v2 · Registration · Camera calibration

1 Introduction

Kinect from Microsoft is a very popular depth sensor. Thanks to its low price, it has been widely adopted for various computer vision applications, including 3D reconstruction, object recognition, and object tracking. To adopt the Kinect for more sophisticated computer vision applications, however, RGB and depth sensors

C. Kim · S. Yun · C.S. Won (✉)

Department of Electronics and Electrical Engineering, Dongguk University, Seoul, Korea
e-mail: cswon@dongguk.edu

C. Kim
e-mail: Kimchda210@naver.com

S. Yun
e-mail: smyun@dongguk.edu

S.-W. Jung
Department of Multimedia Engineering, Dongguk University,
30, Pildong-ro 1gil, Jung-gu, Seoul 100-715, Korea
e-mail: swjung83@dongguk.edu

Table 1 Comparative specifications of Kinect v1 and Kinect v2

	Kinect v1	Kinect v2
Resolution of color image	640×480 (pixel)	1920×1080 (pixel)
Resolution of IR and depth image	320×240 (pixel)	512×424 (pixel)
Field of view of color image	$62^\circ \times 48.6^\circ$	$84.1^\circ \times 53.8^\circ$
Field of view of IR and depth image	$57.5^\circ \times 43.5^\circ$	$70.6^\circ \times 60^\circ$
Maximum skeletal tracking	2	6
Method of depth measurement	Light coding	Time of Flight
Working range	0.8–3.5 m	0.5–8 m

of Kinect need to be precisely calibrated. The calibration for the first version of the Kinect (Kinect v1) has been proposed [1].

Recently, the second version of Kinect (Kinect v2) has been released. The Kinect v2 adopts a different sensing method for the depth measurement and provides higher image resolutions. Specifically, the Kinect v2 uses ‘Time of Flight’ (TOF) method instead of ‘light coding’ of the Kinect v1 for the depth measurements. Also, the image resolutions for both RGB and depth of the Kinect v2 are higher than those of Kinect v1. See Table 1 for more comparisons between Kinect v1 and v2 [2]. Because of the differences listed in Table 1, the calibration parameters developed for the Kinect v1 are not applicable for the Kinect v2 sensors. In this paper, we correct the image distortions caused by the lens of the Kinect v2 and provide its calibration matrix for the correspondence between the RGB, depth, and IR (Infrared) images.

This paper is composed of the following sections. In Sect. 2, the radial distortions of the Kinect are corrected. The holes in the depth image are filled and the correspondence matrix between the RGB and the depth images is provided. In Sect. 3 we show the accuracy of our method by comparing with the results by SDK tools.

2 Kinect v2 Calibration

2.1 Distortion Correction

Camera lens causes image distortions, where Kinect v2 is not an exception. Both IR and color cameras have distortions. Calibration tools such as Camera Calibration Toolbox for Matlab are available to determine the intrinsic parameters of the cameras. A set of checkerboard images from both color and IR cameras are used to identify the corners of the checkerboard pattern in RGB and IR images. Then, by solving the equations from the correspondences of the corner points and by using the non-linear optimization technique to reduce the reprojection errors, the intrinsic camera parameters such as focal length, principal points, and skew can be

Table 2 Camera parameters for Kinect v2

Parameter	Camera	
	RGB	IR
Focal length (f_c)	$[1053.622 \ 1047.508] \pm [4.6884 \ 4.5323]$	$[376.6518 \ 371.4936] \pm [1.8265 \ 1.8015]$
Principal points (c_c)	$[950.3941 \ 527.3442]$	$[265.5583 \ 206.6131]$
Skew (α_c)	$[0.000] \pm [0.000] \rightarrow \text{angle of pixel} = 90.00^\circ$	$[0.000] \pm [0.000] \rightarrow \text{angle of pixel} = 90.00^\circ$
Distortion (k_c)	$[0.0042 \ -0.0019 \ -0.0038 \ -0.00260] \pm [0.0038 \ 0.0033 \ 0.0007 \ 0.00080]$	$[-0.0094 \ -0.0431 \ 0.0004 \ -0.00030] \pm [0.0094 \ 0.0144 \ 0.0015 \ 0.00170]$

determined [3]. Then, the radial distortions of the color camera can be corrected. The results of our calibration parameters for the Kinect v2 are listed in Table 2.

Zoomed images before and after the correction of the Kinect v2 distortions are shown in Fig. 1. As one can see in Fig. 1a, c, the images captured by the Kinect v2 sensor suffer from the radial distortions. These distortions are corrected by the calibration parameters in Table 2 (see Fig. 1b, d).

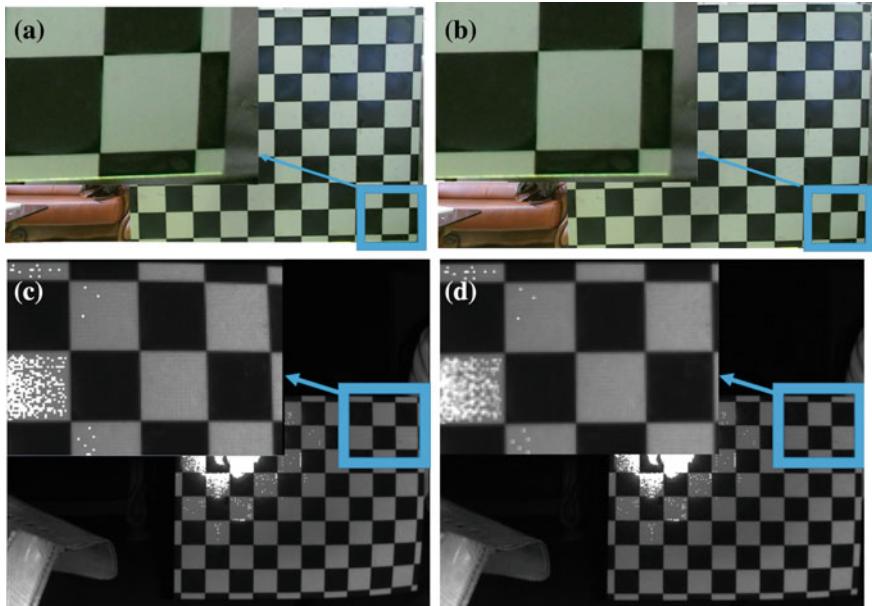


Fig. 1 Images with distortions **a** and **c** and after distortion correction **b** and **d**. *Above* images are taken by color camera and *below* ones are IR images

2.2 Correspondence of Color and IR (Depth) Images

The correspondence between the RGB sensor and the IR (depth) sensor can be done by a transformation matrix between them. Since the depth image is generated by the IR sensor, we can use either the depth image or the IR image for the registration. Here, since the depth image cannot show the pattern on the planer checkerboard, our registration is based on the checkerboard images of the RGB and the IR images.

The projective matrix which converts the RGB image coordinate (x, y) into the IR image coordinate(X, Y) is given as follows [4].

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = T \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (1)$$

Matrix equation (1) with eight unknown parameter values can be rewritten as follows

$$X = a_1x + a_2y + a_3 - a_7xX - a_8yX \quad (2)$$

$$Y = a_4x + a_5y + a_6 - a_7xY - a_8yY \quad (3)$$

Four corresponding pairs of (x, y) and (X, Y) must be known to solve Eqs. (2) and (3). If we use more than four pairs, the parameters can be found using a least square method like the direct linear transform. We use the checkerboard to find the pairs of corresponding points between the RGB and IR images as shown in Fig. 2. The corresponding points are selected manually. Using these points, we calculate the eight parameters for the transformation matrix. Specifically, we use 515 pairs of points extracted from indoor images of limited distances to calculate Eqs. (2) and (3), and we get the transformation matrix as shown in (4). So, the matrix in (4) is good for the calibration of Kinect v2 images captured in the near distances, say up to 3 m.

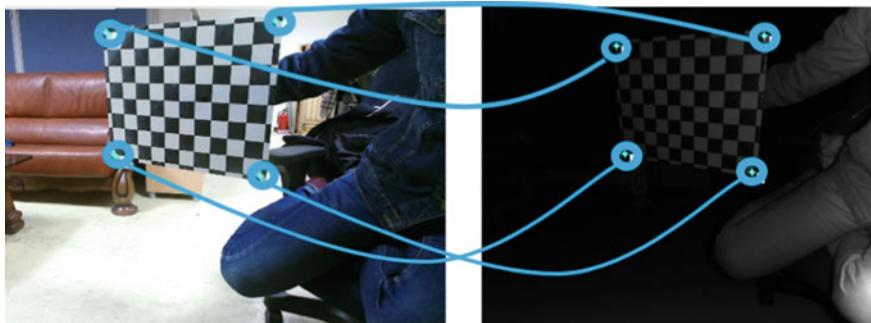


Fig. 2 Corresponding points in RGB and IR images



Fig. 3 *Above* original color and IR images. *Below* registered color and IR images obtained by the transformation and cropping

$$T = \begin{bmatrix} 0.3584 & 0.0089 & 0.0000 \\ 0.0031 & 0.3531 & 0.0001 \\ -101.5934 & 13.6311 & 0.9914 \end{bmatrix} \quad (4)$$

Note that the size of RGB image is bigger than the IR image in Kinect v2. Also, they have different field of views (FOV). Therefore, color images need to be cropped after registration. Figure 3 shows an example of registration results. Above images are raw images and below ones are the results after the registration and cropping. Color image is cropped at both sides to fit the FOV of the IR image. The IR image is also cropped at top-bottom parts because the top-bottom FOV of color image is bigger than IR image. After the calibration and cropping the size of IR image and color image is changed to 512×360 . Original size of color is 1920×1080 and that of IR is 512×424 .

2.3 Hole Filling

Since the depth image is captured by the IR sensor in Kinect v2, the registration between the RGB and the IR image automatically yields the registration among the RGB, IR, and depth images. So, after the registration and cropping, we can generate

the pixel-by-pixel correspondence images of RGB, IR, and depth from the Kinect v2 sensor. Here, to make a perfect correspondence, the holes in the depth image are to be filled.

As in Kinect v1, Kinect v2 has depth holes with missing depth measurements. Although the holes of the Kinect v2 along the object boundary are usually thinner than those of the Kinect v1, the holes near the object boundary in the Kinect v2 depth image can be still filled by the method of Kinect v1. In particular, since the RGB image is already aligned with the depth image, we can exploit the edge information in the RGB image to determine the direction of the hole filling in the depth image [5].

3 Results

Figure 4a shows the superimposed images after the correction of the radial distortion and the registration. Compared to the result of Fig. 4b obtained by the software development kit (SDK) [2] function of “*MapColorFrameToDepthSpace*”, the mismatch errors of our method are much smaller than those of the SDK method. Specifically, the checkerboard lines in the left image (Fig. 4a) are well aligned compared to the checkerboard lines in right image (Fig. 4b).

We also calculate the pixel distances of correspondence points of checkerboard in the registered color and IR images to compare the pixel position errors. The results using three sets of color and IR images are shown in Table 3. Also there are images used for calculating pixel position errors in Fig. 5. The average distance of the SDK method is 26.2354 and that of our result is 5.5121. We can notice that our method is about five times more accurate than the SDK method in terms of the pixel mismatches.

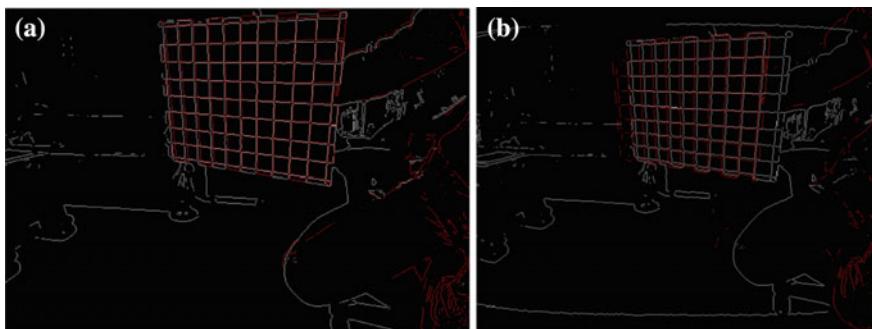
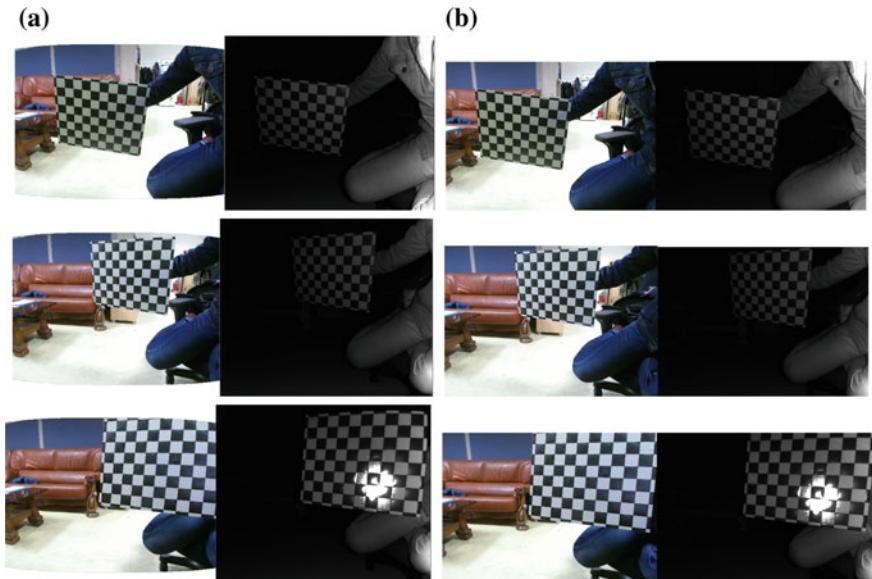


Fig. 4 Superimposed images (red lines are edges of IR image and white ones are edges of color image): **a** our method, **b** the SDK function

Table 3 Comparison of pixel distance

Image set	Method	
	SDK	Ours
A	32.5038	5.7118
B	21.3204	3.0104
C	24.8821	7.8142
Average	26.3254	5.5121

**Fig. 5** Three sets of color and IR images that are used in pixel position error calculation: **a** Result images from the SDK and **b** Result images from our method

4 Conclusion

In this paper, we performed a case study of aligning color, IR, and depth images using Kinect v2 sensor. We obtained the intrinsic parameters for Kinect v2 and calibrated the RGB and IR sensors via the transformation matrix between the two sensors. High accuracy of the proposed registration was confirmed by comparing to the results of the SDK function visually and numerically. As a future work we need a universal calibration matrix for all near and far distances covered by the Kinect v2.

Acknowledgments This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2013R1A1A2005024) and by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (NIPA-2015-H0301-14-4007) supervised by the NIPA (National IT Industry Promotion Agency).

References

1. Smisek, J., Jancosek, M., Pajdla, T.: 3D with Kinect. Consumer Depth Cameras for Computer Vision, pp. 3–25. Springer, London (2013)
2. Microsoft, Kinect for Windows. <http://www.microsoft.com/en-us/kinectforwindows>
3. Zhang, Z.: A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**, 1330–1334 (2000)
4. Rothwell, C., Forsyth, D.A., Zisserman, A., Mundy, J.L.: Extracting projective structure from single perspective views of 3D point sets. In: Fourth International Conference on Computer Vision, pp. 573–582. IEEE Press (1993)
5. Le, A.V., Jung, S.W., Won, C.S.: Directional joint bilateral filter for depth images. *Sensors* **14**, 11362–11378 (2014)

Equivalent Test Model of Wireless Optical Communication System for Automotive Environment

Sang Yub Lee, Jae Kyu Lee, Duck Keun Park and Jae Jin Ko

Abstract Driver's demands are increasing to connect with in-vehicle network using their smart devices for the media playing service needed to use high bandwidth and display quality in vehicular environment. Especially, either on environment of wire or wireless communication, customer wants to be in seamless network status. Recently, on wired communication, MOST (Media Oriented System Transport) network system satisfies as an in-vehicle network has been adopted in all kinds of cars. MOST networks uses optical fiber as the means of communication. As the optical network system is applied, it is contented with high data rates, low weight and high reliability without interference of electro-magnetic fields. Under the condition of being linked with MOST networks, proposed system which exchanges the data by wireless optical communication will take role of connectivity with MOST networks equipped in car. In this paper, it has been validated the interconnection of hetero network system and defined equivalent test model for automotive environment.

Keywords In-vehicle network systems · MOST · WOC · Automotive multimedia system

S.Y. Lee (✉) · J.K. Lee · D.K. Park · J.J. Ko
Software Device Research Centre, Korea Electronics Technology Institute, 22,
Daewangpangyo-ro, Bundang-gu, Gyeonggi-do Seongnam-si 463-490, Republic of Korea
e-mail: syublee@keti.re.kr

J.K. Lee
e-mail: jae4850@keti.re.kr

D.K. Park
e-mail: parkdk@keti.re.kr

J.J. Ko
e-mail: jaejini@keti.re.kr

1 Introduction

Lately, conventional car media system has been adopted to wireless communications system such as Bluetooth or Wireless LAN in order to transit the media data source, but the environment of exchanging multimedia streaming data in car is tough to be operated without ceasing link connection because vehicular space is small and its material cannot progress wireless signals. Proposed MOST-WOC system does not interfere with radio frequency systems and avoids EM (Electro-Magnetic) compatibility problems, thus it can be the only solution for streaming service tool in automotive media environment. Before being implemented automotive network for seamless connectivity combined with wired and wireless communication, equivalent model has to be considered as a valid means for communication in car. Though multimedia network is not effect on ECU (Electronic Control Unit) problem related on vehicle movement and safety in car, proposed WOC has to be checked as a network access point like the MOST networks which has been already equipped in car.

MOST is the de-factor standard for efficient and cost effective networking of automotive multimedia and infotainment system [1, 2]. The current MOST standards released MOST150, 150 means that 150 Mbps network bandwidth with quality of services is available. To meet the demands from various automotive applications, MOST network system provides three different message channels: control, synchronous used in streaming service and asynchronous channel only for packet data transfer. In describe in Fig. 1, proposed network system is consist of MOST devices with WOC module as a ring network. As it has been discussed before [3, 4], conventional WOC system had limitations of data bandwidth. It was a



Fig. 1 System model of MOST-WOC

maximum 2 Mbps Visible light downlink and 125 Kbps infrared uplinks such as light sensors. But, proposed MOST-WOC network system which is linked with MOST network system has the same as those data rate. It can be 150 Mbps uplink and downlink that make a variety of utilizations as a means of internal wireless communications in car.

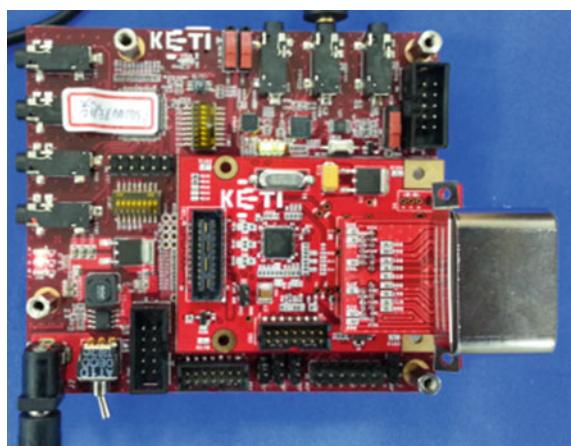
2 System of MOST-WOC

2.1 System Model of MOST-WOC

In describe in Fig. 1, proposed network system is consist of MOST devices with WOC module as an access device in ring network. The proposed MOST-WOC network system which is linked with MOST network system has the same as their data rate. It can be 150 Mbps uplink and downlink that make a variety of utilizations as a means of internal wireless communications in car.

Within a WOC device, the processing module possesses the role of a two way bridge functions between optoelectronic transceiver and the interior network of the device. Figure 2 shows the platform for the implementation of a typical processor style WOC device. The WOC device provides the transmission of converting data between optical signal and electrical one. Simultaneously, this allows sending and receiving the MOST data via EHC (External Host Controller) in an efficient way. The basic system architecture of MOST-WOC unit is shown in Fig. 3. It is described a device with an optic interface as the MOST side. The MOST optic interface transforms the light signal into an electrical signal which transfers the data routing module. When processing data are accessed to define memory block, MOST frame involved in control, streaming and packet data are transmitted to Head

Fig. 2 WOC network interface module with Fiber Optic Transceiver



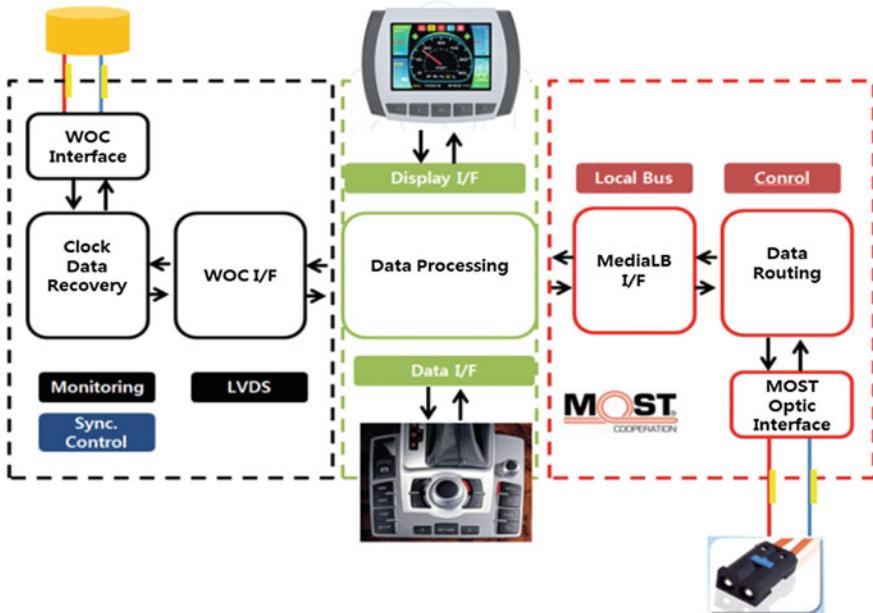


Fig. 3 Data processing linked with MOST and WOC

Unit via internal interface bus system. And, the data processing is realized in accessing to the WOC interface module.

3 Equivalent Test Model

3.1 Wired and Wireless Optical Network Model

The integration of processing platform and optic modules is particularly challenging due to their compatibility. An effective measurement of compliance is to be adopted a certification test process. It is called MOST compliance test. As described in Fig. 4, MOST compliance test process is divided into three courses: device test on physical layer, higher communication levels and application layer.

With regard to the physical layer compliance, the measurement point is to check the signal characteristics for constancy.

For the normal behavior, power, error, ring break diagnosis and network management, higher communication level compliance tests are performed.

The conventional physical layer compliance test on MOST networks system aims at testing complete device and can be considered as a black-box test, since only test points SP2 and SP3 can be tested. Figures 4 and 5 show the setup for physical layer compliance test.

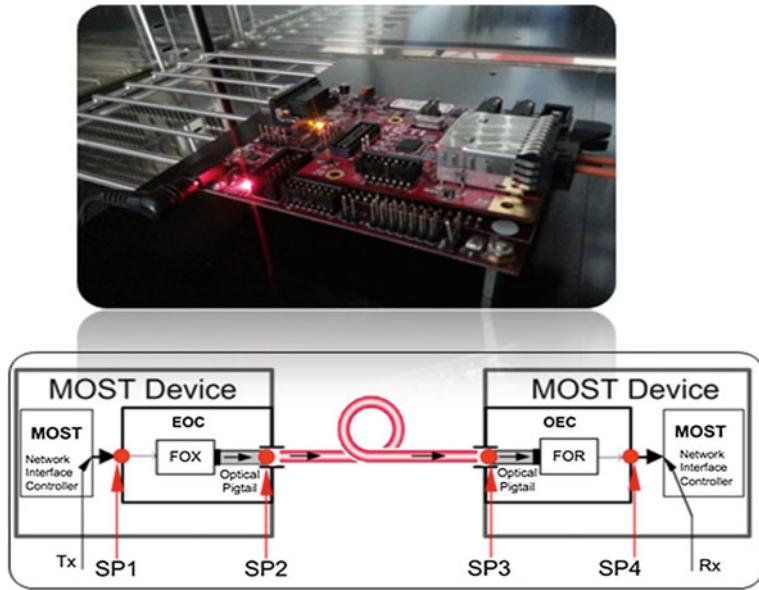


Fig. 4 Conventional physical layer test for MOST Compliance

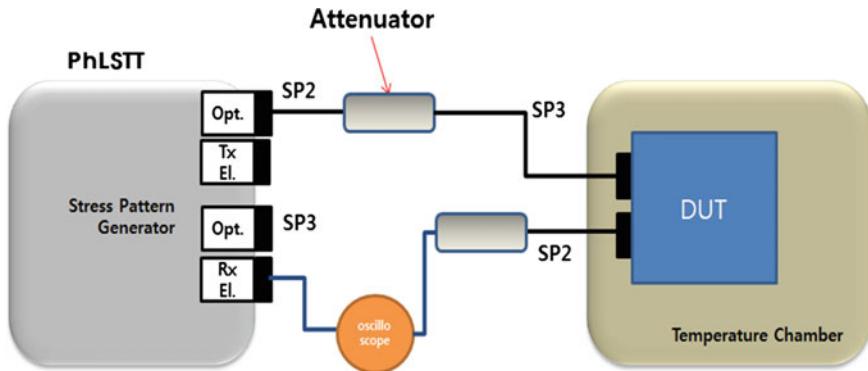


Fig. 5 Setup of general compliance test for wired optical network

Through the MOST compliance test setup, it is validated checking lock capability for variation of the optical input power and data consistency by comparison of the DUT input and output data stream. For the case of proposed experimental model, instead of the connected optical cable, optical wireless network devices are used as the test devices. To examine and demonstrate compliance test, it is defined attenuation level by interval replicated by power attenuation between the test devices. As compared with Figs. 6 and 7, it is presented how equivalent experimental model is comprised as same as MOST compliance test. The stress pattern

generator used in this experiment generates optical signal satisfying with MOST standardized data frame on SP2 as it is shown in Fig. 6. Comparable test model has a different point on SP2 that the stress pattern generator transmits the electrical MOST start frame signal and through the developed MOST-WOC linked platform, it exchanged electrical and optical signal equivalently.

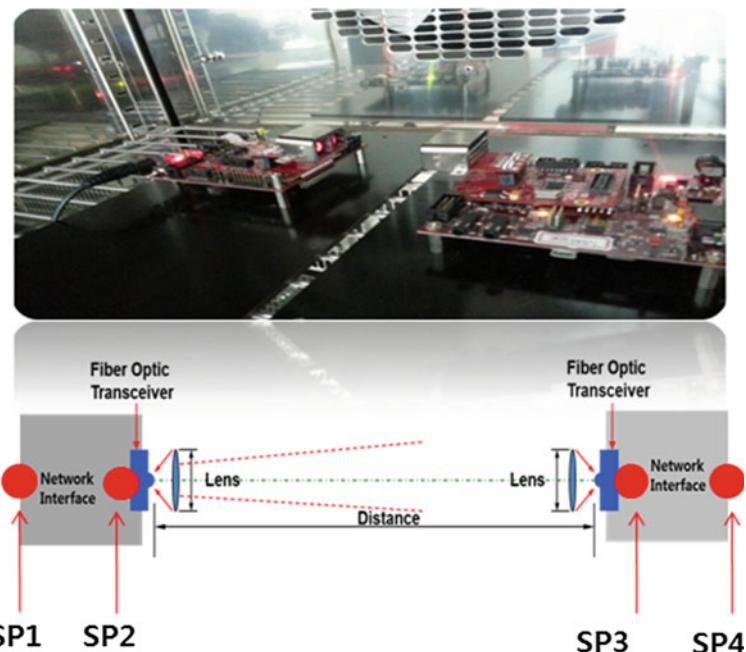


Fig. 6 Proposed WOC equivalent model for MOST compliance

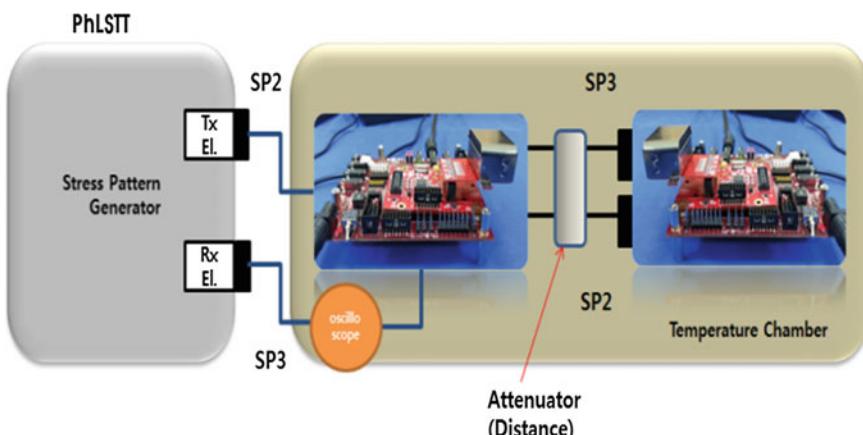
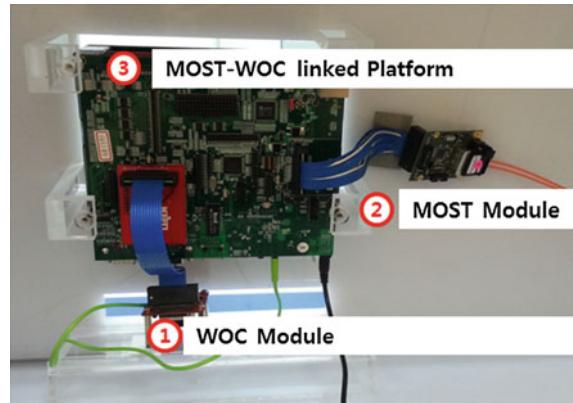


Fig. 7 Setup of proposed compliance test for WOC network

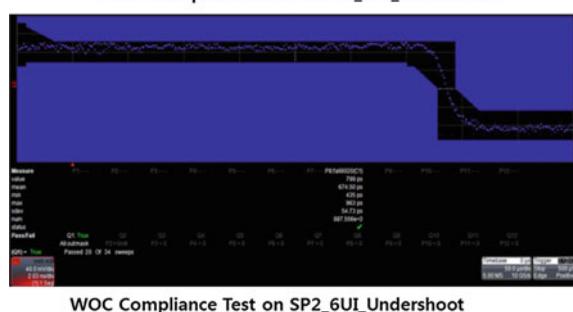
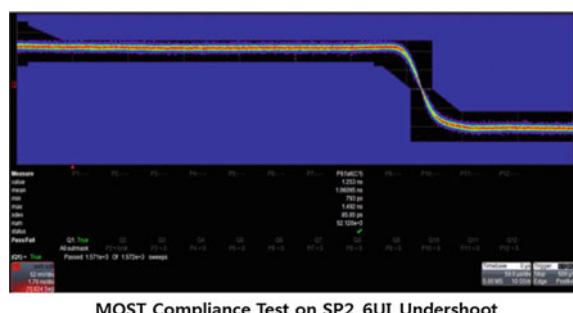
Fig. 8 Demonstration of MOST and WOC network system



4 Demonstration and Results

During the experimental model test, the MOST-WOC components are tested in the context of the final vehicle system. The focus of the proposed system test is to test the interoperability of all MOST components and WOC module. As shown in Fig. 8, the compatibility of the interlinked components under real life conditions. Through demonstration set, it covers as many situations as possible that occur in vehicular environment. In this case, it is important that the system results difference is caused carrying out WOC compliance test is the chronological synchronization of the different timing in access of interface bus and memory described in Fig. 9.

Fig. 9 Comparison results of MOST and WOC compliance test on SP2_UI_Undershoot



5 Conclusion

For the trend of car infotainment system is moving to the high quality audio sound system, this paper is introduced the development of the audio streaming service based on wireless optical network system called as WOC. Particularly, MOST-WOC, to be optimized streaming data transmitting without ceasing communication on vehicle environment is satisfied with reducing the weight and ensuring the reliability for the free of electro-magnetic problems. Especially, in this paper, it is introduced the equivalent test model of designed system based on MOST-WOC by setting up communication link. And it defines the test process of WOC system which transmit the packets linked with MOST networks. It is verified as an experimental model for the MOST compliance test.

Acknowledgments This work was supported by a grant from the IT R&D program of MOTIE/KEIT [10048285, Improving Industrial Infrastructure through embedded system research and development]

References

1. Grzembra, A.: MOST Books from MOST25 to MOST150. MOST Cooperation. FRANZIS, Germany (2012)
2. Strobel, O., Rejeb, R., Lubkoo, J.: Communication in automotive system principles, limits and new trends for vehicles, airplanes and vessels. In: IEEE ICTON, pp. 1–6 (2007)
3. Rufo, J., Quintana, C., Delgado, F., Rabadan, J., Perez-Jimenez, R.: Considerations on modulations and protocols suitable for visible light communications channels. In: IEEE CCNC, pp. 362–364 (2011)
4. Godavarty, S., Broyles, S., Parten, M.: Interfacing to the on-board diagnostic system. In: IEEE VTC, pp. 24–28 (2000)

Feasibility Study of Non-linear Apodization for IVUS B-mode Imaging

**Jin Ho Sung, Seon Mi Ji, Chan Yuk Park, Sung Yun Park,
Sung Min Kim, Won Seuk Jang, Byeong Cheol Choi
and Jong Seob Jeong**

Abstract This paper evaluates the performance of the non-linear apodization technique at the ultrasound B-mode (Brightness mode) image and indirectly demonstrates feasibility at the IVUS (Intravascular ultrasound) imaging application by changing coordinate condition. We conducted simulation with dedicated sound field programs (Field II and MATLAB), and the rectangular window and the Kaiser windows with two different control parameters were used for tri-apodization as an example of the non-linear apodization. The measurements of -6 dB main-lobe width and side-lobe level at the region of interest (ROI) were performed, and the point target simulation results show that tri-apodization had almost similar -6 dB main-lobe width and 6 dB lower side-lobe level in comparison of the result of the rectangular window. Therefore, this technique can be applied to IVUS B-mode image to improve image quality.

Keywords Ultrasound • Non-linear apodization • Window function • IVUS

1 Introduction

Intravascular ultrasound (IVUS) imaging technique has been frequently used for diagnosing and monitoring therapeutic process of cardiovascular diseases. Since IVUS transducer is injected into the interior vessel directly, IVUS imaging

J.H. Sung · S.M. Ji · C.Y. Park · S.Y. Park · S.M. Kim · J.S. Jeong (✉)

Department of Medical Biotechnology, Dongguk University, Seoul 100-715,

Republic of Korea

e-mail: jjsspace@dongguk.edu

W.S. Jang

Severance Hospital Clinical Trials Center for Medical Devices, Yonsei University,
Seoul 120-752, Republic of Korea

B.C. Choi

Department of Biomedical Engineering, Choonhae College of Health Science,
Ulsan 698-784, Republic of Korea

can provide various information about inside of the blood vessel such as diameter, thickness, region of lesion, or tissue component. Generally, the center frequency of the IVUS imaging is much higher than conventional frequency range and thus it suffers from low intensity and high noise components. Therefore, increasing the quality of IVUS image is very important [1]. Reducing side-lobe level is one of the popular ways to increase quality of ultrasound image. The high side-lobe level makes it difficult to distinguish small targets resulting in degraded image quality, so many researches have been studied to solve this problem. Linear apodization is one of the methods using window function to reduce side-lobe level, but it expands main-lobe width causing reduction of spatial resolution, specifically lateral resolution. To overcome this trade-off, non-linear apodization technique has been suggested [2].

In comparison of linear apodization, non-linear apodization technique uses unspecific shape of weight factor so as to maintain main-lobe width with restrained side-lobe level. The notable example of non-linear apodization is multi-apodization which uses various window functions at least more than two. It is simple way but the research has not been sufficient studied with IVUS imaging application. Therefore, in this study, we evaluated the performance of multi-apodization at the conventional B-mode image and changed coordinate condition from Cartesian to polar coordinate so as to demonstrate feasibility of multi-apodization at the IVUS imaging application.

2 Method

In this section, the features of used window functions and multi-apodization method was discussed by using impulse response (IPR). After that, we explained the method of multi-apodization technique for IVUS imaging. To demonstrate the performance of multi-apodization, we simulated tri-apodization as an example. The used window functions were the rectangular window and the Kaiser windows with two different control parameters. They were chosen because the rectangular window has the narrowest main-lobe width among the other window functions where -6 dB main-lobe width of 1.21 bins and the highest side-lobe level is -13 dB, and the Kaiser window has control parameter (α) which can be regulated by user's own purpose. The Kaiser window can be expressed as below [3, 4].

$$w(n) = \frac{I_0\left(\pi\alpha\sqrt{1.0 - \left(\frac{2n}{N}\right)^2}\right)}{I_0(\pi\alpha)}, \quad \left(0 \leq |n| \leq \frac{N}{2}\right) \quad (1)$$

where α is the control parameter and $I_0(X)$ is the zero-order Bessel function shown in the below

$$I_0(X) = \sum_{k=0}^{\infty} \left(\frac{\left(\frac{X}{2}\right)^k}{k!} \right)^2 \quad (2)$$

The discrete Fourier transform (DFT) of the Kaiser window is defined by

$$W(k) = \frac{N}{I_0(\alpha\pi)} \frac{\sinh\left(\sqrt{\alpha^2\pi^2 - \left(\frac{Nk}{2}\right)^2}\right)}{\sqrt{\alpha^2\pi^2 - \left(\frac{Nk}{2}\right)^2}}, \quad (0 \leq |k| \leq \pi) \quad (3)$$

The parameters of the Kaiser window were chosen as 0.5 and 1.0, and the reason of choosing parameter was discussed at the third paragraph in this section. Figure 1 shows IPRs of the rectangular window, the Kaiser windows with α of 0.5 and 1.0 respectively, and tri-apodization. The rectangular window had main-lobe width of 1.2 bins at the -6 dB and the highest side-lobe level was -13 dB. When control parameters of the Kaiser window were 0.5 and 1.0, the main-lobe widths and the highest side-lobe levels were 1.3 and 1.5 bins, and -17 and -25 dB respectively. The -6 dB main-lobe width of tri-apodization was same with that of the rectangular window and the highest side-lobe level of tri-apodization was measured as -22 dB.

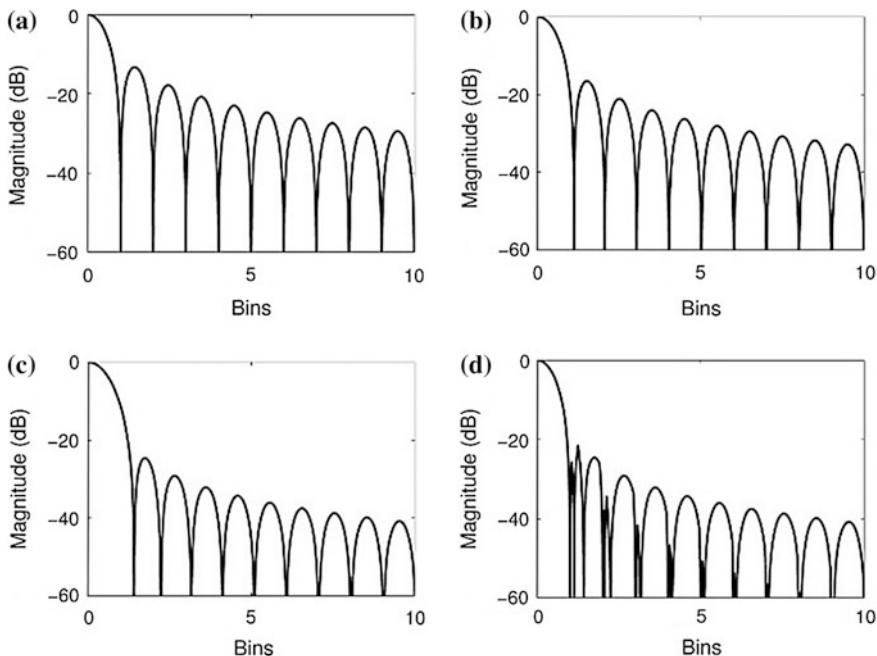


Fig. 1 Impulse responses (IPRs) of **a** rectangular window (R.W), **b** Kaiser window with $\alpha = 0.5$ (K.W ($\alpha = 0.5$)), **c** Kaiser window with $\alpha = 1.0$ (K.W ($\alpha = 1.0$)), and **d** Tri-apodization (T.A)

Table 1 -6 dB main-lobe widths of rectangular, and Kaiser windows with two different control parameters and tri-apodization

	Rectangular window [μm]	Kaiser window ($\alpha = 0.5$) [μm]	Kaiser window ($\alpha = 1.0$) [μm]	Tri-apodization [μm]
1st target	84.9	91.3	113.4	84.9
2nd target	181.8	190.2	222.0	190.6
3rd target	167.3	172.5	181.7	175.6
4th target	243.0	259.6	291.9	261.1
5th target	382.5	422.7	499.4	408.8

To obtain impulse response of the tri-apodization, three different windows; the rectangular, the Kaiser windows with two different control parameters, were applied to the signal respectively resulting in three different versions. After that, minimum value of intensity was selected at each spatial point. In the B-mode image, each windowed signal went through envelope detection process and minimum value of intensity was selected for tri-apodization at the comparator. After that, the log compression, scan converter processes were applied before display. Since the tri-apodization selects minimum value of intensity, the main-lobe width of additional window function should be cross under the highest side-lobe level of existing window function. Therefore, we chose the parameters of the Kaiser window as 0.5 and 1.0.

For the conventional B-mode simulation, Field II and MATLAB programs were used. 20 MHz, 128-element linear array transducer was employed and the number of elements of subaperture was 32. The total scanlines and pitch were 180 and 75 μm each. We assumed sound velocity of medium is 1500 m/s. The five point targets were located at the 2, 4, 6, 8 and 10 mm, and third target was located at the transmit focal point. After numerical measurement of -6 dB main-lobe width and side-lobe level, we changed Cartesian coordinate to polar coordinate to build similar condition for IVUS imaging (Table 1).

3 Results

The result of conventional B-mode image is shown in Fig. 2 with 60 dB dynamic range. Figure 2a had severe side-lobe artifact, and Fig. 2b, c had blurred target boundary instead of serious side-lobe artifact. Figure 2d shows the image applied tri-apodization, and it had low side-lobe artifact and clear boundary. To evaluate the performance of each windows and tri-apodization quantitatively, we used beam projection method in the lateral direction. It was to calculate side-lobe level and main-lobe width more accurately, since the locations that side-lobes occur were different among images. The axial range at the each point target was set to include

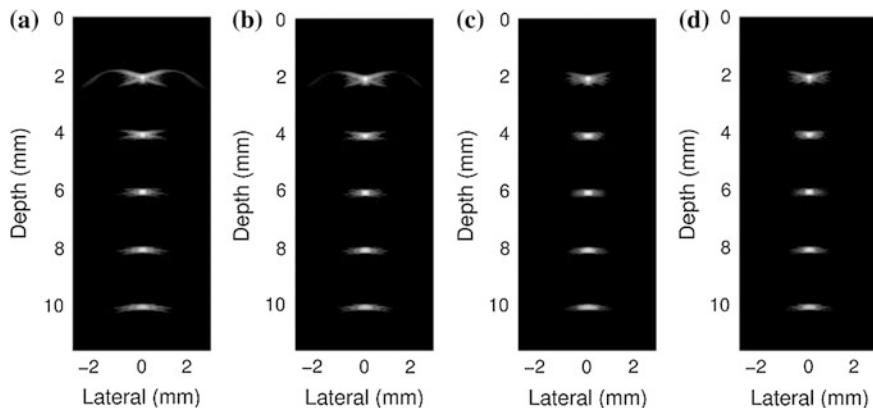


Fig. 2 B-mode image in Cartesian coordinate condition using **a** rectangular window, **b** Kaiser window with $\alpha = 0.5$ (K.W ($\alpha = 0.5$)), **c** Kaiser window with $\alpha = 1.0$ (K.W ($\alpha = 1.0$)), and **d** Tri-apodization (T.A)

sufficient side-lobe component. The main-lobe width was measured at the -6 dB, and the region of interest (ROI) for side-lobe was defined by 0.3 – 0.5 mm in the lateral direction. Figure 3 is lateral beam projection at the focal point. It shows that tri-apodization can maintain narrow -6 dB main-lobe width, and side-lobe level was averagely 6 dB lower than that of the rectangular. The numerical results were written at the Table 2. Figure 4 is an image that converted from Cartesian to polar coordination. To show side-lobe artifact more clearly, we changed dynamic range from 60 to 85 dB, and the results show tri-apodization can be employed to IVUS B-mode imaging application.

Fig. 3 Lateral beam projection image using **a** rectangular window, **b** Kaiser window with $\alpha = 0.5$ (K.W ($\alpha = 0.5$)), **c** Kaiser window with $\alpha = 1.0$ (K.W ($\alpha = 1.0$)), and **d** Tri-apodization (T.A)

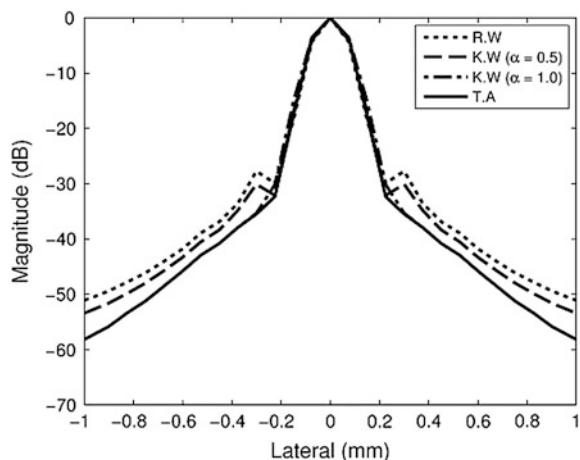


Table 2 Side-lobe level of rectangular, and Kaiser windows with two different control parameters and tri-apodization

	Rectangular window [dB]	Kaiser window ($\alpha = 0.5$) [dB]	Kaiser window ($\alpha = 1.0$) [dB]	Tri-apodization [dB]
1st target	-26	-29	-35	-35
2nd target	-29	-31	-34	-34
3rd target	-34	-36	-39	-39
4th target	-23	-26	-28	-29
5th target	-17	-19	-17	-20

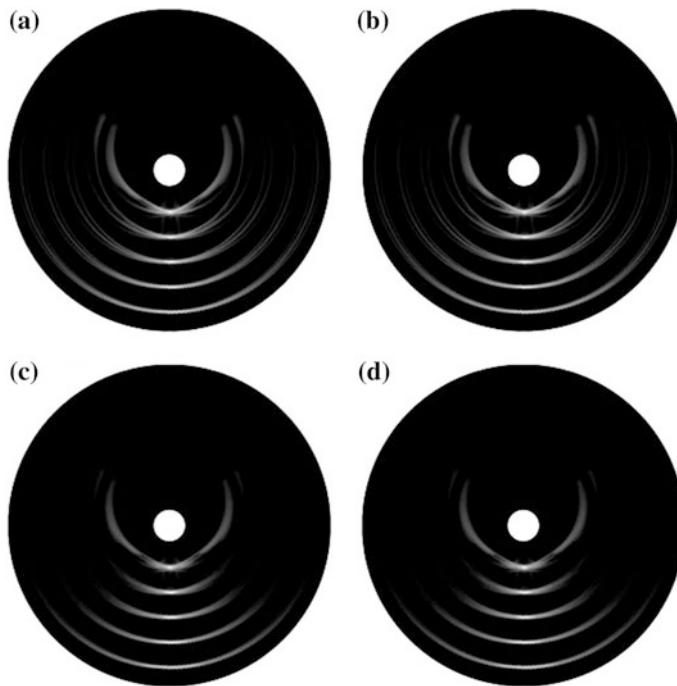


Fig. 4 B-mode image in polar coordinate condition using **a** rectangular window, **b** Kaiser window with $\alpha = 0.5$ (K.W ($\alpha = 0.5$)), **c** Kaiser window with $\alpha = 1.0$ (K.W ($\alpha = 1.0$)), and **d** Tri-apodization (T.A)

4 Discussion and Conclusion

In this paper, the performance of the non-linear apodization was demonstrated quantitatively at the conventional B-mode image. Subsequently, we changed image coordinate from Cartesian to polar coordinate condition to verify feasibility of

non-linear apodization at the IVUS B-mode imaging application indirectly. As previously mentioned, -6 dB main-lobe width and side-lobe level were calculated, and beam projection method was used for precise evaluation of the performance. The results show that tri-apodization as an example of non-linear apodization, can effectively reduce side-lobe level with narrow main-lobe width. The -6 dB main-lobe width was hardly expanded and side-lobe level at each point target was reduced about 6 dB on average with standard deviation of 4.7 . In the simulation, maximum intensity of each image applied to window functions was different, so the -6 dB main-lobe width from second to fifth targets had slightly larger main-lobe width than the rectangular window. It is because of simulation error, but it had still narrow main-lobe width in comparison of that of Kaiser window ($\alpha = 1.0$) with the lowest side-lobe level among the images. In the multi-apodization, the more window used, the better side-lobe level reduction can be achieved. However, it selects minimal intensity value of image, so additionally used window function should cross the highest side-lobe level of existing windowed image. Therefore, the control parameter should be carefully determined considering above-mentioned issues.

Acknowledgments This work was supported by R&D Program of Ministry of Trade, Industry and Energy/Korea Evaluation Institute of Industrial Technology (Grant No. MOTIE/KEIT 10048528) and the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the C-ITRC (Convergence Information Technology Research Center) support program (NIPA-2014-H0401-14-1002) supervised by the NIPA (National IT Industry Promotion Agency), and International Collaborative R&D Program (N01150049) funded by the Ministry of Trade, Industry & Energy (MOTIE), Korea.

References

1. Li, X., Wu, W., Chung, Y., Shih, W.Y., Shih, W.H., Zhou, Q., Shung, K.K.: 80-MHz intravascular ultrasound transducer using PMN-PT free-standing film. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **58**, 2281–2288 (2011)
2. Stankwitz, H.C., Dallaire, R.J., Fienup, J.R.: Nonlinear apodization for sidelobe control in SAR imagery. *IEEE Trans. Aerosp. Electron. Syst.* **31**, 267–279 (1995)
3. Harris, F.J.: On the use of windows for harmonic analysis with the discrete Fourier transform. *Proc. IEEE* **66**, 51–83 (1978)
4. Thomas, G., LoVetri, J., Chamma, W., Kashyap, S., Louie, A.: Sidelobe apodization for high resolution of scattering centres in ISAR images. *IEEE Antennas Prop. Soc. Int. Symp.* **2**, 232–235 (2001)

An Empirical Study of Impacts of User Intention for Smart Wearable Devices and Use Behavior

Yoo-Jin Moon, Young-Ho Hwang and Sungkap Cho

Abstract The paper utilized the method of questionnaires, and on the empirical level analyzed impacts on intention to use smart wearable devices and use behavior based on UTAUT. The analysis showed that intention to use smart wearable devices depended on the level of performance expected by the consumer in utilizing smart wearable devices, on the hedonic experiences that the consumers enjoy, on the social influence that the consumer referents exert, and on the facilitating conditions available. Also it indicated that the actual use of smart wearable devices depended on the intention to use and the facilitating conditions. For management and marketing strategies of smart wearable device providers, implications of the two factors of hedonic motivation and performance expectancy were that consumers should experience the devices with enjoyment and get benefits by utilizing them. And the marketing strategies should appeal to consumers by positioning the using experience as an adventure or a way to reduce their stress and change a negative mood.

Keywords Smart wearable devices · Intention to use · Performance expectancy · Hedonic motivation · Social influence

This work was supported by Hankuk University of Foreign Studies Research Fund of 2014.

Y.-J. Moon (✉)
Hankuk University of Foreign Studies, Seoul, South Korea
e-mail: yjmoon@hufs.ac.kr

Y.-H. Hwang
Kunsan National University, Gunsan-si, South Korea
e-mail: yhwang@kunsan.ac.kr

S. Cho
Korea University, Seoul, South Korea
e-mail: skc1777@naver.com

1 Introduction

Recently smart wearable devices are recognized as a new trend in mobile applications. A smart healthcare system with smart wearable devices can be provided as a collection of healthcare applications through a smartphone. This healthcare system uses advanced technology of terabyte memories by synchronizing with cloud systems, to intellectually monitor and manage patient conditions anytime anywhere, and to provide real-time customized services by analyzing patients' personal and disease information. Using smart care services is on the rise [1–3].

Smart wearable devices are related with the advanced technology, which is the core subject of the information technology acceptance theory [4]. The information technology acceptance theory is related to decision making of human's willingness to accept new technology. So technology advancement study includes variables related to human attitude. This study sets the research model using the extended UTAUT which integrates the existing theories.

UTAUT proposed by Venkatesh et al. [5, 8] includes three variables (performance expectancy, effort expectancy, and social influence) that affect intention to use, one variable (facilitating conditions) that affects usage behavior, and four controlled variables (sex, age, experience, and voluntariness). This study includes additional exogenous variables such as hedonic motivation and personal innovativeness, which become crucial when applying technology of smart wearable devices to UTAUT.

2 The Research Model and Hypotheses

2.1 The Research Model

The extended UTAUT provides an explanation for the acceptance and use of ICTs by consumers [5, 8]. To explain the future intention to use the technology and the past and present use of the technology in organizational contexts, the UTAUT posits that performance expectancy, effort expectancy, social influence, and facilitating conditions affect intention to use the technology, and that intention to use and facilitating conditions influence the actual use of the technology. Figure 1 provides two new constructs incorporated into the extended UTAUT: hedonic motivation and personal innovativeness [6, 9]. The research model revises the definitions of the six constructs, taking into account the four previously cited constructs of the UTAUT together with two additional constructs, as illustrated in Fig. 1 and Table 1.

In this study, the extended UTAUT is applied to analyze acceptance and use of smart wearable devices by consumers.

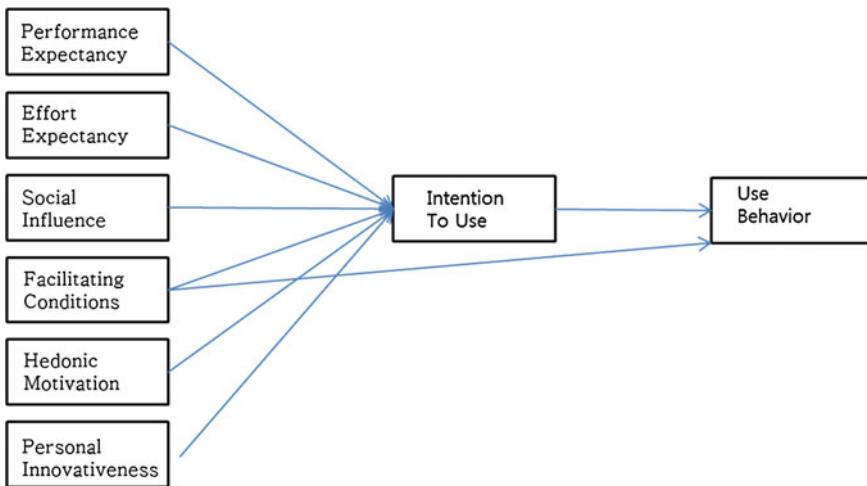


Fig. 1 The research model

Table 1 Operationalized definition

Constructs	Operationalized definitions
Performance expectancy	The degree to which using a new technology will provide benefits to consumers in utilizing smart wearable devices
Effort expectancy	The degree of ease/effort associated with consumer use of smart wearable devices
Social influence	The consumers perceive that important people (e.g. family or friends) believe that they should use a particular technology
Facilitating conditions	Consumer perception of the resources and support available to use smart wearable devices
Hedonic motivation	Pleasure or enjoyment derived from using a new technology
Personal innovativeness	The degree to which consumers search for and prefer use of smart wearable devices
Intention to use	The degree of intention to use smart wearable devices in the future
Use behavior	How often you use smart wearable devices

2.2 *Hypotheses Setting*

Taking into account the relationships and constructs of the extended UTAUT, we put forward the following hypotheses in respect of smart wearable devices by consumers.

- H1. The performance expectancy in the use of smart wearable devices would positively affect intention to use smart wearable devices.
- H2. The effort expectancy in the use of smart wearable devices would positively affect intention to use smart wearable devices.
- H3. The social influence regarding the use of smart wearable devices would positively affect intention to use smart wearable devices.
- H4. The facilitating conditions perceived in the use of smart wearable devices would positively affect intention to use smart wearable devices.
- H5. The hedonic motivation experienced in the use of smart wearable devices would positively affect intention to use smart wearable devices.
- H6. The personal innovativeness regarding the use of smart wearable devices would positively affect intention to use smart wearable devices.
- H7. The facilitating conditions perceived in the use of smart wearable devices would positively affect use behavior.
- H8. The intention to use smart wearable devices would positively affect use behavior.

2.3 Samples and Data Collection

Non-probability samplings were made to select the samples, since the population size for this study was unknown. The samples were three groups: university students at C University in Seoul and D University in the non-capital region of Korea, and E district officers of F City. Just before we distributed questionnaires, we explained those questioned about current smart wearable devices specifically. They completed self-reported questionnaires and voluntarily participated in responding the questionnaires. From October 5, 2014 to October 25, 2014 the final questionnaire was administered to 300 individuals. The total number of questionnaires completed by the individuals surveyed was 229 (76.3 % of the total surveyed), of which the total number of questionnaires usable for the data analysis was 225 (75 % of the total surveyed) after eliminating those questionnaires that had not been completed correctly.

There were more males (58.2 %) than females (41.8 %). The largest proportion of the respondents (64.4 %) was aged between 20 and 29. The largest proportion of the academic background (57.8 %) was college students, followed by college graduates (30.7 %). The most occupation (58.2 %) was students, followed by public servants (33.8 %). There were experienced users with smart wearable devices (16 %).

3 Hypotheses Verification and Empirical Analysis

A regression analysis of latent variables was used in this research, based on the optimization technique of the partial least squares (PLS) to elaborate a model representing relationships between the eight proposed constructs measured by many items. The PLS is a multivariate technique to test the structural model, estimates the model parameters to minimize the residual variance of the whole model dependent variables, does not require any parametric conditions, and is recommended for small samples.

The data analysis took place through a two-stage methodology. The first stage was to develop and evaluate the measurement model and the second stage was to develop the full structural equation model.

3.1 Reliability Analysis

The study verified reliability of the model using collected data ($n = 225$). To measure the internal coherence of all the indicators in relation to the constructs, it tested reliability between multi-item scales on 22 measurement items using SPSS 18 program. The Cronbach coefficient α values are shown in Table 2. They are all above 0.7, which is recommended for confirmatory research.

3.2 Validity Analysis

The study performed the exploratory factor analysis about items of the questionnaire measuring constructs of the research model. A factor extraction method was based on principal component analysis and Varimax rotation with Kaiser-normalization [7].

As illustrated in Table 3, results of the exploratory factor analysis showed that all seven initially intended factors were extracted: F1 'personal innovativeness',

Table 2 Results of reliability analysis

Variables	No. of items	Cronbach α	Standardized cronbach α
Performance expectancy	3	.873	.883
Effort expectancy	3	.872	.882
Social influence	3	.868	.880
Facilitating conditions	3	.884	.898
Hedonic motivation	3	.870	.878
Personal innovativeness	4	.901	.908
Intention to use	3	.868	.883

Table 3 Results of the exploratory factor analysis

Factors	Items	F1	F2	F3	F4	F5	F6	F7
Personal innovativeness	V1	.809	.141	.154	.065	.054	.273	.111
	V2	.894	.137	.057	.106	.081	.068	.114
	V3	.834	.000	.077	.057	.208	-.004	.131
	V4	.735	.043	.203	.181	.047	.119	.357
Effort expectancy	V5	.265	.247	.238	.138	.103	.177	.772
	V6	.285	.181	.175	.250	.277	.128	.723
	V7	.280	.229	.157	.245	.158	.349	.708
Social influence	V8	.149	.291	.217	.754	.262	.228	.187
	V9	.153	.272	.193	.783	.221	.250	.211
	V10	.151	.294	.179	.757	.239	.273	.222
Performance expectancy	V11	.087	.787	.189	.270	.264	.218	.204
	V12	.113	.765	.138	.278	.274	.222	.190
	V13	.131	.783	.187	.236	.235	.243	.208
Hedonic motivation	V14	.124	.329	.175	.277	.286	.669	.212
	V15	.169	.216	.183	.240	.226	.798	.197
	V16	.214	.266	.238	.286	.248	.718	.226
Facilitating conditions	V17	.165	.161	.839	.186	.108	.040	.156
	V18	.091	.198	.781	.193	.116	.296	.178
	V19	.157	.078	.811	.075	.256	.140	.116
Intention to use	V20	.174	.355	.210	.241	.700	.256	.181
	V21	.141	.272	.224	.205	.779	.257	.201
	V22	.161	.223	.189	.244	.814	.185	.120
Eigen value		3.25	2.79	2.62	2.61	2.58	2.48	2.33
Explained variance (%)		14.770	12.699	11.885	11.855	11.719	11.287	10.592
KMO (%)		84.807						

F2 ‘performance expectancy’, F3 ‘facilitating conditions’, F4 ‘social influence’, F5 ‘intention to use’, F6 ‘hedonic motivation’, and F7 ‘effort expectancy’. Each factor showed that Eigen value was above 1 and the rate of cumulative variance showed 84.807 % of total variance. This study found that multi-collinearity did not exist. Items measuring the same construct represented prominently higher factor loadings on a single construct than other constructs (boldface).

3.3 Hypotheses Verification

1. T-test verification regarding gender and user experience

Results of the T-test verification on samples showed that no statistically significant difference between genders existed. Among six variables three variables—effort expectancy, hedonic motivation and intention to use—showed

Table 4 Results of T-test between users and non-users

	Levine's equal variance		T-test on identity of mean	
	F	α	t	α (two-tail)
Performance expectancy	.051	.822	1.948	.057
Effort expectancy	1.194	.276	3.798	.000
Social influence	.002	.967	-1.466	.149
Facilitating conditions	.089	.766	1.525	.134
Hedonic motivation	.005	.941	3.432	.001
Personal innovativeness	.359	.550	1.921	.061
Intention to use	.176	.675	2.9425	.007

Table 5 Summary of test results for the structural model

Hypothesis	Path	β	P-value	Accepted	Construct	R^2
H1	PE → IU	.308	.000	Yes	Intention to use	.652
H2	EE → IU	.028	.683	No		
H3	SI → IU	.184	.009	Yes		
H4	FC → IU	.126	.021	Yes		
H5	HM → IU	.227	.001	Yes		
H6	PI → IU	.056	.261	No		
H7	FC → UB	.250	.001	Yes	Use behavior	.610
H8	IU → UB	.367	.000	Yes		

statistically significant differences between users and non-users at the level of $\alpha = .05$, as shown in Table 4.

2. Hypotheses verification using the structural model

A PLS analysis was performed to test H1 through H8. The regression parameters were based on a bootstrapping of 100 samples but not on a sample estimator, which facilitated the computation of the t-student for each hypothesis and generalization of the results. The results shown in Table 5 indicate relationship between the different constructs. All the R-squares are higher than .10, which means that the productive capability of the model is satisfactory.

Firstly, the results show that the main predictors of intention to use, in order of importance, are performance expectancy, hedonic motivation, social influence and facilitating conditions. That is, intention to use smart wearable devices depends on the level of performance expected by the consumer in utilizing smart wearable devices, on the hedonic experiences that the consumers enjoy when using the smart wearable devices, on the social influence that the consumer referents exert, and on the facilitating conditions available. However, in the hypothesized causal

relationships in the proposed model, impact of effort expectancy on intention to use and that of personal innovativeness on intention to use were not supported.

Secondly, the results highlight that main predictors of actual use behavior, in order of importance, are intention to use and facilitating conditions. Therefore, it can be stated that the actual use of smart wearable devices depends on the intention to use and the facilitating conditions.

4 Conclusions

Based on results of this research, practical recommendation can be made to the providers of smart wearable devices regarding appropriate management and marketing strategies for improving key parts of their business model. This research explains how consumers behave regarding use of smart wearable devices. In particular this research aims to analyze factors that influence both consumers' intention to use and their actual use of smart wearable devices, for which we utilize a new adapted and extended version of the UTAUT.

The results show that the factor of facilitating conditions influences the intention to use and the actual use behavior, which coincides with the results of the previous studies [5, 8, 10] and contrasts with the findings obtained by others [6, 9]. This means that consumer perceptions of the support and the resources for using smart wearable devices influence both intention to use and actual usage. Consequently, by making support and resources available the providers had better facilitate customers to purchase the smart wearable devices, so that consumers can make access to the means necessary to resolve any problem that they may encounter during the using process.

Implications of the two factors of hedonic motivation and performance expectancy can be headed for the management and marketing strategies of smart wearable device providers, which is that consumers should experience the devices with enjoyment and get benefits by utilizing them. Thus, the smart wearable devices have to be elaborated in the pleasant and beneficial way. Marketing strategies should appeal to consumers by positioning the using experience as an adventure or a way to reduce their stress and change a negative mood.

Social influences affect intention to use; a finding that coincides with the results from some studies [5, 8, 10] and contrasts with the findings from others [6, 9], which means that consumers form intention to use because they imagine that their referents, such as friends, family and colleagues, think that they should use. Therefore, one of the marketing strategies for smart wearable devices should be reputation-building, in order to gain a favorable opinion from referents, whether they are existing users or not, so that these persons can actively recommend using the devices to others.

Finally, the results indicate the effect of intention to use on use behavior: the greater the intention to use, the higher the probability of use behavior. Clearly, providers should aim to strengthen consumers' intention to use, which will lead to

more actual use of the devices. Accordingly, they need to take action on particular variables like performance expectancy, hedonic motivation, social influence and facilitating conditions to increase the intention to use.

Future studies should examine that the model proposed can be applied for other kinds of smart products and services. It would also be of interest to analyze the possible cross-cultural differences in the determinant factors that influence consumers in their intention to use and their actual use of smart wearable devices. In addition, the influence of social demographic variables as moderator variables might be examined. And it would be helpful to conduct a longitudinal analysis to determine how these variables change over time.

References

1. Takuji, S., Hirokazu, T., Shigenobu, M., Hiroshi, Y., Takashi, M.: Wearable wireless vital monitoring technology for smart health service. In: 2013 7th International Symposium on Medical Information and Communication Technology (ISMICT), pp. 1–4 (2013)
2. Ishaq, I., Hoebelke, J., Rossey, J., Poorter, E.D., Moerman, I., Demeester, P.: Enabling the web of things facilitating development, discovery and resource access to IoT objects using embedded web services. *Int. J. Web Grid Serv.* **10**(2/3), 218–243 (2014)
3. Peck, J.L., Stanton, M., Reynolds, G.E.: Smartphone preventive health care: parental use of immunization reminder system. *J. Pediatr. Health Care* **28**(1), 35–42 (2014)
4. Wallace, L.G., Sheetz, S.D.: The adoption of software measures: a technology acceptance model (TAM) perspective. *Inf. Manag.* **51**, 249–259 (2014)
5. Lescevica, M., Ginters, A., Mazza, R.: Unified theory of acceptance and use of technology (UTAUT) for market analysis of FP7 CHOReOS products. *Procedia Comput. Sci.* **26**, 51–68 (2013)
6. Parameswaran, S., Kishore, R., Li, P.: Within-study measurement invariance of the UTAUT instrument: an assessment with user technology engagement variables. *Inf. Manag.* **52**(3), 317–336 (2015)
7. Hwang, Y., Hwang, W., Moon, Y.: A study of factors influencing intra-organizational potential users' intention to use N-screen service. *KCGR* **16**(1), 89–114 (2012)
8. Venkatesh, V., Thong, J.I.I., Xu, X.: Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. *MIS Q.* **36**(1), 157–178 (2012)
9. Escobar-Rodriguez, T., Carvajal-Trujillo, E.: Online purchasing tickets for low cost carriers: an application of the unified theory of acceptance and use of technology (UTAUT) model. *Tour. Manag.* **43**, 70–88 (2014)
10. Aizstrauma, D., Ginters, E., Eroles, M.P.: Applying theory of diffusion of innovations to evaluate technology acceptance and sustainability. *Procedia Comput. Sci.* **43**, 69–77 (2015)

Lightweight Context-Aware Activity Recognition

Byung Gill Go, Asad Masood Khattak, Babar Shah
and Adil Mehmood Khan

Abstract In ubiquitous environments, it is important to recognize the situation and deliver services accordingly. In addition, it is equally important to have a fast response time. The existing context-aware activity recognition engines have good recognition rates; however, they consume lots of time to produce feasible results. Our focus in this research is to reduce the time required by eliminating the need for ontology matching (in context-aware activity manipulation engine) and extend the rules. In addition, we incorporate the sliding time window concept to retain activities for a longer duration and maintain their relevance using ontological data for a better accuracy. The proposed scheme has increased the overall accuracy against the existing system by 12.6 % for individual activities relevance and 6 % for high level activities.

Keywords Activity recognition · Ontology · Knowledgebase

1 Introduction

The computing paradigm has evolved a great deal. Giving a static response to an explicit input from a user is a thing of the past. Today's computers are smart enough to sense and understand the user's context and reacting accordingly. The recog-

B.G. Go · A.M. Khattak (✉) · B. Shah

College of Technological Innovation, Zayed University, Abu Dhabi, United Arab Emirates
e-mail: asad.khattak@zu.ac.ae

B.G. Go
e-mail: sugujang@naver.com

B. Shah
e-mail: babar.shah@zu.ac.ae

B.G. Go · A.M. Khan
Department of Computer Science, Innopolis University, Kazan, Russia
e-mail: a.khan@innopolis.ru

nition of user's context is dependent on time and location [1]. Typically, users' context consists of their current location, time, proximity to other users, status of their home appliances, their behavior, work histories, and so on [2, 3]. To recognize context, sensors are deployed to sense and report the context of user. The recognized context is then analyzed to provide recommendations and services [2, 4–7].

In any context-aware computing system, obtaining context in an accurate and reliable way is of utmost importance and a large number of studies have been conducted in this regard [5]. Management of such collected information is another challenge, which is necessary to provide better services to the user in his/her given context [5, 8, 9]. To achieve these objectives, we propose context-aware activity manipulation engine (CAME) that works using ontological data and their matching process to facilitate users in their context [7]. CAME works in two phases: (1) Ontology Based Match Making: the knowledgebase ontology is matched against the temporary ontology created from the newly detected activities and if a match is found it is retrieved for further processing. Furthermore, this process also retrieves unrelated activities from the knowledgebase. (2) Rule Based Filtering: rule based filtering works with a knowledgebase of rules and an inference engine [1]. The rules part is responsible for expressing the knowledge and if it results in a true value by establishing relationships then the actions are executed on well-defined objectives. Due to this two phase nature, CAME is very time consuming, especially for high level context recognition. Our recent studies have shown that the same results are also achievable even if we skip the match making step of CAME (which is the most time consuming phase), by increasing rules in the rule-base. This idea helps us in creating a lightweight context-aware activity recognition scheme which is a customization of CAME.

In this research, the detected activities are first collected and preprocessed for the purpose of representation in a unified representational format developed using ontology. This recognized and formally represented activities/context is then used in reasoning engine with the help of semantic web rule language (SWRL) rules to recognize the related higher level context. For instance, if activities (context) like jumping, sitting, standing, walking, and running are recognized (in any order) then the rule-based reasoning produces the related high level activity (context) as exercise [7]. To achieve the same level of accuracy, we have introduced a sliding time window in the reasoning process so that the list and sequence of activities are maintained for a longer duration inside the sliding window and are used for determining appropriate high level activities. It is obvious that after eliminating the ontology based match making process the proposed scheme will take less time as compared to CAME, so we are not presenting results for proposed system's efficiency. The experiments and results we have conducted in this research are for accuracy of the proposed system against the existing system [8]. The proposed system has achieved 12.6 % overall better results as compared to the existing system.

This paper is organized as follow: Sect. 2 presents the proposed lightweight context-aware activity recognition system with details on the working of our proposed scheme. Section 3 presents detailed results of the proposed scheme against the existing system. We conclude our research paper in Sect. 4 and talk about future directions.

2 Rule-Based Lightweight Context-Aware Activity Recognition

In a ubiquitous environment, sensor devices work in a similar way as human senses do to sense and update us about our surroundings. The ubiquitous environment is configured to transfer data from the sensor devices to the system, where the data are processed and analyzed for recommendations and services execution [7]. To facilitate the overall process, data collection from the sensors is very important; however, the collected data, most of the times, are not meaningful. Due to the diverse nature of sensors, the data are mostly in different type and also with different structure. The Preprocessing of Data module (in Fig. 1) is responsible to

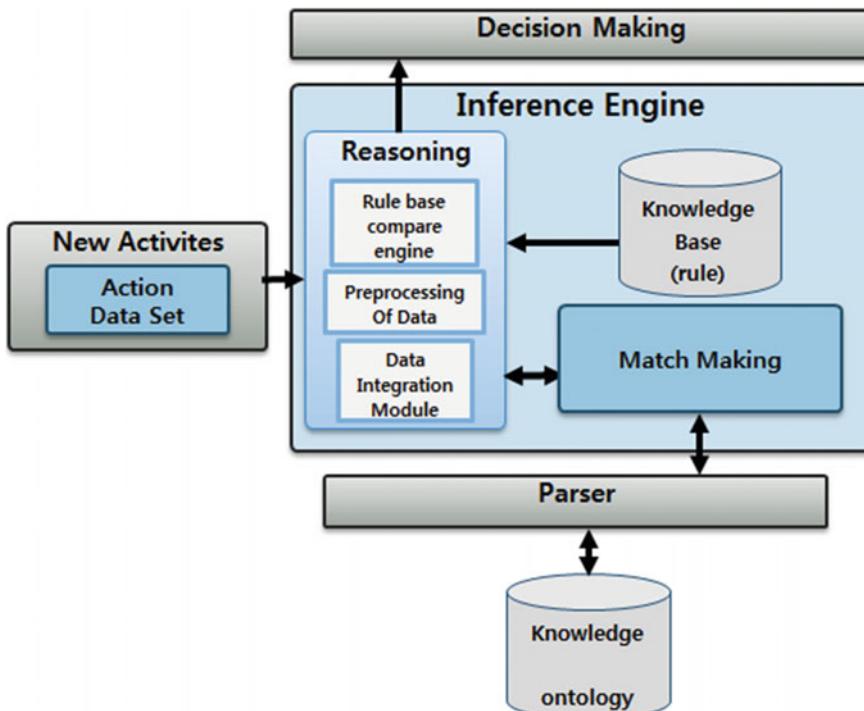


Fig. 1 Overall architecture of the proposed lightweight context-aware activity recognition system

convert the collected data into a unified form constrained to a unified structure so that the process of match making and rule-based engine can work properly.

This research mainly deals with the optimization of context-aware activity manipulation engine [7], where ontology matching part has been removed and the overall system is compressed into a lightweight application to work on smart devices as well [9, 10]. After the newly recognized activity is represented in a proper representational format, it is stored in the Knowledgebase developed using OWL ontology. Parser module (as shown in Fig. 1) is mainly responsible for the storage of the new activities and retrieval of stored information for decision making. The Match Making module in this architecture is mainly responsible for constructing relevant sparql queries (not ontology matching) when required and execute them on the Knowledgebase. The abstract working of proposed system is shown in Fig. 2 with a scenario of how the execution is taking place.

The Reasoning module of the proposed architecture is activated for any newly recognized activities. After formal representation of the activity relevant information from the Knowledgebase is extracted. The extracted information and the detected activities are integrated for further processing with SWRL rules for reasoning to infer the actual situation and the related high level activity (context). SWRL rules are compiled and used by the Reasoning module on the aggregated activities information to filter and recognize the high level activities (context). These high level activities are later used for Decision Making where these decisions can be the initiation of recommendations/services or just storage of the high level context in the knowledgebase, which might be useful in the future [5, 10].

As mentioned above, the ontology matching process is eliminated from the current version in order to optimize it. The elimination is due to the reason that the

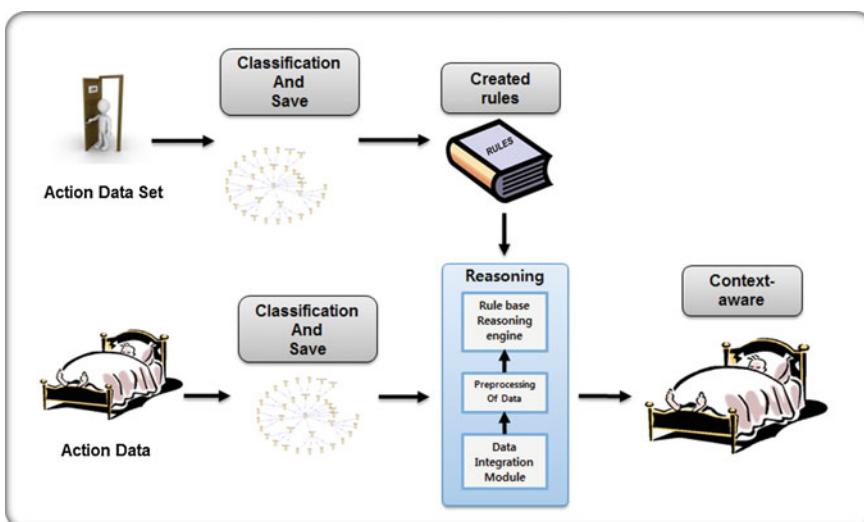


Fig. 2 Abstract view of the proposed scheme working in a given scenario

ontology matching process needs extra memory and computation time and these are the limited resources on a mobile device. To overcome any issues that might arise due to this elimination, the number of required rules is increased and the rules are fine-tuned. For the rules to execute properly, we have used a sliding time window procedure that facilitates in restricting rules to a defined time, which is not explained here due to space limitation.

3 Experiments and Results

This section presents the experiments and results of our proposed system. Online available dataset is used in the experiments which was also used in [8]. The dataset contains 245 instances of high level activities (shown in Table 1) which are realized from a set of low level activities.

The details on sensors used and their deployment in the experimental environment is available in the dataset. In addition, the start time, end time, position and location information is also provided that facilitates in understanding the purpose and operation of these. The existing method is compared with our proposed method and with rules based filtering a significant amount of change is realized using our proposed scheme. The results of existing system against the proposed research is given in Table 2. The main reason for the change in results is because of the size of sliding window and introduction of retaining the recognized activities for relevance and association with next incoming activities; whereas, the existing system only focused on the pre-existing relationship of the recognized activity with information stored in the knowledgebase. Using our proposed system, we recognized 12.61 % of overall change (activity to activity association) in results which was not recognized by the existing system.

In the next experiments, to work on user's behavior, a set of 140 activities is selected from a list of 245 and 100 rules are defined to work on these activities. When the results are compared, the proposed system is able to recognize 6 % more high level activities compared to the actual data as shown in Table 3. The reason for this is the use of ontology to retain relevance among stored data and then facilitate them to remain inside the sliding time window due to their relevance [7].

Table 1 List of activities and their instances used in experiments

Activity	Number of instances
Leaving	34
Toileting	114
Showering	23
Sleeping	24
Breakfast	20
Dinner	10
Drink	20

Table 2 System comparison against existing system for the rate of change recognized with the proposed system

Activities Data		Higher level activities using existing system	Higher level activities using proposed system	Total change in higher level activities	Percentage of change (%)
Use_toilet	Bathroom_door	202	215	13	6.44
Use_toilet	Toilet_door	51	80	29	56.86
Go_to_bed	Bedroom_door	34	35	1	2.94
Use_toilet	Flush	127	132	5	3.94
Take_shower	Toilet_door	58	61	3	5.17
Drink	Cup	21	24	3	14.29
Breakfast	Fridge	40	44	4	10.00
Drink	Fridge	35	39	4	11.43
Dinner	Fridge	35	42	7	20.00
Dinner	Pans	26	28	2	7.69
Breakfast	Groceries	29	35	6	20.69
Dinner	Cup	6	10	4	66.67
Breakfast	Cup	2	5	3	150.00
Total		666	750	84	12.61

Table 3 Result of proposed system against the existing while applying a set of 100 rules

Activity	Actual set	Existing scheme	Proposed scheme	Change ratio (%)
Leave	22	15	16	5
Used toilet	69	56	62	9
Shower	12	9	10	8
Sleeping	11	8	8	0
Breakfast	8	6	6	0
Dinner	7	5	6	14
Drinking	11	8	8	0
Total	140	107	116	6

For instance, the activity of cup is relevant to the activity of breakfast. However, if it is not included in the sliding time window then the generated result will be very much different.

4 Conclusion

The recognition of high level activities greatly depends on the activity time, the sequence of activities and their relevance to pre or post activities. To recognize the context/high level activities, it is confirmed that there is a need for relevant data within the referred time. This relevant data is used in the proposed system with ontology for more accurate context/situation analysis. In addition, creation and execution of rules has significant impact on the overall system working in terms of both time for execution and maintaining sliding time window of activities for better recognition. This actually helped to eliminate the ontology matching (time consuming) procedure from the system and make it a lightweight system with lesser need for resources. In future, we are looking forward to extend this concept to behavior analysis of a user.

References

1. Choi, J., Lee, G., Moon, J.: Web context classification based on information quality factors. *J. Univers. Comput.* **16**, 2232–2251 (2010)
2. Yoon, J., Lee, S., Suh, Y., Ryu, J., Woo, W.: Information integration system for user recognition and location awareness in smart environment. *KHCI* (2002)
3. Davies, N., Cheverst, K., Mitchell, K., Efrat, A.: Developing a context sensitive tour guide. In: *Proceedings of 1st Workshop on Human-Computer Interaction for Mobile Devices* (1998)
4. Kortuem, G., Segall, Z., Bauer, M.: Context-aware, adaptive wearable computers as remote interfaces to ‘intelligent’ environments. In: *The Proceedings of the 2nd International Symposium on Wearable computers*, pp. 58–65 (1998)
5. Khattak, A.M., Akbar, N., Aazam, M., Ali, T., Khan, A.M., Jeon, S.K., Hwang, M.G., Lee, S. Y.: Context representation and fusion: advancements and opportunities. *J. Sens.* **14**(6), 9628–9668 (2014)
6. Banaver, G., Bernstein, A.: Issues and challenges in ubiquitous computing: software infrastructure and design challenges for ubiquitous computing applications. *Commun. ACM* **12**, 92–96 (2002)
7. Khattak, A.M., Truc, P.T.H., Hung, L.X., Vinh, L.T., Dang, V.H., Guan, D., Pervez, Z., Han, M.H., Lee, S.Y., Lee, Y.K.: Towards smart homes using low level sensory data. *J. Sens.* **11** (12), 11581–11604 (2011)
8. Kasteren, T., Noulas, A., Englebienne, G., Kroese, B.: Accurate activity recognition in a home setting. In: *UbiComp’08*, Seoul, Korea, 21–24 Sept 2008
9. Tabatabaei, H., Amir, S., Gluhak, A., Tafazolli, R.: A survey on smartphone-based systems for opportunistic user context recognition. *ACM Comput. Surv.* **45** (2013)
10. Khan, A.M., Lee, Y.K., Lee, S.Y., Kim, T.S.: A triaxial accelerometer-based physical activity recognition via augmented features and a hierarchical recognizer. *IEEE Trans. Inf. Technol. Biomed.* **14**, 1166–1172 (2010)

Efficient Sliding Window Join in Data Stream Processing

Hyeon Gyu Kim

Abstract Our previous work compared two possible approaches to process sliding window join over continuous data streams. It showed that using multiple hash tables per stream source provided better performance than using a single hash table. On the other hand, performance of the single-table scheme can be improved by reducing the windowing cost to deal with expired tuples. In this paper, we discuss how to reduce the cost in the single-table scheme algorithm as a solution for efficient sliding window join.

Keywords Window join • Sliding windows • Symmetric hash join • Data streams

1 Introduction

Recently, real-time analytics on continuously updated data sets has become essential to meet many enterprises [1, 2]. Numerous applications must process on-the-fly data, often with minimal latency and high accuracy. For example, an application that monitors the Twitter Firehose for an ongoing earthquake may want to report relevant information within a few seconds of when a tweet appears, and must handle drastic spikes in the tweet volumes [3]. Similar applications can be found in monitoring the stream of data items, such as stock exchanges, network measurements, telecommunication logs, web page visits, sensor readings, and so on.

Sliding window join is popular in processing continuous data streams [4]. A well-known example of the window join query is packet identification in a network path which can be used for network intrusion detection. Consider a query to identify packets flowing in a network path over the last 5 min, where the path

H.G. Kim (✉)

Department of Computer Engineering, Sahmyook University, Seoul 139-742,
Republic of Korea

e-mail: hgkim@syu.ac.kr

consists of three routers, R1, R2 and R3. The result needs to be updated every minute in this example. To identify the packets, an equijoin can be performed among the three router streams over a common attribute *pid*, denoting a packet ID, as follows.

```
Q1.  SELECT *
      FROM R1[RANGE 5 mins, SLIDE 1 min],
           R2[RANGE 5 mins, SLIDE 1 min],
           R3[RANGE 5 mins, SLIDE 1 min]
      WHERE R1.pid = R2.pid and R2.pid = R3.pid
```

Above, RANGE and SLIDE denote the size and the slide interval of a window, respectively. The query is evaluated every minute, i.e., whenever windows are updated.

Suppose that it is evaluated at t seconds. Then, all window intervals are first updated to $(t - 300, t]$, and then *expired tuples* whose timestamps lie outside the window interval are discarded from memory. After the windowing is finished, join is performed over the tuples in memory.

The window slid by time (i.e., updated periodically) is referred to as a *time-slide window* [5, 6]. One characteristic of the time-slide window is that adjacent windows can overlap during the query execution, from which duplicate processing can be incurred [7]. This happens when the window slide interval is set smaller than the window size (e.g., *Q1*) in order to help users to receive query responses more frequently.

Our previous work [5, 6] compared two possible join algorithms when the time-slide windows are given in a query. One is a conventional approach where a single hash table is allocated for each stream source, which is called a *single-table scheme*. In the example of *Q1*, three hash tables are used in this scheme, where each table is allocated for R1, R2 or R3 (Fig. 1a). The other is a new approach where multiple hash tables are allocated for each stream. In this *multi-table scheme*, a hash table is allocated for each set of tuples arriving for a window slide interval. In case of *Q1*, five hash tables will be allocated for each stream (Fig. 1b).

The previous work discussed that the two schemes have pros and cons in terms of join and windowing costs, and showed that the multi-table scheme provided better performance. The performance gap was resulted from heavier cost of the

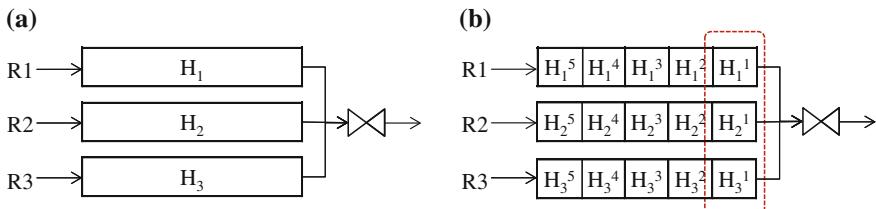


Fig. 1 Hash table organization of **a** the single-table scheme (left) and **b** the multi-table scheme (right) for *Q1*

single-table scheme that is required to find and discard expired tuples from the hash tables. In this paper, we propose a method to reduce the cost to deal with expired tuples in the single-table scheme, and show that the modified algorithm provides better performance than the multi-table scheme.

2 Preliminaries

In this section, we describe the single-table and multi-table schemes for sliding window join. Suppose we want to join m input data streams over a common attribute A . Let the i th input stream be S_i ($1 \leq i \leq m$) and its window size be W_i (i.e., the value of a RANGE parameter). At time t , a tuple s belongs to a windowed substream $S_i[W_i]$ if s has arrived on S_i in the time interval $(t - W_i, t]$. An m -way window equijoin can be represented as $S_1[W_1] \bowtie_A S_2[W_2] \bowtie_A \dots \bowtie_A S_m[W_m]$. The output of the join consists of all m -tuples, (s_1, s_2, \dots, s_m) , satisfying $s_1 \cdot A = s_2 \cdot A = \dots = s_m \cdot A$, where $s_i \in S_i[W_i]$.

In this paper, a discrete time domain is assumed. Unless otherwise stated, time is denoted in seconds. All input streams are timely synchronized. The scope of this paper is limited to window join where SLIDE parameters are explicitly defined in window specifications. Join without SLIDE parameters have been discussed in many previous studies including [8, 9], and the single-table scheme can be used in this case. To simplify the discussion, we only consider the windows whose SLIDE values are all equal to T , for some $T \in \mathbb{Z}$. Their RANGE values are also assumed to be equal to W ($W \in \mathbb{Z}$), satisfying $W \bmod T = 0$. The windows in $Q1$ are an example.

Algorithm 1 describes the single-table scheme, where $f(\cdot)$ denote a hash function and H_i denote the hash table for storing tuples arriving on S_i . Each entry of H_i has a list of tuples with the same join attribute value. To get the list of tuples with hash value v from H_i , a function $get(v)$ can be used. Below, L_i denotes the list of tuples returned from H_i .

Algorithm.1 *Single-table scheme*

Whenever T seconds expire (at time t) ...

1. (Join) For each tuple s arriving on S_k for the last T seconds
 - 1.1. (Hash) $v \leftarrow f(s)$, and insert s to L_k
 - 1.2. (Probe) For each S_i ($1 \leq i \leq m, i \neq k$)
 - 1.2.1. $L_i \leftarrow H_i.get(v)$
 - 1.2.2. If L_i is empty, go to Step 1.4
 - 1.3. (Output) Generate a cross-product of L_i ($1 \leq i \leq m$)
 - 1.4. (Add) Insert s to H_k
2. (Windowing) For each tuple s in H_i ($1 \leq i \leq m$),
 - 2.1. If $s.ts < t - W + T$, remove s from H_i

For each tuple $s \in S_k$ arriving for the previous T seconds, the algorithm calculates the hash value v and initiates L_k with the tuple s (Step 1.1). Then, hash tables of all other streams are scanned to find partners for s (Step 1.2). If there is no partner, producing output is skipped (Step 1.2.2). If all tables have partners for

s and all L_i s are filled with them, the algorithm generates a cross-product of L_i s as join results (Step 1.3). The tuple s is then added to its hash table (Step 1.4).

After the join (Step 1) is finished, expired tuples are found and discarded from the hash tables. To see which tuples are outdated, arrival timestamps of all tuples in H_i are checked (Step 2.1). If a tuple's timestamp, denoted ts , is smaller than $t - W + T$ at a certain time t , the tuple is considered outdated and is removed.

In the multi-table scheme, a hash table is allocated for each set of tuples arriving for T seconds. For simplicity, the notion of *window panes* [7] (also, called *basic windows*) is adopted. A window pane is the largest disjoint segment of a window. The number of panes in a window can be calculated as W/T . For example, a window in $Q1$ can be divided into five panes, each of which is 1 min long.

At time t , the id of a window pane can be obtained by t/T . Let the j th pane of stream S_i be P_i^j . Then, a tuple s belongs to P_i^j if its arrival time is in the interval of $[jT, (j + 1)T]$. For $Q1$, whose slide interval is 60 s, P_i^0 will consist of tuples whose arrival timestamps are between 0 and 59 s, P_i^1 will have tuples with timestamps between 60 and 119 s, and so on.

Based on the notion of panes, the window join of the multi-table scheme can be described as Algorithm 2. Below, n is the number of panes, such that $n = W/T$. A hash table for tuples in P_i^j is denoted as H_i^j ($j \in \mathbb{Z}, j \geq 0$), and τ is used to denote an index of the latest window pane.

Algorithm.2 *Window join for the multi-table scheme*

Whenever T seconds expire (at time t) ...

1. (Join) For each tuple $s \in P_k^\tau$,
 - 1.1. (Hash) $v \leftarrow f(s)$, and insert s to L_k
 - 1.2. (Probe) For each S_i ($1 \leq i \leq m, i \neq k$)
 - 1.2.1. Clear L_i
 - 1.2.2. For each H_i^j ($\tau - n < j \leq \tau$),
 - 1.2.2.1. $L_i \leftarrow H_i^j.get(v)$
 - 1.2.3. If L_i is empty, go to Step 1.4
 - 1.3. (Output) Generate a cross-product of L_i ($1 \leq i \leq m$)
 - 1.4. (Add) Insert s to H_k^τ
2. (Windowing) Discard the oldest tables $H_i^{\tau-n+1}$ ($1 \leq i \leq m$)

The above algorithm is different from Algorithm 1 in two respects. The first is in Step 1.2, where n hash tables are scanned for each stream S_i . The result from each table scan is added to L_i (Step 1.2.2.1). The second is in Step 2, where the windowing is performed simply by discarding hash tables of the oldest window panes.

3 Proposed Method

Among the two schemes, the multi-table scheme provides better performance in dealing with expired tuples. This is because windowing can simply be done by discarding m hash tables in this scheme. On the other hand, its join cost becomes

higher. In the multi-table scheme, $n(m - 1)$ table scans are required to join a tuple, whereas the single-table scheme requires only $(m - 1)$ table scans for each tuple. Regarding this, our previous work [5, 6] showed that the windowing cost (tuple expiration handling cost) is more dominant than the join cost in overall performance.

On the other hand, if the windowing cost can be reduced, the single-table scheme can provide better performance. Inefficiency of the single-table scheme is due to the brute-force scanning of the hash tables (to check whether each of tuples is expired), as shown in Algorithm 1. To resolve this issue, we group IDs of tuples according to their window panes, and use them to find expired tuples selectively from the hash tables. In this approach, whenever the window slide interval elapses, tuple IDs included in the oldest window pane in memory are used to extract expired tuples from the tables.

Let Δ_i^j denote the set of tuple IDs in the j th window pane of stream S_i . Then, the algorithm utilizing Δ_i^j for efficient windowing can be described as follows. Below, $\Delta_i^{\tau-n+1}$ denotes the set of tuple IDs in the oldest window pane kept in memory, where τ is an index of the latest window pane.

Algorithm.3 *Updated single-table scheme*

Whenever T seconds expire (at time t) ...

1. (Join) Same as Algorithm 1

2. (Windowing) For each tuple ID δ in $\Delta_i^{\tau-n+1}$ ($1 \leq i \leq m$),

 2.1. $H_i.\text{remove}(\delta)$

To check the performance, Algorithms 1–3 were implemented in Java. A window join operator is located in a server side and is organized to receive tuples from three remote clients. A tuple arrival rate, denoted λ , was set to 10,000 tuples per second in each stream. The size of an input tuple was about 120 bytes. The experiments were conducted on an Intel Core i3-2120 3.3 GHz machine, running Windows 7 with 4G main memory.

Figure 2 shows the performance of the three algorithms, where the proposed method is denoted as the *updated scheme*. Figure 2a shows the average time to join λ tuples in each algorithm, while Fig. 2b compares the average times to finish windowing tuples. In this result, the one-table and updated scheme provided better

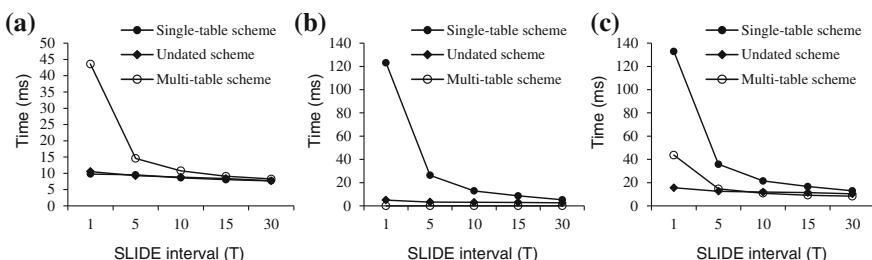


Fig. 2 Performance of single-table, multi-table, and proposed (updated) schemes, when $0.1 \leq \sigma \leq 0.2$. **a** Average join time. **b** Average windowing time. **c** Average execution time

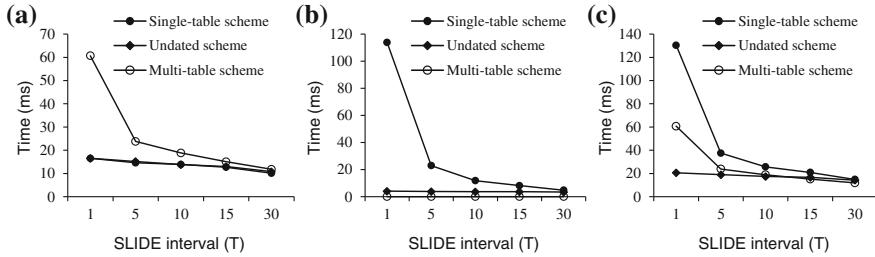


Fig. 3 Performance of single-table, multi-table, and proposed (updated) schemes, when $0.5 \leq \sigma \leq 0.8$. **a** Average join time. **b** Average windowing time. **c** Average execution time

performance in terms of the join cost. This is because the number of hash table scans in Algorithms 1 and 3 is smaller than that of Algorithm 2. On the other hand, the multi-table scheme showed smaller windowing cost because tuple expiration can be done simply by discarding the oldest hash tables.

Figure 2c shows the sum of the join and windowing times in each algorithm. In the figure, the updated scheme shows the best performance when $T \leq 10$, whereas the multi-table scheme is the best in other cases. This implies that reduction of the windowing cost in the proposed method does not give significant influence on overall performance when join time is relatively small.

We increased the *join selectivity* (i.e., number of join outputs/number of inputs), denoted σ , then monitored its influence on the performance (Fig. 3). In this experiment, the proposed method provides better performance than the others in more test cases (i.e., when $T \leq 15$), compared with the results shown in Fig. 2. The above results show that the proposed method can properly be used when a window slide interval is small or when the join selectivity is relatively large.

4 Conclusion

In this paper, we proposed a method to reduce the windowing cost of the single-table scheme for efficient sliding window join. The basic idea is to group IDs of tuples according to their window panes, and use them to find expired tuples selectively from hash tables. In this approach, whenever the window slide interval elapses, tuple IDs included in the oldest window pane in memory are used to extract expired tuples from the tables. Our experimental results show that the proposed method provides better performance than the existing single-table and multi-table schemes when a window slide interval is small or when the join selectivity is relatively large.

References

1. Kim, H.G., Kim, M.H.: A review of window query processing for data streams. *J. Comput. Sci. Eng.* **7**(4), 220–230 (2013)
2. Fernandez, R., Migliavacca, M., Kalyvianaki, E., Pietzuch, E.: Integrating scale out and fault tolerance in stream processing using operator state management. In: Proceedings of the 2013 ACM SIGMOD, pp. 725–736 (2013)
3. Lam, W., et al.: Muppet: MapReduce-style processing of fast data. *Proc. VLDB Endowment* **5**(12), 1814–1825 (2012)
4. Zhang, R., et al.: A highly optimized algorithm for continuous intersection join queries over moving objects. *VLDB J.* **21**, 561–586 (2012)
5. Kim, H.G., Park, Y.H., Cho, Y.H., Kim, M.H.: Time-slide window join over data streams. *J. Intell. Inf. Syst.* **43**, 323–347 (2014)
6. Kim, H.G.: A structure for sliding window equijoins in data stream processing. In: Proceedings of the 16th IEEE CSE, pp. 100–103 (2013)
7. Li, J., et al.: Semantics and evaluation techniques for window aggregates in data streams. In: Proceedings of the 2005 ACM SIGMOD, pp. 311–322 (2005)
8. Viglas, S.D., Naughton, J.F., Burger, J.: Maximizing the output rate of multi-way join queries over streaming information sources. In: Proceedings of the 29th VLDB Conference, pp. 285–296 (2003)
9. Kwon, T.H., Lee, K.Y., Kim, M.H.: Load shedding for multi-way stream joins based on arrival order patterns. *J. Intell. Inf. Syst.* **37**, 245–265 (2011)

The Effect of the User Interface Design of Smartphone Applications on Users' Individual Experiences of Performance

Wonjin Jung

Abstract Even though an awareness of the necessity of studying the importance of User Interface (UI) design in the development of smartphone applications has risen among researchers, a current review of the relevant IS literature reveals the existence of a scant understanding of the topic in empirical research terms. Such a review shows that only a small amount of empirical findings have been published on the effect that the UI design of smartphone applications has on the usability of applications, as well as users' individual experiences of performance with a given application. This study aims to examine: (1) the direct effect of UI design on users' individual experiences of performance with a given application, (2) the indirect effect of UI design on users' individual experiences of performance through a mediating variable, which is the usability of applications as perceived by users. The data collected through a survey was analyzed by Structural Equation Modeling (SEM). The results of the analysis show the significant effect of the UI design on both perceived usability and users' individual experiences of performance. In addition, the perceived usability of applications also had a positive impact on users' individual experiences of performance.

Keywords Smartphone • Application • Design • Usability • Performance

1 Introduction

The number of mobile applications in use worldwide has skyrocketed since the iPhone was introduced onto the market in 2007. In 2014, there were over one million applications available in both Apple's App Store and Google Play [5–9]. One of the main reasons for the rapid penetration of mobile applications is that they are useful in our daily lives. They are not only used for voice calls, but also for a

W. Jung (✉)

The School of Business and Economics, Dankook University, 152, Jook-Jun-Ro, Soo-Ji-Goo, Young-In, Kyung-Ki-Do 448-701, Korea
e-mail: jungw@dankook.ac.kr

wide variety of other purposes including online search, the exchange of text messages, photography, the playing of music, the viewing of TV programs, the playing of games, and online shopping, among others; furthermore, the use of such applications is location independent, which means that they can be used in external environments or while in transit. The capabilities of smartphone applications are continually developed on a constant basis.

Through the introduction of innovative technological advances in the smartphone industry, today's smartphone applications provide greater functionality. However, corresponding concerns regarding the usability of smartphone applications have also gained prominence. Along with the addition of innovative functions, applications have become increasingly complicated. In general, complicated applications are not only less usable, but are also difficult to understand. Moreover, users believe that such a reduction in usability will have a negative impact on their individual experiences of performance with the given application.

Smartphone applications therefore need to be usable, which means that they should be easy to both understand and use. Otherwise, users will become confused when they are faced with a complicated application—in no way does overloading an application with functions improve its usability. For this reason, UI design is considered critical in the development of any smartphone application.

Even though studying the importance of UI design in the development of smartphone applications has become an increasingly pertinent need among researchers, a comprehensive literature review reveals that only a small amount of research on UI design in relation to smartphone applications exists in the IS literature. Specifically, there is a lack of empirical research regarding the direct effect of UI design of smartphone applications on users' individual experiences of performance with a given application. In addition, the indirect effect of UI design on individual experiences of performance through a mediating variable, which is perceived usability, has also been seldom explored. This study therefore aims to examine: (1) the direct effect of UI design on users' individual experiences of performance with a given smartphone application, (2) the indirect effect of UI design on users' individual experiences of performance through a mediating variable, which is the usability of applications as perceived by users.

In sum, the literature review and the discussion above imply that the UI design of smartphone applications appears to be critical not only for users' individual experiences of performance, but also for the perceived usability of applications. This study therefore proposes the following hypotheses for empirical examination.

H1: The UI design of smartphone applications positively affects the usability of the applications as perceived by smartphone users

H2: The usability of smartphone applications as perceived by smartphone users positively affects their individual experiences of performance with a given application

H3: The UI design of smartphone applications as perceived by smartphone users positively affects their individual experiences of performance with a given application

2 Research Methodology, Data Analysis, and Results

This study aims to examine the effect of the UI design of smartphone applications on smartphone users' individual experiences of performance with a given application through a mediating variable, which is the perceived usability of the application. A survey was conducted to collect data. A total of 236 students and practitioners volunteered to take part in the survey: 66.9 % were undergraduate students majoring in business administration, economics, or computer science at three universities in Korea, and 33.1 % of the participants were practitioners. Of the participants, 50.4 % were male and 49.6 % were female, and 80.1 % were aged between 20 and 29 years. The application type that 61.9 % of the participants had used immediately before answering the survey questions was social networking- and communication-related.

Structural Equation Modeling was used to test the proposed research model and SPSS Statistics with AMOS ver. 18 was the software used for the analysis. First, the proposed measurement model was tested by examining the reliability of all individual instrument items that were surveyed for the latent variables (see Table 1). In this respect, the loadings of the items on their respective constructs should be 0.6 or higher [1, 2]. The results of analyses showed that all of the loadings were 0.6 or higher, thus indicating adequate reliability (see Table 2).

Next, the convergent validity of all latent variables was tested by examining the composite reliability (CR) and the average variance extracted (AVE) of the

Table 1 Survey items for latent variables

Latent variables		Questions
User interface (UI) design	DS1	The application that I used most recently provided information and content consistently in terms of structure
	DS2	The application that I used most recently was simple enough to easily access a certain function in the application
	DS3	The meanings of all of the graphic images including buttons and icons in the interface of the application that I used most recently were easy enough to understand
Perceived usability	PU1	The functions that I tried to find in the application were quickly found
	PU2	The information that I tried to find in the application was easily found
	PU3	The information that I tried to find in the application was provided quickly
Individual experience	PF1	The smartphone application that I used most recently helped me to easily achieve my personal goals
	PF2	The smartphone application that I used most recently was the most efficient way to achieve my personal goals
	PF3	The smartphone application that I used most recently was generally beneficial for me

Table 2 Standardized regression weights of observable variables, composite reliability (CR), and average variance extracted (AVE)

Latent variables	Estimates	Variance C.R.	Composite reliability	AVE
User interface design	.667	8.634	.793	.506
	.714	7.937		
	.750	7.262		
Perceived usability	.868	5.837	.905	.667
	.804	7.744		
	.776	8.328		
Individual performance	.809	9.597	.961	.754
	.811	9.576		
	.897	7.884		
	.947	5.097		

constructs in the model. To satisfy the requirements for the convergent validity, all constructs should have the values of CR and AVE be greater than 0.7 and 0.5 respectively. The values for CR and AVE were not provided in the system, so they were manually calculated with the formulas suggested by Fornell and Larcker [3] and Hair et al. [4]. The results showed that the values of CR for all constructs were of 0.7 or higher than the recommended cutoff of 0.7. In addition, all constructs also had the values of AVE equal or greater than the recommended tolerance of 0.5 (see Table 3). Therefore, the constructs in the proposed model demonstrated satisfactory convergent validity.

Then, the discriminant validity of the constructs was also examined to test the proposed measurement model. To meet the requirements for the discriminant validity, the square root of all constructs' AVEs should be greater than the correlations with other constructs in the model [2]. The results of analyses showed that this was the case for each construct in the model (see Table 3). Thus, this study also confirmed the discriminant validity for the constructs in the model.

The goodness of fit was evaluated to test the structural model. The indices for the goodness of fit include χ^2/df , GFI, AGFI, NFI, TLI, CFI, and RMSEA, and the results were as follows: $\chi^2/df = 1.519$, GFI = .963, AGFI = .936, NFI = .967, TLI = .984, CFI = .988, and RMSEA = .047, which indicate that the model has a fairly good fit in general.

Finally, to test the proposed structural model, the significance and the strength of the relationships between the variables were assessed by examining the estimates of the path coefficients. User interface design had a significant influence on perceived

Table 3 Correlation coefficient value between constructs and AVE

Constructs	AVE	σ^2	σ^2	σ^2
User interface design	.506	.367	.462	1.000
Perceived usability	.667	.377	1.000	
Individual performance	.754	1.000		

Table 4 Hypothesis test

	Paths	Coeff.	Stand. Coeff.	P	Results
H1	User interface design → perceived usability	.642	.606	***	Accept
H2	Perceived usability → individual performance	.339	.320	***	Accept
H3	User interface design → individual performance	.544	.485	***	Accept

usability ($\beta = .642$, $p = .000$) as well as on individual performance ($\beta = .544$, $p = .000$). Perceived usability also had a positive impact on individual performance ($\beta = .399$, $p = .000$). Therefore, all hypotheses were supported. Table 4 below shows the results of the analysis.

3 Discussion and Conclusion

This study empirically examined the direct effect of UI design on smartphone users' individual experiences of performance, as well as the indirect effect through a mediating variable that is the perceived usability of an application. The results of this study show the significant effect of UI design on perceived usability as well as on users' individual experiences of performance. In addition, the perceived usability of an application also had a positive impact on the users' individual experiences of performance.

In terms of the contribution that this study makes to the IS literature, this study explored the following UI design attributes in a smartphone application context: simplicity, consistency, and metaphor. Thus, smartphone application users assess these three design attributes during use and, based on the assessment of these three attributes, users perceive different levels of usability. According to the results of this study, when the usability of applications as perceived by users was low, their individual experiences with the given application's performance were also low. Alternatively, when the users' assessments were high, their individual experiences of the application's performance were also high. Therefore, it is important to meet users' standards and expectations in each of the above UI-design categories to escalate or to maintain the level of the users' individual experiences of performance at high. These findings highlight the importance of UI design, especially in terms of the three design attributes, not only for the usability of applications, but also for the users' individual experiences of an application's performance.

The results of this study also suggest that UI design should be simple and consistent for smartphone applications. Recently, a large amount of functions and information have been added to smartphone applications including the overuse of metaphors in the application interface. These make the applications more

complicated and confusing, resulting in a negative impact on perceived usability and individual experiences of performance. Therefore, this study also proposes to practitioners the importance of minimizing complicating and confusing factors in the UI of applications, whereby the simplicity, consistency, and understandability of the UI is maintained. In turn, adherence to this proposal leads to the improvement of usability and the enhancement of the individual experiences of the performance of a given application.

Finally, since this study shows that the UI of applications has both a direct and indirect impact on usability, as well as on users' individual experiences of performance, we also suggest that developers seek to not only improve their UI design, but to also utilize it as a competitive strategy to differentiate their applications.

References

1. Barclay, D., Higgins, C., Thompson, R.: The partial least squares (PLS) approach to causal modeling: personal computer adoption and use as an illustration. *Technol. Stud.* **2**, 285–324 (1995)
2. Chin, W.W.: The partial least squares approach for structural equation modeling. In: Marcoulides, G.A. (ed.) *Modern Methods for Business Research*, pp. 295–336. Lawrence Erlbaum, Mahwah (1998)
3. Fornell, C., Larcker, D.F.: Evaluating structural equation models with unobservable variables and measurement error. *J. Mark. Res.* **18**, 39–50 (1981)
4. Hair, J.F., Black, B., Babin, B., Andersong, R.E., Tatham, R.L.: *Multivariate Data Analysis*, 6th edn. Pearson Prentice Hall, Upper Saddle River (2006)
5. Keach, S.: Microsoft says windows phone now touts 300,000 Apps. <http://www.t3.com/news/microsoft-says-windows-phone-now-touts-300000-apps>
6. Perez, S.: Mobile App usage increases in 2014, as mobile web surfing declines. <http://techcrunch.com/2014/04/01/mobile-app-usage-increases-in-2014-as-mobile-web-surfing-declines/>
7. Perez, S.: iTunes App store now has 1.2 million Apps, has seen 75 billion downloads to date. <http://techcrunch.com/2014/06/02/itunes-app-store-now-has-1-2-million-apps-has-seen-75-billion-downloads-to-date/>
8. Reisinger, D.: Worldwide smartphone user base hits 1 billion. <http://www.cnet.com/news/worldwide-smartphone-user-base-hits-1-billion/>. Retrieved 25 Oct 2014
9. Yang, H.C.: Bon appétit for Apps: young american consumers' acceptance of mobile applications. *J. Comput. Inf. Syst.* **53**, 85–96 (2013)

Data Hiding for H.264/AVC Based on the Motion Vector of 16 Grids

Cheng-Hsing Yang, Yih-Kai Lin, Chun-Hao Chang and Jin-Yi Chen

Abstract Due to the development of digital techniques and the Internet, lots of digital multimedia are created, stored, and spread over the Internet. Because digital video can be modified easily and copied illegally, we need to pay attention to the issues of data security and copyright protection during transmission. Digital watermark, which embeds meaningful marks or words into the digital video, is a kind of methods used as the video copyright protection or the secret communication. In this paper, we propose a data hiding technique specific to the H.264/AVC video format based on the motion vectors. We achieve the purpose of data hiding by modifying motion vectors to their neighboring points within 16 grids. Experimental results show that our method can maintain a good video quality after the secret data are embedded. Moreover, our method has a large embedding capability and the embedded data can be easily extracted.

Keywords H.264/AVC · Motion vector · Data hiding · Copyright protection

1 Introduction

Due to the rapid change of the information technology and the development of the Internet, data can be transmitted more quickly and easily. However, it also has brought about network security issues. Therefore, the data security and the

C.-H. Yang (✉) · Y.-K. Lin · C.-H. Chang · J.-Y. Chen

Department of Computer Science, National Pingtung University, Pingtung 900, Taiwan
e-mail: chyang@mail.nptu.edu.tw

Y.-K. Lin
e-mail: yklin@mail.nptu.edu.tw

C.-H. Chang
e-mail: bm102101@mail.nptu.edu.tw

J.-Y. Chen
e-mail: bbe103119@mail.nptu.edu.tw

copyright protection of the multimedia have become important issues. The information hiding or the digital watermark, which embeds meaningful graphs or words into multimedia, is a kind of methods used as copyright protection or integrity authentication of multimedia.

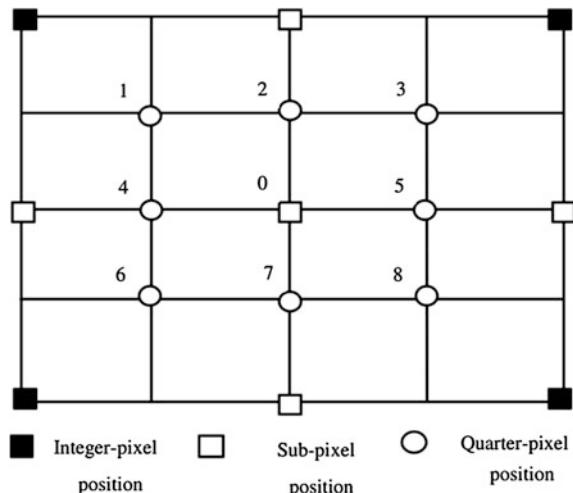
H.264/AVC video format is the latest video coding standard jointly developed by ITU-T and ISO/IEC. In the recent information hiding technologies of H.264, they can be categorized as follows: to embed into the intra-prediction modes or the coefficients of the frequency domain [1–3], to embed into the MV (Motion Vectors) of the inter-prediction [4–6], and to embed by modifying the CAVLC (Context-Based Adaptive Variable Length Coding) codes [7, 8].

In this paper, we propose a method based on the characteristic of the motion vector to hide the information. In order to hide the information, we modify motion vectors to their neighboring points according to the hidden secret bits. In addition, we add some limited conditions to avoid modifying the motion vector with value 0, thus it can reduce the impact on the size of the video file after embedding. The remainder of this paper is organized as follows. Section 2 introduces some information hiding techniques of H.264. In Sect. 3, our scheme is proposed. In Sect. 4, we present the experimental results. Finally, the conclusion is proposed in Sect. 5.

2 Related Works

In 2010, Zhu et al. [5] proposed an information hiding technique of H.264 based on the quarter-pixel motion estimation. As shown in Fig. 1, they divided the search points into two groups M and N by their parities, where $M = \{0, 1, 3, 6, 8\}$

Fig. 1 Search graph of the quarter-pixel motion estimation



and group $N = \{2, 4, 5, 7\}$. For an embedded bit w , if $w = 0$, the searching group is M ; otherwise, the searching group is N .

The embedding process of a bit w by modifying the search points of the quarter-pixel motion estimation is as follows:

1. Choose the searching group S from M or N according to the bit w .
2. Use the rate distortion optimization to find the best search point i in S .
3. Modify the MV to the search point i .

In 2011, Hu et al. [6] proposed an information hiding algorithm, which embeds 2-bit data by modifying the LSB bits of each MVD (Motion Vector Difference). This method has little impact on the quality of video. Furthermore, this method can extract the hidden information quickly because it only needs to decode the MVD . If a LSB bit of MVD is 0, a bit 0 is extracted and if a LSB bit of MVD is 1, a bit 1 is extracted.

3 Our Proposed Method

Our data hiding method embeds data by modifying the motion vectors to their neighboring points of the 4×4 region. Give a motion vector $(MV_x, MV_y) = (u, v)$, where its neighboring points of the 4×4 region are shown in Fig. 2. The embedding method is to modify (u, v) to a neighboring point. The detailed algorithm is as follows.

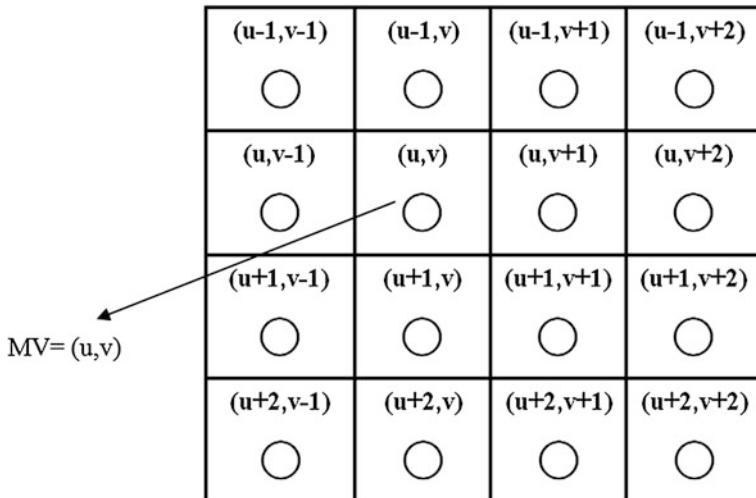


Fig. 2 The 4×4 neighboring points of a motion vector (u, v)

Algorithm EmbeddingInput: Original video H , secret data S Output: Watermarked video H' Step 1: Read a motion vector (u, v) from H and embed four-bit data of S by the following two steps.Step 2: For the 4×4 neighboring points of (u, v) , calculate the function values of the following equation:

$$F(u', v') = [4 \times (u' \bmod 4) + (v' \bmod 4)] \bmod 16, \quad (1)$$

$$\text{where } u - 1 \leq u' \leq u + 2, v - 1 \leq v' \leq v + 2.$$

Step 3: Read 4-bit secret data from S and transform it into a decimal value val .Choose the neighboring point (u', v') satisfying $F(u', v') = val$. Change motion vector (u, v) into (u', v') .Step 4: Continue to execute the above steps until all secret data are embedded. Output watermarked H' .

We add some limit rules to improve the bit-rate of our embedded results. Note that H.264 uses special encoding mode to shrink the bit-rate when the motion vector $(MV_x, MV_y) = (0, 0)$. If the motion vector $(MV_x$ or $MV_y)$ is modified from zero values to nonzero values during the data hiding, it would lead to the significant increase of the bit-rate. So, we skip to embed secret data into the original MV_x or MV_y equal to zero. Let (u, v) be a motion vector, the embedding method with the limit rules is as follows.

1. If u is 0 and v is not 0, use v to hide 2-bit data by modifying v into v' which has the following function value equal to the value of the 2-bit data:

$$F(v') = (v' \bmod 4), \quad (2)$$

where $-1 \leq v' \leq v + 2$. If the selected v' equals to 0, v' need to be changed into a nonzero value by increasing 4 or decreasing 4.

2. If u is not 0 and v is 0, use u to hide 2-bit data similar to Rule 1.
3. If u is not 0 and v is not 0, use Eq. (1) to hide 4-bit data into (u, v) . If the selected u' and v' equal to 0, u' and v' need to be changed into a nonzero value by increasing 4 or decreasing 4, respectively.

4 Experiment Performance and Analysis

In this section, we implemented our method in the H.264/AVC reference software version JM13.2 [9]. Figure 3 shows the ten QCIF format standard video sequences used in this experiment. The QCIF resolution is 176×144 pixels.

In each video sequence, we take out the first 30 frames to do the experiment. Tables 1 and 2 show the embedding capacity (EC), the bit size of the embedded

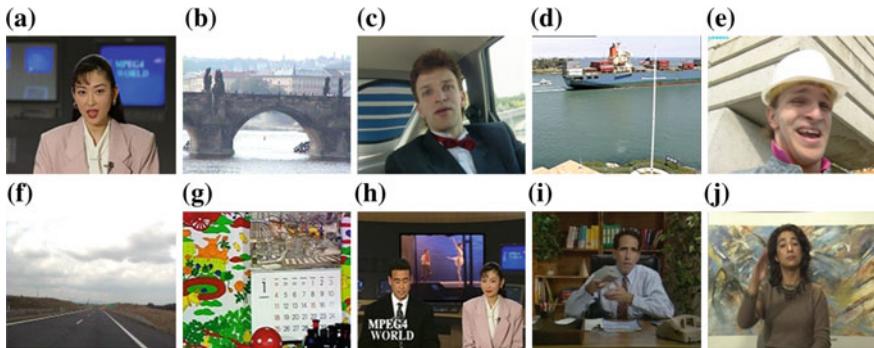


Fig. 3 The original frame: **a** Akiyo, **b** Bridge-Close, **c** Carphone, **d** Containe, **e** Foreman, **f** Highway, **g** Mobile, **h** News, **i** Salesman, **j** Silent

Table 1 Comparisons between our results without the limit rules and the original video

Video sequence	EC (bits)	Video size	BRI (%)	PSNR Y	PSNR U	PSNR V
Foreman	182848	158768	+10.96	+0.04	+0.03	+0.02
Akiyo	183232	55784	+48.39	+0.12	-0.1	+0.03
Carphone	181056	133000	+9.95	+0.03	+0.06	+0.04
Container	182912	66776	+32.87	+0.04	-0.04	+0.01
Highway	181248	86760	+16.91	+0.02	+0.01	+0.01
Mobile	183616	531968	+5.81	+0.03	+0.02	-0.01
News	181696	102968	+23.72	+0.07	+0.01	+0.02
Salesman	183616	112360	+22.79	+0.05	+0.04	+0.04
Silent	181632	109554	+20.18	+0.01	+0.01	0
Bridge-close	183744	136544	+22.10	+0.07	+0.04	+0.01

Table 2 Comparisons between our results with the limit rules and the original video

Video Sequence	EC (bits)	Video size	BRI (%)	PSNR Y	PSNR U	PSNR V
Foreman	124784	154032	+7.65	+1.98	+0.04	+0.17
Akiyo	12694	38560	+2.58	0	+0.02	+0.01
Carphone	102480	128008	+5.82	+0.02	+0.04	+0.11
Container	13244	50232	-0.05	0	-0.03	+0.03
Highway	67240	78200	+5.38	0	0	-0.01
Mobile	148390	526656	+4.75	+0.02	-0.02	-0.03
News	20578	84520	+1.56	0	-0.04	+0.01
Salesman	23944	94432	+3.20	-0.01	0	-0.02
Silent	26196	93320	+2.37	-0.03	-0.01	-0.02
Bridge-close	7944	111600	-0.21	0	-0.01	-0.01

Table 3 The comparisons between our approach and Hu et al. [6]

Video		PSNR Y	PSNR U	PSNR V	BR (%)	EC (bits)
Bridge-close	Proposed	35.01	37.24	37.92	-0.21	7944
	Hu [8]	35.00	37.27	37.96	1.26	6396
Carphone	Proposed	37.55	40.55	41.05	5.82	102480
	Hu [8]	37.50	40.49	40.94	0.6	13260
Foreman	Proposed	38.86	40.70	42.36	7.65	124784
	Hu [8]	36.87	40.65	42.19	1.31	18255
News	Proposed	37.05	39.76	40.37	1.56	20578
	Hu [8]	37.05	39.80	40.34	0.89	7630
Silent	Proposed	36.15	39.00	40.00	2.37	26196
	Hu [8]	36.18	38.99	40.03	0.83	9432

video, and the comparisons between our results and the original video, including Bit-Rate-Increase (BRI) and the change of PSNR. The results show that the PSNR values of the videos decrease or increase slightly. For the dynamic videos, such as Carphone, Foreman, and Mobile, to use our program without the limit rules would be suitable. By contrast, for the static videos, such as Akiyo, News, and Salesman, it is better to avoid modifying the motion vectors with zero values.

One interesting phenomenon can be found in the experimental results: the bit size of the embedded video is smaller than that of the embedded data. For example, in Table 1, the data embedded into Akiyo is 183,232 bits. However, the size of the embedded Akiyo is only 55,784 bits.

We compare the experimental results to the methods proposed by Hu et al. [6] and Zhu et al. [5]. Our results are using the method with the limit rules. Table 3 shows the PSNR values and embedding capacity (EC) of the methods of ours and Hu et al. Our embedding capacities are much better than that of Hu et al. But, our BRIs are slight larger than that of Hu et al. The PSNR values are similar. Table 4 shows the comparisons between the results of ours and Zhu et al. The embedding capacities of our method are much larger than that of Zhu et al. However, our BRIs increase from 1.56 to 7.65 %.

Table 4 The comparisons between our approach and Zhu et al. [5]

Video	Proposed		Zhu et al. [7]	
	BRI(%)	EC (bits)	BRI(%)	EC (bits)
Foreman	+7.65	124784	-0.10	18554
Carphone	+5.82	102480	-0.01	12000
Container	-0.05	13244	+0.2	1688
Highway	+5.38	67240	-0.17	9478
News	+1.56	20578	+0.05	7130
Silent	+2.37	26196	-0.02	7564
Bridge-close	-0.21	7944	+0.03	3456

5 Conclusions

In this paper, we propose a method to realize the technique of information hiding by using the characters of the motion vectors of H.264 inter-prediction. We hide the secret data by modifying motion vectors to their neighboring domains. For each motion vector, it can hide 4-bit data at most. The experimental results show that, there is no obvious change of the video quality after the secret data are embedded into the video by our method. The BRI (change of bit-rate) is obvious increased. However, the sizes of the embedded data are smaller than the increasing sizes of the embedded video.

Acknowledgments This research was partially supported by the Ministry of Science and Technology of the Republic of China under the Grants MOST 103-2221-E-153-005.

References

1. Ma, X.J., Li, Z.T., Tu, H., Zhang, B.: A data hiding algorithm for H.264/AVC video streams without intra-frame distortion drift. In: *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1320–1330, Oct 2010
2. Yang, G., Li, J., He, Y., Kang, Z.: An information hiding algorithm based on intra-prediction modes and matrix coding for H.264/AVC video stream. *AEU—Int. J. Electr. Commun.* **65**, 331–337 (2011)
3. Su, P.C., Chen, W.Y., Shiao, S.Y., Wu, C.Y., Su, A.Y.S.: A privacy protection scheme in H.264/AVC by data hiding. In: *2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pp. 1–7 (2013)
4. Wang, P., Zheng, Z., Ying J.: A novel video watermark technique in motion vectors. In: *ICALIP International Conference on Audio, Language and Image Processing*, 2008
5. Zhu, H., Wang, R., Xu, D.: Information hiding algorithm for H.264 based on the motion estimation of quarter-pixel. In: *2010 2nd International Conference on Future Computer and Communications (ICFCC)*, Vol.1, pp. 423–427 (2010)
6. Hu, L., Wang, R.: Large capacity information hiding for H.264/AVC. In: *2011 International Conference On Electronics, Communications and Control (ICECC)*, pp. 878–882 (2011)
7. Liao, K., Ye, D., Lian, S., Guo, Z., Wang, J.: Lightweight Information Hiding in H.264/AVC Video Stream. In: *International Conference on Multimedia Information Networking and Security, 2009. MINES'09.*, pp. 578–582, 18–20 Nov 2009
8. Li, X., Chen, H., Wang, D., Liu, T., Hou, G.: Data hiding in encoded video sequences based on H.264. In: *2010 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT)*, Volume 6, pp. 121–125 (2010)
9. JM. <http://iphome.hhi.de/suehring/tm/> Accessed 18 Jun 2014

Idle-Time Processing in Time-Slide Window Join

Hyeon Gyu Kim

Abstract When memory is exhausted from bursty tuple arrivals, some portion of tuples in memory can be evicted to disk, and those tuples can be recalled to complement earlier results when a system becomes idle. This paper presents a method for the idle-time processing when sliding window join queries are given to deal with continuous data streams. Regarding this, we discuss (i) how to determine the priority of tuples for victim selection when memory is not enough and (ii) how to ensure that no result tuple is generated twice when join is conducted with disk-resident tuples in idle time.

Keywords Window join • Sliding windows • Idle-time processing • Data streams

1 Introduction

Sliding window join is common in processing continuous data streams [1, 2]. For example, consider a query to identify packets flowing in a network path over the last 5 min; the result needs to be updated every minute. Suppose that the network path consists of three routers, R1, R2 and R3. To identify packets in the network path, an equijoin can be performed among the three router streams over a common attribute *pid*, denoting a packet ID. This requirement can be described as a join

H.G. Kim (✉)

Department of Computer Engineering, Sahmyook University, Seoul 139-742
Republic of Korea

e-mail: hgkim@syu.ac.kr

query shown in *Q1*. Below, RANGE and SLIDE denote the size and the slide interval of a window, respectively. For convenience, the window slid by time (i.e., updated periodically) is referred to as a *time-slide window* [3] in this paper.

Q1. SELECT *
 FROM R1[RANGE 5 mins, SLIDE 1 min],
 R2[RANGE 5 mins, SLIDE 1 min],
 R3[RANGE 5 mins, SLIDE 1 min]
 WHERE R1.pid = R2.pid and R2.pid = R3.pid

If the arrival rate of input tuples is too high, memory might be insufficient to accept all input tuples in the windows. To cope with the arrival rate and produce results in a timely manner, many existing approaches simply discard some portion of tuples from the memory to shed load [4, 5]. On the other hand, it is also possible to save the evicted tuples into disk for later use. Those tuples can be recalled to complement the previous join results when a system becomes idle (e.g., when one or more stream sources experience delay).

In this paper, we discuss the idle-time processing when time-slide windows are used to join multiple data streams. One characteristic of time-slide windows is that adjacent windows can overlap, from which duplicate processing can be incurred [6]. This happens when the window slide interval is set smaller than the window size (e.g., *Q1*) in order to help users to receive query responses more frequently. When overlapping windows are used, victim selection and idle-time processing can be performed in a different manner, compared with existing approaches.

The remaining part of this paper is organized as follows. Section 2 discusses online window join where join is conducted over tuples kept in memory. It also discusses how to determine a tuple's priority for victim selection when memory is not enough. Section 3 presents the idle-time processing where join is conducted over disk-resident tuples. A method to avoid duplicate join results is also discussed. Section 4 concludes the paper.

2 Online Window Join

In our query model, a discrete time domain is assumed, and a join is performed over m input streams. For each stream, S_i ($1 \leq i \leq m$), a variable number of tuples may arrive in each unit of time. The windows considered here are *time-based*, where W_i denotes a window size of stream S_i in time units (e.g., seconds). At time t , a tuple s belongs to $S_i[W_i]$ if s has arrived on S_i in the time interval $(t - W_i, t]$. For join, we consider an m -way sliding window equijoin [7], which can be denoted as $S_1[W_1] \bowtie_A S_2[W_2] \bowtie_A \cdots \bowtie_A S_m[W_m]$, where A is a common attribute for join. The output of the join consists of all m -tuples, denoted (s_1, s_2, \dots, s_m) , satisfying $s_1.A = s_2.A = \cdots = s_m.A$, where $s_i \in S_i[W_i]$.

Note that windows considered here are time-slide windows whose boundaries are updated periodically (i.e., slid by time). As mentioned above, overlapping windows are common when time-slide windows are used for data stream processing. Let the window slide interval of stream S_i be T_i . Then, if $T_i < W_i$, adjacent windows overlap. To simplify discussion, we only consider the windows such that $W_i = W$ and $T_i = T$ for some integer W and T . $Q1$ can be an example of this window.

2.1 Join with Basic Windows

When overlapping windows are used in a query, duplicate processing may incur. For instance, in $Q1$, all input tuples will join five times. To avoid such duplicate processing, the notion of *basic windows* [8] is adopted in this paper. A basic window can be viewed as the largest disjoint segment of overlapping windows. The size of a basic window can be calculated as $\text{GCD}(W, T)$, where GCD denotes the greatest common divisor. The number of basic windows can be calculated as $W/\text{GCD}(W, T)$.

At time t , the *id* of a basic window can be obtained by t/T . Let the j -th basic window of stream S_i be B_i^j . Then, a tuple s belongs to B_i^j if its arrival time is in the interval of $[jT, (j+1)T]$. For $Q1$, where the slide interval T is given to 60 s, B_i^0 will consist of tuples whose arrival timestamps are between 0 and 59 s, B_i^1 will have tuples with timestamps between 60 and 119 s, and so on.

Figure 1 shows an example of the window join for query $Q1$. Suppose that the latest join has performed at τ minutes from the start of the query. Then, input tuples arriving until $\tau + 1$ min are added to the buffer of B_i^τ (Fig. 1a). At $\tau + 1$ min, the next join is performed for the tuples in B_i^τ . Its result, denoted R^τ , can be represented as follows.

$$R^\tau = \{ B_1^\tau \bowtie S_2[W_2] \bowtie S_3[W_3] \} \cup \{ B_2^\tau \bowtie (S_1[W_1] - B_1^\tau) \bowtie S_3[W_3] \} \cup \{ B_3^\tau \bowtie (S_1[W_1] - B_1^\tau) \bowtie (S_2[W_2] - B_2^\tau) \}$$

After the join, windowing is performed to remove expired tuples. This can be done easily by discarding extents of the oldest basic windows, $B_i^{\tau-4}$, from all input

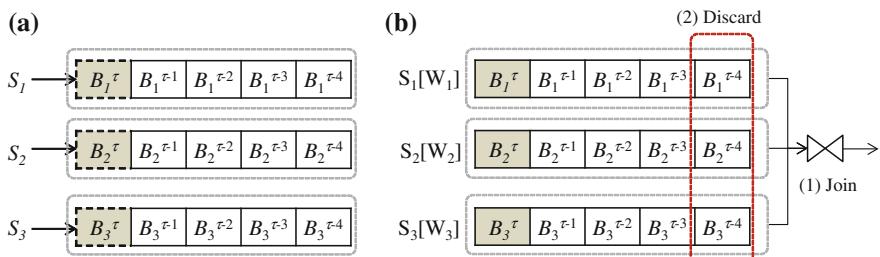


Fig. 1 Time-slide window join for query $Q1$. **a** Before $\tau + 1$ minutes. **b** At $\tau + 1$ minutes

streams (Fig. 1b). If tuples in a basic window are grouped in a single hash table, tuple expiration can be performed in a more efficient way [8].

2.2 Memory Overflow Handling

When memory is exhausted during the join execution, some portion of tuples in memory needs to be moved out to disk. To determine which tuples go out, a mechanism to decide the priority of each tuple is required.

In the proposed method, the priority of a tuple r is determined based on the id of a basic window to which r belongs. This is from intuition that missing of the latest tuples gives more influence to accuracy of query results than missing of earlier tuples. For example, suppose that tuples in B_1^τ are evicted from memory in Fig. 1. Then, join results cannot be complete until the next 4 min. On the other hand, if tuples in $B_1^{\tau-4}$ are evicted, only the latest join results (at time τ) will be incomplete.

In this way, whenever memory is not enough, the proposed method evicts tuples in a basic window with the smallest id from memory. Priorities of all tuples in a basic window are the same. Also, priorities of tuples in windows with the same id, B_i^τ ($1 \leq i \leq m$), are the same. In both cases, any tuples in the windows can be evicted. To simplify discussion, we assume that victim selection and eviction are performed in the unit of basic windows; a smaller unit can also be adopted for fine control, where its discussion is omitted due to lack of pages.

Suppose that B_i^x ($1 \leq i \leq m$, $x < \tau$) is evicted at time τ . Then, the latest basic windows, $B^\tau = \bigcup B_i^\tau$ ($1 \leq i \leq m$), need to be written together into disk in our method. Such *coordinate flushing* is necessary to produce join results from the disk-resident tuples in idle time. For example, suppose that memory becomes exhausted when accepting input tuples in Fig. 1a, and $B_1^{\tau-4}$ is evicted to make a room. Then, the latest windows B^τ should be backup together in order to produce $B_1^{\tau-4} \bowtie B^\tau$ later (i.e., when the system becomes idle). Below, those backup windows are called *coordinate extents*.

3 Idle-Time Processing

To determine when to start the join for disk-resident tuples, the proposed method monitors run-time parameters such as average arrival rate of input tuples per second, average join time per tuple, and average disk access time per tuple. Let these three parameters be λ , μ , and δ , respectively. Since join is performed every T slide interval, total λT tuples arrive in the system during the join. Thus, join time can be calculated as $\mu\lambda T$. In contrast, idle time can be obtained by $T - \mu\lambda T$.

From the idle time, it is possible to estimate the number of tuples that can be loaded into memory and participated in the join. Let the number be n . Then, time to

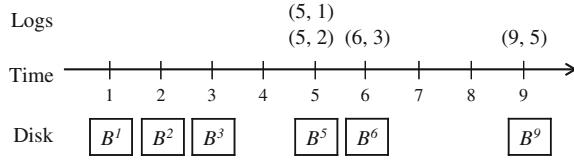


Fig. 2 An example of disk status for query Q1

load and join n disk-resident tuples can be represented as $n(\mu + \delta)$. This time must be smaller than the idle time.

$$n(\mu + \delta) < T - \mu \lambda T$$

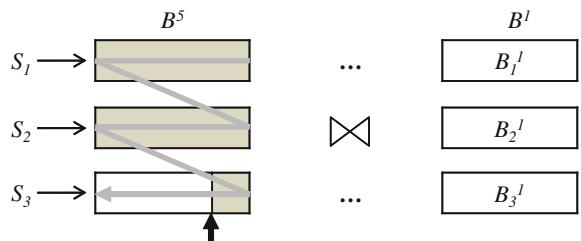
As a solution for n , the largest value satisfying the above inequality is chosen in our method. With this value, we determine which tuples can be transferred into memory. Let $B^{f(x)}$ denote the coordinate extent of a victim B^x in disk. Then, we check whether the size of B^x and $B^{f(x)}$ exceeds n . If their size is smaller than n , tuples of B^x and $B^{f(x)}$ are loaded into memory and participated in the join.

Figure 2 shows an example of disk status to illustrate the idle-time processing for query *Q1*. In the figure, victims and coordinate extents are stored in the unit of basic windows B^x , where $B^x = \bigcup B_i^x (x \geq 1, 1 \leq i \leq m)$. To identify which one is a victim or a coordinate extent, logs of coordinate flushing are maintained, which are called *victim logs*.

A victim log consists of the ids of two basic windows designating a victim and its coordinate extent, respectively. For example, $(5, 1)$ denotes that B^1 is evicted at unit time 5, which also indicates that B^1 and B^5 are stored in disk as a victim and its coordinate extent, respectively. In the figure, from the four times of coordinate flushing, six window extents are stored in the disk. If a chosen victim or a coordinate extent already exists in disk, it is not stored twice. For example, when the second victim B^2 is evicted at time 5, its coordinate extent B^5 is not stored again. Similarly, when the fourth victim B^5 is evicted at time 9, only its coordinate extent B^9 is stored in the disk.

In the proposed method, idle-time processing can be conducted whenever the online join is finished. At that point, we first estimate n , then check victim logs to see how many tuples can be loaded and participated in the join. In the example of Fig. 3, the log $(5, 1)$ is first examined. If the size of B^1 and B^5 is smaller than n , their tuples

Fig. 3 An example of disk status for query Q1



are transferred into memory. Then, the next log is checked to see if its tuples can be transferred. This process continues until the total size of transferred tuples exceeds n . After the transfer is finished, join is performed over the transferred tuples.

Now, let us discuss how to avoid duplicate join results in the idle-time processing; no result tuple must be generated twice. If some tuple is generated twice in the proposed method, this is when a basic window of the tuple goes to disk during the join execution. Otherwise, duplicate tuple generation does not happen; the window is evicted before join or discarded after the join in our method.

The problem can simply be resolved by keeping the position where the last join is performed before the eviction is conducted. To keep the position in a constant way, join should be performed in a predefined order (e.g., Z-order) over multiple input streams shows its example, where the dark arrow denotes the position where the last join is performed before eviction.

The position can be saved in victim logs. For instance, a log $(5, 1, 235)$ denotes that B^1 is evicted at unit time 5 after join for the 235-th tuple in B^5 is finished. Given this log, join can restart from the 236-th tuple in B^5 later, so we can guarantee that no tuple is generated twice even when the eviction occurs during the join execution.

The proposed method to avoid duplicate outputs is different from the existing approaches that utilize tuples' timestamps [9]. Those approaches may provide inaccurate results when many input tuples have the same timestamps; such situation frequently occurs when the arrival rate of input tuples is high. On the other hand, the proposed method is free from the timestamp issue.

4 Conclusion

In this paper, we proposed a new method for the idle-time processing when time-slide windows are used to join multiple data streams. The main idea of the proposed method is to use basic windows as a unit for memory overflow handling and idle-time processing, where a basic window is the largest disjoint segment of overlapping windows. Whenever memory is not enough to accept input tuples, tuples in a basic window with the smallest id kept in memory are evicted into disk. This victim selection is due to that missing of the latest tuples gives more influence to accuracy of query results than missing of earlier tuples. When a system becomes idle, disk-resident tuples are recalled to complement earlier join results. To ensure that no result tuple is generated twice, victim logs are utilized which consists of the ids of victim windows and the position where the last join is performed before eviction.

Acknowledgments This paper was supported by the Sahmyook University Research Fund in 2014.

References

1. Kim, H.G., Kim, M.H.: A review of window query processing for data streams. *J. Comput. Sci. Eng.* **7**(4), 220–230 (2013)
2. Zhang, R., et al.: A highly optimized algorithm for continuous intersection join queries over moving objects. *VLDB J.* **21**, 561–586 (2012)
3. Kim, H.G., Park, Y.H., Cho, Y.H., Kim, M.H.: Time-slide window join over data streams. *J. Intell. Inf. Syst.* **43**, 323–347 (2014)
4. Das, A., Gehrke, G., Riedewald, M.: Approximate join processing over data streams. In: *Proceedings of the ACM SIGMOD*, pp. 40–51 (2003)
5. Kwon, T.H., Lee, K.Y., Kim, M.H.: Load shedding for multi-way stream joins based on arrival order patterns. *J. Intell. Inf. Syst.* **37**, 245–265 (2011)
6. Li, J., et al.: Semantics and evaluation techniques for window aggregates in data streams. In: *Proceedings of the 2005 ACM SIGMOD*, pp. 311–322 (2005)
7. Viglas, S.D., Naughton, J.F., Burger, J.: Maximizing the output rate of multi-way join queries over streaming information sources. In: *Proceedings of the 29th VLDB Conference*, pp. 285–296 (2003)
8. Kim, H.G.: A structure for sliding window equijoins in data stream processing. In: *Proceedings of the 16th IEEE CSE*, pp. 100–103 (2013)
9. Urhan, T., Franklin, M.J.: XJoin: a reactively-scheduled pipelined join operator. *IEEE Data Eng. Bull.* **23**, 42–48 (2000)

A Large-Scale Object-Based Active Storage Platform for Data Analytics in the Internet of Things

**Quanqing Xu, Khin Mi Mi Aung, Yongqing Zhu
and Khai Leong Yong**

Abstract In this paper, we propose a large-scale object-based storage platform, named *Gem*, for data analytics in the Internet of Things (IoT). In *Gem*, a region covered by an IoT network is partitioned into sub-regions, each of which can be identified by a unique ID and known to all participants, which is automatic and economical. *Gem* can preserve object locality using type and location sensitive hashing, as well as dynamically distribute objects among a server cluster to keep load balancing. All data from the IoT can be stored as objects with attributes, methods and policies in Object Store Devices (OSDs). For some applications such as data analytics, application-specific operations are executed by OSDs, and only the results are returned to clients, rather than data files are read by the clients. Thus, the platform *Gem* is able to greatly reduce the overhead of data analytics applications in the IoT.

Keywords Object-based storage • Data analytics • Internets of things

1 Introduction

The Internet of Things (IoT) [1] is the network that connects objects to the Internet through kinds of devices, such as sensors and smart phones. In the IoT, there are a large number of perception devices that continuously and automatically collects

Q. Xu (✉) · K.M.M. Aung · Y. Zhu · K.L. Yong
Data Storage Institute, A*STAR, Singapore, Singapore
e-mail: Xu_Quanqing@dsi.a-star.edu.sg

K.M.M. Aung
e-mail: Mi_Mi_AUNG@dsi.a-star.edu.sg

Y. Zhu
e-mail: ZHU_Yongqing@dsi.a-star.edu.sg

K.L. Yong
e-mail: YONG_Khai_Leong@dsi.a-star.edu.sg

Table 1 Use case examples in IoT

Questions	Queries
How many data files are collected from a device A in Region B from 20/10/2014 to 25/10/2014?	Dir = /10001, Name = A, when = 20/10/2014:25/10/2014
What videos are from two cameras A and B in the same building C from 2 p.m. to 3 p.m. in 28/10/2014?	Dir = /11001/C/A, Dir = /11001/C/B, when = 2 p.m.:3 p.m. in 28/10/2014
Which device text files are accessed within a day?	Type = txt, atime < 1 days
Which videos will be expired and deleted?	Retention time = expired, ctime > 1 month
How much storage do the video files consume in all the devices in a building A?	Sum size where dir = /11100/A, type = video
What is the average degree for all sensors in region A from 10 a.m. to 11 a.m. in 23/11/2014?	Dir = /10011, files = sensors*.txt, when = 10 a.m.:11 a.m. in 23/11/2014

information, leading to a rapid expansion of data scale. With the IoT, all the ordinary physical objects that can be individually addressed are able to exchange information with each other, and ultimately achieve the goal of intelligent recognition and management. The IoT represents the next evolution of the Internet, taking a huge leap in its ability to collect, store, analyze and distribute data. The “things” of the real world are seamlessly integrated into the virtual world, enabling anytime and anywhere connectivity. Data generated from the IoT will be so huge that we have to build a scalable storage platform. In the environment of the IoT, data is from different kinds of devices and represents billions of objects.

Table 1 demonstrates some popular use case examples in the IoT. IoT exponentially increases the volume, variety, and velocity of data. The burden falls on IT to solve data storage, data integration, and data analytics dilemmas caused by the IoT. We cannot use current approaches because the data to be captured, managed, and exploited is even more diverse, and the use cases are even more varied. The data comes from a variety of sensors and cameras attached to many kinds of devices and objects. Object-based storage is utilized for data analytics applications, in which large data sets are scanned for patterns of varying complexity. It can efficiently support many On-Line Analytical Processing (OLAP) operations, with improvements comparable to purely scan-based functions. It allows a large class of data intensive applications to exploit OSDs [2] with changes to only a few database primitives. Many more novel optimizations are explored to improve the scheduling knowledge available when applications are executed close to the OSDs.

2 Gem Overview

In this section, we proceed to illustrate the system architecture of large-scale object-based storage for data analytics in the IoT.

2.1 Region Partitioning

For an administrative region covered by an IoT network, we divide it into a number of sub-regions based on system requirements in a recursive manner for maintenance [3]. The region is first divided into two half sub-regions (R_e and R_w) based on the north/south direction (i.e., longitude), where R_e and R_w are respectively represented by 1 and 0; and then for R_e and R_w , they are also divided into two half sub-regions (R_n and R_s) based on the east/west direction (i.e., latitude), where R_n and R_s are respectively represented by 1 and 0 as well. The above procedure shown in Fig. 1 is recursively processed until the differences in longitude and latitude of a sub-region are both less than given thresholds LO and LA . Consequently, the whole administrative region is divided into multiple geographical sub-regions. Each sub-region is represented by a unique region ID, and each device uses this embedded service to keep the location information of all sub-regions in the whole IoT network. There are three cases: case (a) most of a building is in a region R (e.g., 11011), and this building belongs to R , case (b) halves of a building are in two regions R_e and R_w , and it belongs to R_w (e.g., 10010), case (c) halves of a building are in two regions R_s and R_n , and it belongs to R_n (e.g., 10000).

This device-to-sub-region mapping is known to all devices and proxy servers, so we present object hierarchical namespace as shown in Fig. 2, in which the folders form a hierarchical tree structure composed by five levels. The first three levels indicate the information of space, i.e., region, build and device. The last two levels include the information of time, i.e., year and month. As shown in Fig. 2, all video

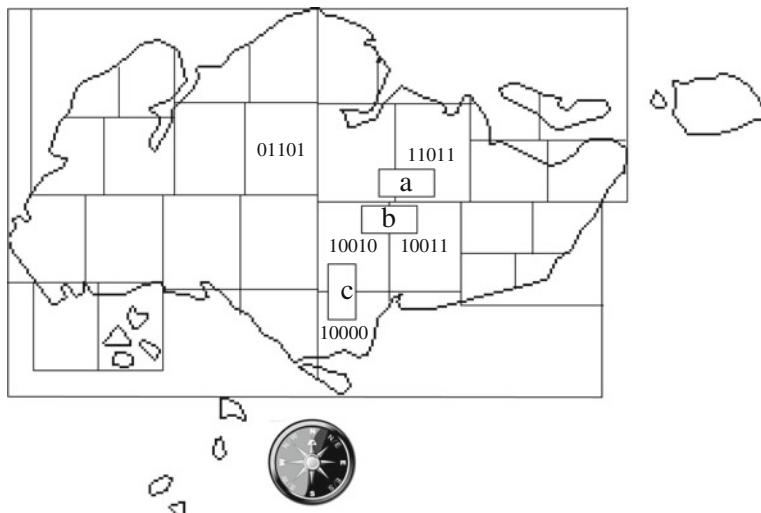


Fig. 1 Region partitioning for Singapore

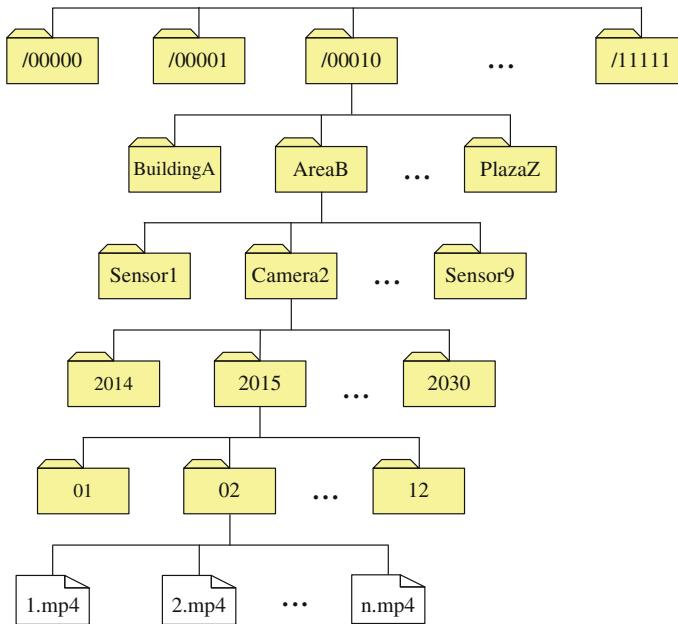


Fig. 2 Object hierarchical namespace

files of a camera listed are generated in February 2015. This architecture could be utilized for distributed file systems, e.g., HDFS [4], in a data center.

2.2 System Architecture

As shown in Fig. 3, we get data from smart city, smart grid, smart building and smart home, and we classify data via data classifier. There are two kinds of data: one is text data, and the other is media data. Text data and media data are stored into our EB-scale storage system via the custom process module. Put/Get connect *Gem* and system interface. *Gem* is a EB-scale storage system that we design, as shown in Fig. 4, where hash-based mapping does not evenly partition the address space into which keys get mapped, causing some OSDs get a larger portion of it. To cope with this problem, virtual OSDs are used as a means of improving load balancing, similar to virtual nodes in [5]. Virtual OSDs make not only re-distribution and cooperative caching [6] become easier, but also scaling out as data grows. When scaling out, more physical OSDs may be added and virtual ones can be moved onto them seamlessly.

Gem can achieve better namespace locality and load balancing by allocating more virtual OSDs per OSD since object IDs are not uniformly distributed. The data structures are typically not so expensive from the perspective of space, thus it is not

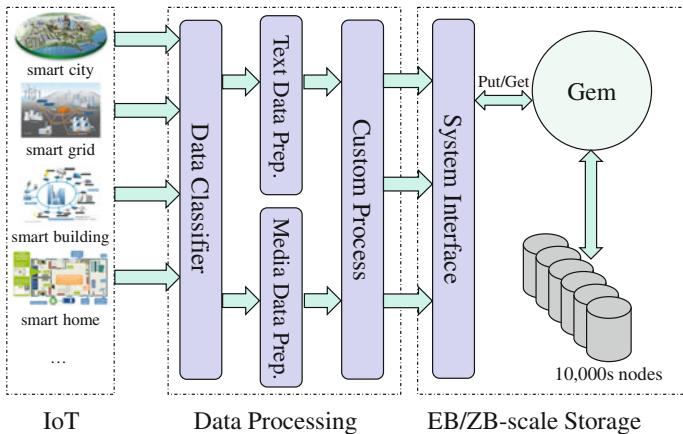
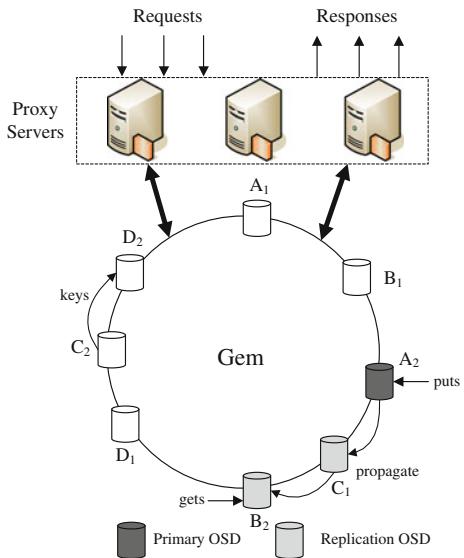


Fig. 3 EB-scale storage in IoT

Fig. 4 Gem system architecture



a serious problem. We have to consider a much more significant problem arising from network bandwidth. In general, to maintain connectivity of the network, every node frequently pings its neighbors to make sure they are still alive, and replaces them with new neighbors if they are not alive any more. There is a multiplicative increase in network traffic because of running multiple virtual OSDs in each OSD, but it is located in a data center with enough bandwidth. Meanwhile, an efficient linearizable consistency mechanism proposed in our previous research [7] keeps excellent replica consistency among OSDs.

3 Detailed Design

3.1 Type and Location Sensitive Hashing

Pathnames are directly used with fixed-size keys, where every lookup message should contain a key as large as the longest path. To limit message overhead without modifying routing mechanisms, we use a more compact key encoding with three continuous fields: type, location and pathname, as shown in Fig. 5. To place the files with the same type together, the type field is encoded first to facilitate retrieval. The first two fields are encoded with 4 bytes respectively, each directory of the third field is encoded with 2 bytes, and the last 4 bytes are allocated for a file name, and they can represent 2^{32} files per directory in theory. Eventually, the 64-byte key enables up to many quintillion files in count with many exabytes and even zettabytes of storage, and it excellently satisfies the requirement of the IoT. The key encoding mechanism provides a good trade-off between key size and file count, and it enables naming of new files and directories. In addition, a file may be moved to a different directory, and its key can be quickly changed to reflect the new path using the encoding mechanism. Furthermore, related objects are organized into a group using it to preserve in-order traversal of file system, e.g., files in the same directory are related.

In order to keep object locality, object keys are no longer distributed uniformly in the key space. Storage load would not be balanced as shown in Fig. 6. OSDs are

Type	Location	2	...	2	reserved	File Id
4 bytes	4 bytes	20 path levels, 40 bytes		12 bytes		4 bytes

Fig. 5 Type and location sensitive hashing

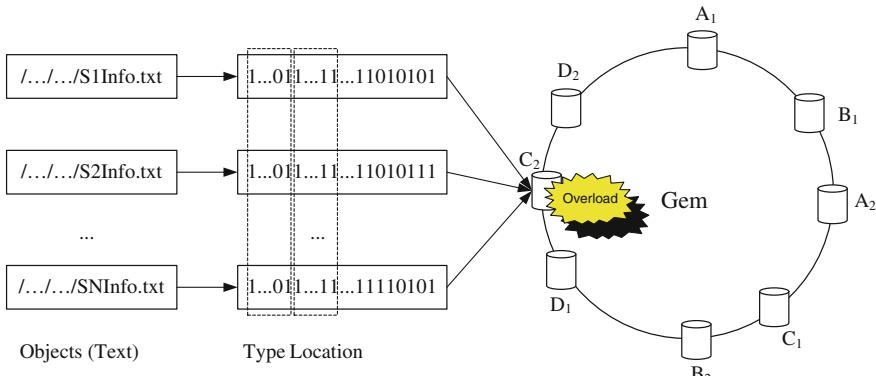


Fig. 6 Load imbalance caused by type and location sensitive hashing

responsible for roughly equal ranges of the key space in Gem. However, load balance is necessary to limit both the maximum storage that each OSD has to provision and the worst case data regeneration cost caused upon failures.

3.2 Dynamic Load Balancing

Each object server periodically contacts its neighbors in the system. Active drive D_i is load balancing if its load satisfies $\frac{1}{t} \leq \frac{L_i}{L} \leq t (t \leq 2)$. The Gem system is said to be load-balanced if the largest load is less than t^2 times the smallest load. Given a set of m OSDs $S = \{s_i, i = 1, \dots, m\}$ and a set of n virtual OSDs $V = \{v_j, j = 1, \dots, n\}$, each virtual drive v_j has a weight w_j that means how many files in a range are maintained by v_j , and each OSD s_i has a remaining capacity (weight) W_i that means the difference between the average storage load (capacity) W and the existing weight in the OSD s_i . The problem can be formulated as a 0–1 Multiple Knapsack Problem [8] (MKP), i.e., it is to determine how to reassign n virtual drives to m OSDs in a way that minimizes the wasted space in the OSDs as follows:

$$\text{minimize } z = \sum_{i=1}^m s_i \quad (1a)$$

s.t.

$$\sum_{i=1}^m x_{ij} = 1, \quad j \in N = \{1, \dots, n\} \quad (1b)$$

$$\sum_{j=1}^n w_j x_{ij} + s_i = W_i y_i, \quad i \in M = \{1, \dots, m\} \quad (1c)$$

$$w_j x_{ij} = \sum_{k=1}^l o_{jk}, \quad i \in M, \quad j \in N \quad (1d)$$

$$x_{ij} \in \{0, 1\}, \quad y_i \in \{0, 1\}, \quad i \in M, \quad j \in N \quad (1e)$$

where

$$x_{ij} = \begin{cases} 1 & \text{if virtual node } j \text{ is reassigned to OSD } i \\ 0 & \text{otherwise} \end{cases}$$

$$y_i = \begin{cases} 1 & \text{if OSD } i \text{ is used} \\ 0 & \text{otherwise} \end{cases}$$

s_i = space left in OSD i

o_{jk} = the storage size of the k th object in virtual OSD j

Constraint (1b) makes sure that each virtual OSD is only assigned to an OSD. Constraint (1c) ensures that the total number of files assigned to each OSD is less than the capacity of OSD. Constraint (1d) means that there are l objects with different sizes, where o_{jk} means the storage size of the k th object in virtual OSD j . Note that *Gem* is different from our previous work [9]: *Gem* stores objects with different sizes, while *DROP* stores objects (i.e., metadata) with fixed sizes. Constraint (1e) states it is a 0–1 knapsack problem. In addition, an adaptive load balancing proposed in our previous work [10] keeps excellent request load balancing among OSDs.

4 Conclusion

In this paper, we explore the use of object-based storage to improve the interaction between data analytics applications and storage systems in the Internet of Things. We present a large-scale object-based storage platform, called *Gem*, for data analytics in the IoT. Current data analytics applications roughly exploit the characteristics of storage systems, while current storage systems have no semantic knowledge of the requirements of data analytics in the IoT, making it difficult for well-informed optimization decisions. In *Gem*, we employ object-based storage interfaces to allow data analytics applications to communicate its storage requirements to the object-based storage system. By complying with current industry-standard OSD specification, *Gem* addresses data at a fine granularity, and it allows data analytics applications to access individual objects and their attributes.

Acknowledgments This work is supported by A*STAR under Grant No. DSI/14-300009.

References

1. Atzori, L., Iera, A., Morabito, G.: The Internet of things: a survey. *Comput. Netw.* **54**, 2787–2805 (2010)
2. Mesnier, M., Ganger, G.R., Riedel, E.: Object-based storage. *IEEE Commun. Mag.* **41**(8), 84–90 (2003)
3. Xu Q., Shen H.T., Chen Z., Cui B., Zhou X., Dai Y.: Hybrid retrieval mechanisms in vehicle-based P2P networks. In: Proceedings of the International Conference on Computational Science (ICCS'09). Lecture Notes in Computer Science, vol. 5544, pp. 303–314. Springer, Berlin (2009)
4. Shvachko, K., Huang, H., Radia, S., et al.: The hadoop distributed filesystem. In: MSST 2010 (2010)
5. Stoica, I., Morris, R., Karger, D.R., Kaashoek, M.F., Balakrishnan, H.: Chord: A scalable peer-to-peer lookup service for internet applications. In: SIGCOMM, pp. 149–160 (2001)
6. Xu, Q., Shen, H.T., Chen, Z., Cui, B., Zhou, X., Dai, Y.: Hybrid information retrieval policies based on cooperative cache in mobile P2P networks. *Front. Comput. Sci. China* **3**(3), 381–395 (2009)

7. Xu, Q., Arumugam, R.V., Yong, K.L., Mahadevan, S.: Efficient and scalable metadata management in EB-scale file systems. *IEEE Trans. Parallel Distrib. Syst.* **25**(11), 2840–2850 (2014)
8. Chekuri, C., Khanna, S.: A polynomial time approximation scheme for the multiple knapsack problem. *SIAM J. Comput.* **35**(3), 713–728 (2005)
9. Xu, Q., Arumugam, R.V., Yong, K.L., Mahadevan, S.: DROP: facilitating distributed metadata management in EB-scale storage systems. In: *MSST*, pp. 1–10 (2013)
10. Xu, Q., Arumugam R.V., Yong K.L., Wen, Y., Ong, Y.S.: C²: adaptive load balancing for metadata server cluster in cloud-scale storage systems. In: *IES*, pp. 195–209 (2015)

Concurrent Regeneration Code with Local Reconstruction in Distributed Storage Systems

Quanqing Xu, Weiya Xi, Khai Leong Yong and Chao Jin

Abstract Reed-Solomon (RS) codes are a standard erasure code choice and their repair cost is so high that it is a penalty for storage efficiency and high reliability. In this paper, we propose a novel class of Concurrent Regeneration codes with Local reconstruction (CRL), that enjoy three advantages: (1) to minimize the network bandwidth for node repair, (2) to minimize the number of accessed nodes, and (3) faster reconstruction in distributed storage systems. We show how they overcome the limitation of RS codes, and we demonstrate that they are optimal on a trade-off between minimum distance and locality. By conducting performance evaluation in both memory and JBOD environments, experimental results demonstrate the performance of the CRL codes.

Keywords Regeneration codes · Local construction codes · Storage system

1 Introduction

Large-scale distributed storage systems [1] are used to store all kinds of data, e.g., multimedia data including audio and video, and they typically use replication to provide reliability. Recently, erasure codes have been used to reduce the storage overhead of three-replicated storage systems, e.g., Microsoft's Windows Azure Storage [2], Facebook's HDFS-Xorbas [3] and Hitchhiker [4]. In a distributed

Q. Xu (✉) · W. Xi · K.L. Yong · C. Jin
Data Storage Institute, A*STAR, Singapore, Singapore
e-mail: Xu_Quanqing@dsi.a-star.edu.sg

W. Xi
e-mail: Xi_Weiya@dsi.a-star.edu.sg

K.L. Yong
e-mail: YONG_Khai_Leong@dsi.a-star.edu.sg

C. Jin
e-mail: Jin_Chao@dsi.a-star.edu.sg

storage system, the simplest form of data redundancy is replication, which replicates data into n replicas and each node stores a replica, incurring n times storage overhead. In addition, how to keep consistence of replicas is also a difficult issue [5]. Traditional erasure codes are suboptimal for distributed storage systems due to the so-called repair problem, i.e., when a single node fails, typically one chunk is lost from each stripe that is stored in that node. The RS (m, k) codes are usually repaired with the simple method that requires transferring m chunks and recreating the original m data chunks even if a single chunk is lost, hence causing a m times overhead in repair bandwidth and disk I/O. Local reconstruction codes [2] are a family of erasure codes that are used to reduce the repair overhead. In addition, it is possible to repair erasures with much less network bandwidth than this naive method [6].

When repairing a failed node, regenerating codes seek to minimize the amount of network bandwidth, while local construction codes attempt to minimize the number of nodes accessed. Apart from making sure reliability, the principal goals in distributed storage systems are node repair and data collection. We have to design an architecture that stores data across n nodes in such a way that a data collector can recover the data by contacting a small number k of nodes. Node repair is completed by downloading a uniform amount of data from each node in a subset of d nodes, where W is a total repair bandwidth. It is interesting to minimize both the repair bandwidth W and the repair degree, i.e., the number d of nodes accessed during repair. Meanwhile, it is desirable to have multiple choices for both node repair and data collection in terms of the set of k or d nodes that one links to. In our another work [7], we propose a class of codes named Regenerating-Local Reconstruction Codes (R-LRC), similar to the CRL codes, but the latter can tolerate more failures and be faster than the R-LRC codes.

2 Background

Two alternative coding approaches were advocated to enable more efficient node repair, i.e., local reconstruction codes [2] and regenerating codes [6].

2.1 Local Construction Codes

A family of Local Reconstruction Codes (LRC) encodes a file into n data chunks in such a way that one can recover any chunk by accessing only r chunks even after some of chunks are erased. LRC (k, l, r) is a class of maximally recoverable codes, and it tolerates up to $r + 1$ arbitrary failures. LRC provides low storage overhead. Among all the codes that can decode single failure from k/l chunks and tolerate $r + 1$ failures, it requires the minimum number of parity chunks. Therefore, the local reconstruction codes are in fact very similar to the CRL codes.

2.2 Regenerating Codes

One important idea behind regenerating codes is the sub-chunk mechanism, and each chunk is composed of a few sub-chunks. When a node storing a chunk fails, other nodes send in some of their sub-chunks for recovery. Efficiency of the recovery procedure is measured in terms of the overall bandwidth consumption. In many cases, regenerating codes can achieve a rather significant reduction in bandwidth consumption. In practice, a considerable overhead is caused because of accessing extra storage nodes. In particular, coding solutions that do not rely on the mechanism of sub-chunk and thus access less nodes are sometimes more attractive.

3 CRL Detailed Design

3.1 Definition

We formally define **Concurrent Regeneration Code with Local Reconstruction** (CRL). A CRL (k, g, l) calculates g global parities from all the data chunks, and divides k data chunks into l groups, where there are $(k + g)/l$ data chunks in each group. It calculates a local parity chunk within each group. Let n be the total number of chunks, i.e., $n = k + g + l$. Therefore, the normalized storage overhead is $n/k = 1 + (g + l)/k$. There is an example: CRL $(8, 2, 2)$ with storage cost of $1 + 4/8 = 1.5\times$, as shown in Fig. 1. We define *repair cost* as the number of chunks

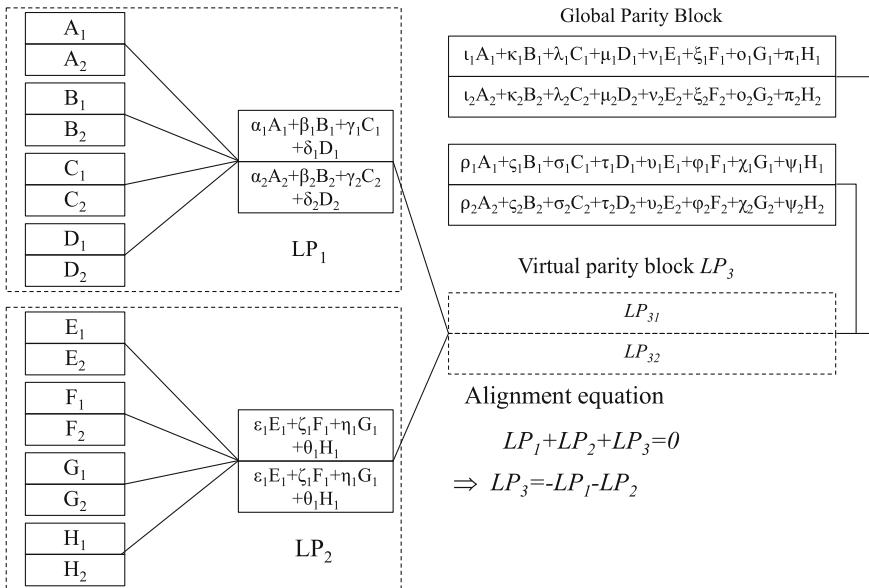


Fig. 1 CRL $(8, 2, 2)$

that are required to repair a failed *data chunk*. The goal of CRL is to reduce the repair cost. It achieves this by calculating some of the parities from a subset of the data chunks. CRL (8, 2, 2) generates 4 parity chunks, in which G_1 and G_2 are *global parities* that are calculated from *all* the data chunks. However, for the other two parities, it divides the data chunks into two equal size groups and calculates a *local parity chunk* for each group. The first local parity L_1 is calculated from the 4 data chunks in the first group, and the second local parity L_2 from the 4 data chunks in the second group.

It is easy to verify that the reconstruction of any single data chunk requires only 4 chunks, which are a half of the number required by the RS code. This CRL example adds one more parity chunk than the Reed-Solomon one, so it might appear that CRL reduces repair cost at the expense of higher storage overhead. In addition, CRL provides more options than Reed-Solomon code, in terms of trading off storage overhead and reconstruction cost. Note that the local parities satisfy an additional *alignment equation* $LP_1 + LP_2 + LP_3 = 0$, so $LP_3 = -LP_1 - LP_2$.

Therefore, we do not store the local parity LP_3 and instead consider it an *implied parity chunk*.

3.2 Encoding

The encoder of CRL initially divides a file into stripes of 8 chunks and calculates 2 global parity chunks. Depending on the file size, the last stripe may contain fewer than 8 blocks. We make incomplete stripes as “zero-padded” full stripes as far as the parity computation is concerned. In CRL, two extra parities are calculated for a total of 12 chunks per stripe, which includes 8 data chunks, 2 global parity chunks and 2 local parity chunks, as shown in Fig. 1. Similar to the computation of the parities in Reed Solomon codes, CRL computes all parity chunks in a distributed way. All the chunks come across a cluster according to a preset chunk placement mechanism.

3.3 Concurrent Repair for Multiple Failures

We start with a repair example for four failures to explain the concurrent repair for multiple failures. Figure 2 shows that there are four failures. Four nodes 1, 2, 3, and 4 download three sub-chunks from live nodes. For example, node 1 downloads the first sub-chunk LP_{11} of LP_1 , the first sub-chunk LP_{21} of LP_2 , and the first sub-chunk GP_{11} of GP_1 . Four variables E_1 , F_1 , G_1 and H_1 will vanish in GP_{11} after using LP_{21} based on interference alignment [8], and two variables C_1 and D_1 are known. In Node 1, there are two unknown variables A_1 and B_1 , and two equations, so A_1 and B_1 can be solved, as shown Fig. 3. Similarly, A_2 and B_2 can be solved in Node 2, E_1 and F_1 can be solved in Node 3, E_2 and F_2 can be solved in Node 4, as shown in

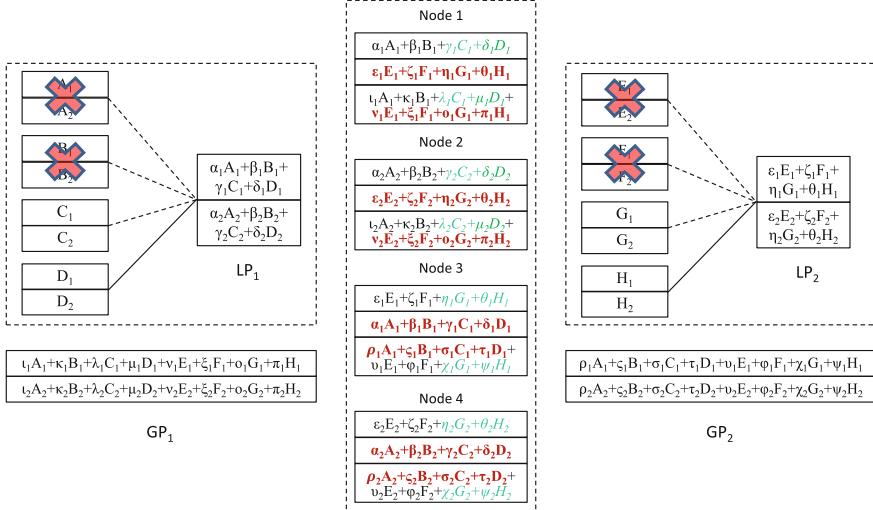


Fig. 2 Four failures

Fig. 3. *Node 1* and *Node 2* interchange A_2 and B_1 , so *Node 1* has A_1 and A_2 , and *Node 2* has B_1 and B_2 . In a similar way, *Node 3* and *Node 4* interchange E_2 and F_1 , so *Node 3* has E_1 and E_2 , and *Node 4* has F_1 and F_2 . Note that four nodes perform the repair process simultaneously, so it is faster than existing erasure codes.

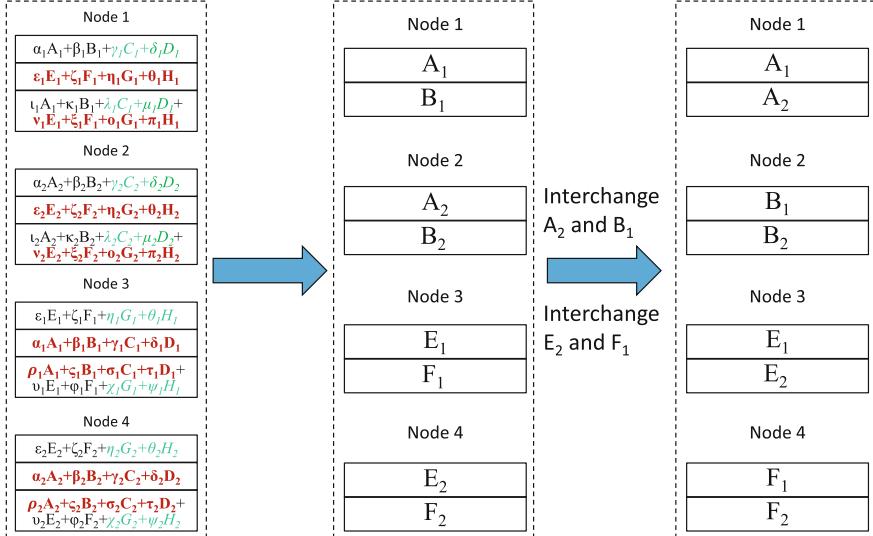


Fig. 3 Repair process

4 Performance Evaluation

In this section, we evaluate CRL(8, 2, 2) and compare its performance with other erasure codes: (1) RS_van that means the RS(8, 4) code based on the Vandermonde matrix, and (2) RS_cauchy that means the RS(8, 4) code based on the Cauchy matrix. We compute the throughput of encoding/decoding as follows: $sizeoff/f/t$, where f is a file that is encoded or decoded and t is running time cost. We measure the encoding/decoding throughput when repairing 1 erasure, 2 erasures and 3 erasures. All experiments are conducted on a Linux Server with Intel(R) Xeon(R) CPU E5-2690 v2 @ 3.00 GHz and 64 GB of RAM, running 64-bit Ubuntu 14.04 over Ext4, and a JBOD (Just a Bunch Of Disks) consisting of 10 1 TB WD Red 3.5" hard disks and 2 1 TB WD1000FYPS 3.5" hard disks. All the experiments are repeated five times, and average results are reported.

4.1 In Memory

Figure 4 presents the encoding and decoding throughputs in memory using different erasure codes. Obviously, CRL is much more efficient than the other two erasure codes when repair 1 erasure because of the local reconstruction property of CRL, as shown in Fig. 4a. Figure 4b illustrates that CRL has higher repair throughput than RS_van and RS_cauchy when repairing 2 erasures. In CRL, 2 erasures happen, and

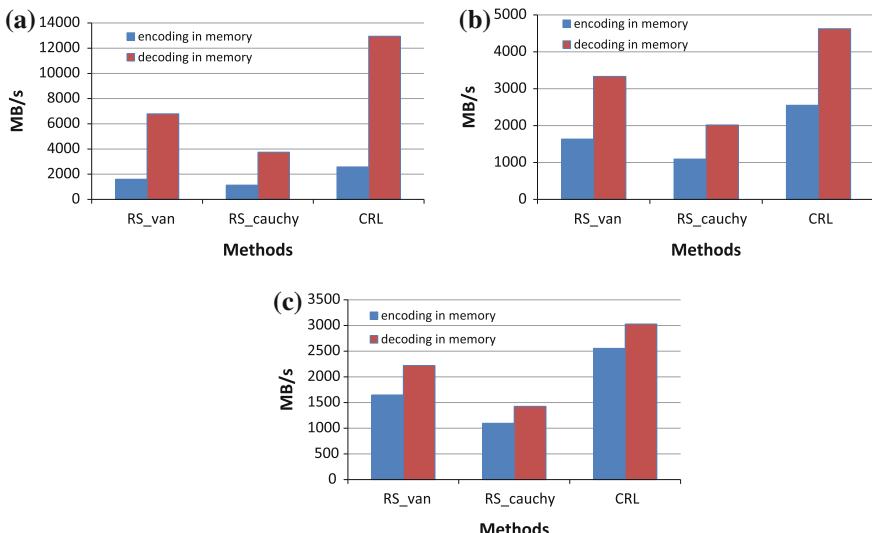


Fig. 4 Encoding/decoding throughputs in memory. **a** 1 erasure (*8). **b** 2 erasures (*4). **c** 3 erasures (*8/3)

they are concurrently repaired by accessing: (1) 8 chunks if both two local parities fail, (2) 4 or 5 chunks otherwise, while they are repaired by accessing 8 chunks in RS_van and RS_cauchy. Thus CRL is faster than RS_van and RS_cauchy in repair speed when repairing 2 failures. Similarly, CRL is faster than RS_van and RS_cauchy when repairing 3 failures, as shown in Fig. 4c.

4.2 In JBOD

In Fig. 5, we present the encoding and decoding throughputs in a JBOD using different erasure codes. Note that we use the file size to compute the decoding throughput instead of the size of erasure(s), so the decoding throughput is larger than the real one writing to the JBOD. There are $k + g + l = 12$ threads that are used to write chunks into 12 disk in the JBOD. As shown in Fig. 5a, CRL is more efficient than than RS_van and RS_cauchy when repairing 1 erasure in the JBOD because of its local reconstruction. Figure 5b demonstrates that CRL is faster than the other two codes because it can repair the failures concurrently due to its concurrent repair mechanism for multiple failures even in the worst case. CRL has similar repair throughput when repairing 2 failures and 3 failures because it uses similar numbers of chunks, as shown in both Fig. 5b, c.

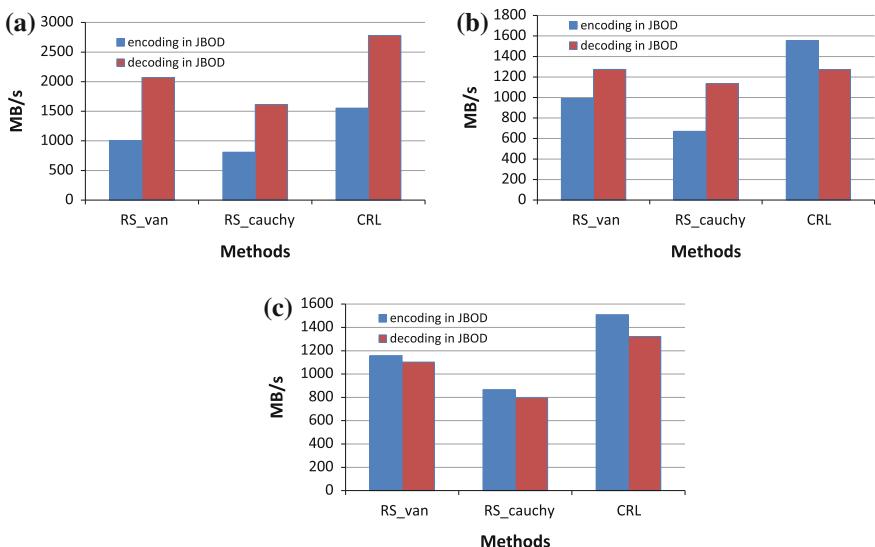


Fig. 5 Encoding/decoding throughputs in a JBOD. **a** 1 erasure (*8). **b** 2 erasures (*4). **c** 3 erasures (*8/3)

5 Conclusion

In this paper, we propose a novel class of erasure codes, named CRL. The CRL codes have three advantages. Firstly, they can minimize the network bandwidth for node repair. Secondly, they can minimize the number of accessed nodes. Last but not least, they have faster reconstruction than the existing erasure codes. We show how the CRL codes overcome the limitation of RS codes, and we demonstrate analytically that they are optimal on a trade-off between minimum distance and locality. By conducting performance evaluation in both memory and JBOD environments, experimental results demonstrate the CRL codes have better performance than the existing erasure codes.

Acknowledgments This work is supported by A*STAR under Grant No. R15GAP-0004.

References

1. Xu, Q., Arumugam, R.V., Yong, K.L., Mahadevan, S.: DROP: facilitating distributed metadata management in EB-scale storage systems. In: MSST, pp. 1–10 (2013)
2. Huang, C., Simitci, H., Xu, Y., Ogu, A., Calder, B., Gopalan, P., Li, J., Yekhanin, S.: Erasure coding in windows azure storage. In: 2012 USENIX Annual Technical Conference, Boston, MA, USA, June 13–15, 2012, pp. 15–26 (2012)
3. Sathiamoorthy, M., Asteris, M., Papailiopoulos, D.S., Dimakis, A.G., Vadali, R., Chen, S., Borthakur, D.: Xoring elephants: novel erasure codes for big data. PVLDB **6**(5), 325–336 (2013)
4. Rashmi, K.V., Shah, N.B., Gu, D., Kuang, H., Borthakur, D., Ramchandran, K.: A “hitchhiker’s” guide to fast and efficient data reconstruction in erasure-coded data centers. In: ACM SIGCOMM’14, pp. 331–342 (2014)
5. Xu, Q., Arumugam, R.V., Yong, K.L., Mahadevan, S.: Efficient and scalable metadata management in EB-scale file systems. IEEE Trans. Parallel Distrib. Syst. **25**(11), 2840–2850 (2014)
6. Dimakis, A., Godfrey, P., Wu, Y., Wainwright, M., Ramchandran, K.: Network coding for distributed storage systems. IEEE Trans. Inf. Theory **56**(9), 4539–4551 (2010)
7. Xu, Q., Ng, H. W., Xi, Q., Jin C.: Regenerating-Local Reconstruction Codes for Distributed Storage Systems. To appear in the 7th ICFCC (2015)
8. Wu, Y., Dimakis, A.G.: Reducing repair traffic for erasure coding-based storage via interference alignment. ISIT **2009**, 2276–2280 (2009)

A Technique for Streaming Multiple Video Parts in Parallel Based on Dash.js

Konthorn Sangkul, Sucha Smanchat and Jo Yew Tham

Abstract Watching video on the Internet has become an alternative to offline video playback on video player software. Videos on the Internet are streamed to video player so that they can be played without having to wait for the whole files to be downloaded. However, when playing a video, some users may skip ahead to different positions or scenes in the video to see whether the video is really of their interest. Most video players only stream videos from the beginning, thus jumping to a part of a video that has not been streamed causes a delay while the video starts to buffer. In this paper, we propose a technique to stream multiple parts of a video simultaneously by modifying the dash.js framework so that users can navigate to different points in a video with minimal delay. The prototype that realizes our concept proves the feasibility of our approach. With further improvement, our technique may contribute to the scene selection function for video streaming, similar to that of video player software.

Keywords Video streaming · Parallel streaming · Dash.js · Scene selection

1 Introduction

Video playback over the Internet has become substantial alternative to offline playback on media player software. Users can watch video from many video streaming providers such as YouTube and Netflix. With the advance in Internet

K. Sangkul (✉) · S. Smanchat

Faculty of Information Technology, King Mongkut's University of Technology North
Bangkok, 1518 Pracharat 1 Rd, Bangsue, Bangkok, Thailand
e-mail: konthorn.sangkul@gmail.com

S. Smanchat

e-mail: sucha.s@it.kmutnb.ac.th

J.Y. Tham

Institute for Infocomm Research (I2R), A*STAR, 1 Fusionopolis Way, Singapore, Singapore
e-mail: jytham@i2r.a-star.edu.sg

technology, watching video online has become easier and faster, however, it does not have full-fledge functionalities of media player software. Specifically, to reduce network traffic, video streaming usually starts from the beginning of a video.

When playing a video, some users may skip ahead to different positions or scenes in the video to see whether the video is really of their interest. With most of the current video streaming, jumping to a part of a video that has not been streamed causes a delay while the video starts to buffer. This makes it less convenient to the users. This paper proposes a technique to alleviate this problem by streaming multiple video parts simultaneously so that users can jump to different positions with minimal delay.

The structure of this paper is organized as follows. Section 2 describes the related technology in video streaming including video streaming techniques and the dash.js framework of which modification is explained in Sect. 3 in order to enable parallel streaming. Section 4 illustrates the evaluation of our prototype and Sect. 5 concludes the paper with our intended future work.

2 Related Work

This section explains the basic concept of video streaming along with existing techniques used in video streaming. The dash.js framework that is used for video playback is subsequently explained to provide the background of our dash.js extension.

2.1 *Video Streaming*

Video streaming is a technique that starts video playback as soon as the video frames arrive without having to wait for the whole file. The simplest way is to progressively download a video and play it on a video player [1]. Videos usually have metadata attached to provide information about the videos. Metadata can include video title, description, tags or keywords, authors and copyright [2, 3]. There are two types of video metadata: (1) auto-generated metadata from software or device such as file format and date (2) manually-created metadata, such as a transcript of dialogue, which can be used for custom action such as search optimization and specifying the URLs and video segments for parallel video streaming.

Streaming and progressive download are two basic types of video delivery. Traditional video streaming uses Real-Time Transport Processing (RTP) [4]—a stateless connection to deliver video content. RTP can achieve short delay and fast delivery as it does not have fault tolerance, so any missing video frame is ignored. On the contrary, in progressive download, videos are delivered via TCP thus any missing video segment is re-transmitted [5]. With HTTP range request, it is possible to download only a part of video and buffer them on video client. Streaming usually

allows some form of feedback from client to provide the ability to change and adapt video stream according to user commands, network status and device capability.

2.2 *Parallel Video Streaming Techniques*

The word parallel video streaming describes the delivery of two or more video stream to a target client at the same time. Lederer [6] used content centric network, a transfer link that needs no client-server connection, to download multiple streams at the same time. Mobile devices can use available connections such as mobile network and Wi-Fi at the same time to achieve parallel streaming over multiple network interfaces [6].

Tu and Sreenan [7] propose an adaptive split transmission algorithm that splits a video into multiple sub-flows. These flows are delivered over multiple unused wireless network channels in parallel without any change in underlying network. Shen and Li [8] uses DHT-based (Distributed Hash Table) and other algorithms to manage peer nodes for peer-to-peer (P2P) video streaming network. A decentralized network is created to serve video segments from appropriate providers. The technique enhances the P2P streaming with improved scalability, availability and latency of the video [8].

Chaurasia and Jagannatham [9] use TCP protocol for video transmission in their “Parallel TCP For Wireless Scalable Video Transmission” technique. By using multiple TCP connections, it efficiently utilizes the bandwidth of wireless channel and improves the overall video streaming quality. A video is encoded into hierarchy of base and enhancement layers, which are transferred through different TCP connections by giving the highest priority to the base layer to be transferred first.

So far, we have not encountered any technique that streams many parts of a video in parallel. Unlike existing work, which splits a video by physical connections [6] and horizontal layers [7, 9], this research focuses on parallel streaming that splits a video into time-based (vertical) fragments over HTTP requests. The proposed technique in this research uses the benefit of Media Source Extension (MSE) [10] and HTML5 media elements [11] to achieve parallel video streaming.

2.3 *Dash.js*

The “Dynamic Adaptive Streaming over HTTP”, or DASH [12] is an adaptive bit rate streaming technique to provide the standard for media delivery over the Internet. Unlike traditional video streaming, with DASH and MSE, most of the processes occur on client side [13]. DASH works by splitting a video file into smaller segments. In HTML video streaming, javascript is used to create a video object that can interact with commands such as play, pause, change video source and insertion of video segments enhancing flexibility of media control through

MPD manifest file [14]. This paper focuses on DASH video on HTML by modifying the dash.js [15], a javascript library based on MSE, to enable parallel video streaming.

3 A Parallel Video Streaming Technique

The proposed technique to achieve streaming multiple video parts involves modifying the MPD file and the modification of fragment scheduling within dash.js framework.

3.1 *Modifying MPD File*

The “Media Presentation Description”, or MPD file, is an xml manifest file based on ISO Base Media File Format (BMFF). MPD describes the segment structure and resource identifiers such as URL, duration and bit rate. As an xml file, MPD allows for custom tags and parameters. This research uses a custom “weight” parameter [16] to add priority to video segment tags to allow for more control over the flow of video segments. With a sequence number and a weight attached to each video segment, we can sort and rearrange the order of the fragments being loaded into buffer. For example, in Fig. 1, two video parts begin from the video segments numbered 0 and 3 with the weight of 1. The first part has 3 segments while the second part has 2 segments.

In this work, a part of a video is loosely defined because it may start from a segment with a weight value higher than 1 (e.g. we want the part near the end to start streaming later). The specification of part, which may require consideration of performance and user specification, is out of the scope of this paper. A simple approach is to divide a video into parts of an equal size. To start streaming from many parts simultaneously, fragment scheduling in dash.js is modified.

```

<SegmentList timescale="11988" duration="35999">
  <Initialization sourceURL="seginit.mp4"/>
  <SegmentURL sequenceNumber = "0" weight="1" media="seg1.m4s"/>
  <SegmentURL sequenceNumber = "1" weight="2" media="seg2.m4s"/>
  <SegmentURL sequenceNumber = "2" weight="3" media="seg3.m4s"/>
  <SegmentURL sequenceNumber = "3" weight="1" media="seg4.m4s"/>
  <SegmentURL sequenceNumber = "4" weight="2" media="seg5.m4s"/>
<SegmentList/>

```

Fig. 1 The modification of MPD file

3.2 *Modifying Fragment Scheduling*

Based on the modified MPD, dash.js handler begins by reading the information entries of all segments. The entries are then sorted into the “Pending Request” list [15] in ascending order based on the assigned weight. The entries with the same weight are further sorted based on sequence numbers in ascending order. This sorting makes the segments with the same minimum weight (i.e. 1) at different parts of the video to appear first (unlike existing players whose buffers grow only from left to right).

Fragment controller in dash.js then progressively downloads the video segments using HTTP requests beginning from the front of the Pending Request list. Once a segment completes downloading, the Buffer Controller checks if the video buffer is busy and if the “Pending Media” list [15], which contains the segments waiting to be appended, is empty. If the buffer is being appended by another segment or the Pending Media list is not empty, the new segment is added to the Pending Media list to be subsequently appended to the video buffer; otherwise the new segment is immediately appended to the buffer. This process can be elaborated in Fig. 2.

In Fig. 2, the Pending Media list may contain the list of segments that are not in order of the sequence number. This happens because the segments may complete their download in an order that is different from that of the Pending Request list.

4 Evaluation Result

The prototype is implemented to simultaneously stream multiple parts of a video file, which are stored locally, to a video player. The video used for the evaluation is approximately 3 min in length. The MPD file is specified to divide the video into 5 parts of an equal size. The length of each part is thus approximately 36 s.

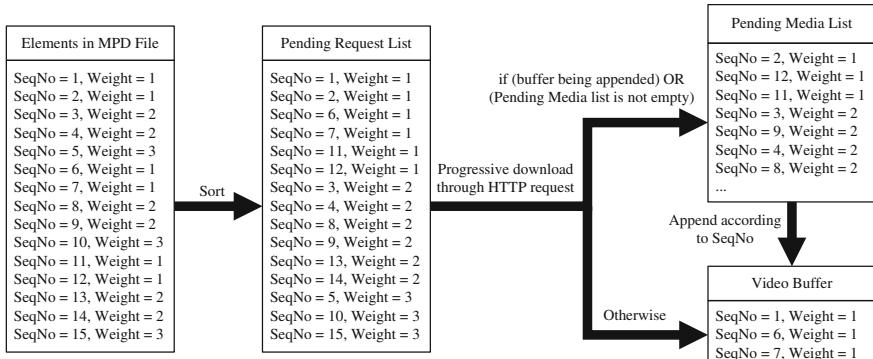


Fig. 2 The process of streaming multiple video parts



Fig. 3 The video progress at the beginning of playback



Fig. 4 The video progress after the streaming starts



Fig. 5 The video progress after more segments are buffered

Figure 3 shows the video progress before starting the parallel streaming. Once the streaming starts, every part is streamed simultaneously as can be seen by the video buffer growing at multiple positions in the video slider in Fig. 4. Users can click to skip the playback to any location that has been buffering without having to wait. Figure 5 depicts the buffer after more segments are appended to it.

As mentioned earlier, the segments may finish downloading in an order that is different from the sorted Pending Request list. Thus, the buffer may consequently grow in a different order. From the result, the prototype verifies the feasibility of our technique based on the modification of dash.js framework. Nevertheless, some technical considerations still need to be resolved to enable efficient parallel streaming such as using HTTP range request to reduce the number of video segments. The bandwidth requirement also needs to be investigated to determine an appropriate number of parts and the offset delay before starting later parts of a video to enable smooth streaming.

5 Conclusion and Future Work

We propose a technique for simultaneously streaming multiple video parts so that users can jump to different parts of a video without experiencing delay. Our approach modifies the dash.js framework to enable such streaming. Our prototype that realizes our concept proves the feasibility of the approach and the full version is being developed to be deployed and tested on the cloud environment.

Further improvement on segment scheduling will be conducted together with bandwidth testing. As a final outcome, each video part may correspond to a scene in the video so that video streaming can be enhanced with the scene selection function similar to that of video player software.

References

1. Ozer, J.: Streaming vs. progressive download vs. adaptive streaming. <http://www.onlinevideo.net/2011/05/streaming-vs-progressive-download-vs-adaptive-streaming/>
2. YouTube: Video metadata. <https://www.youtube.com/yt/playbook/metadata.html>. 17 Nov 2014
3. Video University: Metadata for video. <http://www.videouniversity.com/articles/metadata-for-video>. 17 Nov 2014
4. Schulzrinne, H. et al.: RTP: a transport protocol for real-time applications (RFC3550). <http://tools.ietf.org/html/rfc3550>. 17 Nov 2014
5. Kashyap, A. et al., *Efficient HD Video Streaming Over the Internet* in *Proceedings of the IEEE SoutheastCon*, 2010
6. Lederer, S., et al : Adaptive streaming over content centric networks in mobile networks using multiple links. In: Proceedings of the 2013 IEEE International Conference on Communications Workshops (ICC)
7. Tu, W., Sreenan, C.: Adaptive split transmission for video streams in wireless mesh networks. In: Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC 2008)
8. Shen, H., et al.: A DHT-Aided chunk-driven overlay for scalable and efficient peer-to-peer live streaming. *IEEE Trans. Parallel Distrib. Syst.* **24**(11) (2013)
9. Chaurasia, A., Jagannatham, A.: dynamic parallel TCP for scalable video streaming over MIMO wireless networks. In: Proceedings of the 6th Joint IFIP Wireless and Mobile Networking Conference (WMNC) (2013)
10. Colwell, A. et al.: Media source extension. <https://dvcs.w3.org/hg/html-media/raw-file/tip/media-source/media-source.html>. 11 Nov 2014
11. W3C: HTML5 media elements. <http://dev.w3.org/html5/spec-preview/media-elements.html>. Editor's Draft 22 Aug 2012; 20 Jan 2015
12. ISO/IEC 23009-1:2012: Dynamic adaptive streaming over HTTP (DASH). http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=57623. 17 Nov 2014
13. Miller, K. et al.: Adaptation algorithm for adaptive streaming over HTTP. In: Proceedings of the 19th International Packet Video Workshop (PV) (2012)
14. Dash Industry Forum: DASH-264 javascript reference client. <http://dashif.org/reference/players/javascript/1.2.0/index.html>. 17 Nov 2014
15. Dash Industry Forum: Dash.js. <https://github.com/Dash-Industry-Forum/dash.js/wiki>. 17 Nov 2014
16. Evensen, K. et al.: A network-layer proxy for bandwidth aggregation and reduction of IP packet reordering. In: Proceedings of the IEEE 34th Conference on Local Computer Networks (LCN 2009), Zürich, Switzerland, 20–23 Oct 2009

Prioritized Medical Image Forwarding Over DTN in a Volcano Disaster

Muhammad Ashar, Morihiko Tamai, Yutaka Arakawa
and Keiichi Yasumoto

Abstract In this paper, we propose a method for priority-based medical image forwarding over DTN (delay-tolerant networks), in which an important part (injured part) in a patient's picture is detected and assigned a high priority, and forwarded faster in DTN. In our method, we suppose a disaster scenario where medical doctors in a disaster area take a picture of the injury and send it to the main hospital for getting information of appropriate treatment. DTN is a promising technology used for transmitting images from the disaster area to the hospital in an environment where no high speed cellular networks are available. However, transmitting a high-resolution medical image over DTN is a challenging task. Our proposed method aims to solve this problem by introducing image segmentation and priority forwarding techniques into a conventional DTN. Focusing on eye injury in a volcano disaster, we developed an algorithm to divide each patient's picture into pieces and assign a priority to the pieces containing the injury part in the picture with color marking. We also employ a scheduling methods for priority based message forwarding, in which the intermediate node in DTN forward high priority pieces before other pieces. As a result, an important part of a picture, which is mandatory for diagnosis, will arrive at the main hospital faster than other parts. Through a computer simulation, we confirmed that the proposed method could deliver diagnosable images faster than a conventional method.

Keywords Delay-tolerant networks · Priority forwarding · Medical image delivery · Image segmentation

M. Ashar (✉) · M. Tamai · Y. Arakawa · K. Yasumoto
Information Science, Nara Institute of Science and Technology, Nara, Japan
e-mail: Muhammad_ashar.ls6@is.naist.jp

M. Tamai
e-mail: morihi-t@is.naist.jp

Y. Arakawa
e-mail: ara@is.naist.jp

K. Yasumoto
e-mail: yasumoto@is.naist.jp

1 Introduction

There are many active volcanic mountains in the world and they suddenly erupt and cause a lot of economical and human damages. For example, Indonesia's Mount Merapi erupts periodically. A lot of people are potentially exposed to an acute eye injury from volcanic products such as ash fall, which contains varying proportion of free crystalline silica in short duration (days to week). Because there are no ophthalmologists in a disaster area in general, a local doctor needs to ask how to treat these eye injuries to an ophthalmologist in a city hospital by showing injury images and so on. However, it is difficult to transmit medical images through a cellular network, because high-speed data connection service is not provided. Consequently, the victim does not receive medical treatment quickly.

In case of a mountain eruption, the infrastructure of 3G/4G mobile phone is often destroyed. Therefore, a mobile ad hoc network or DTN (delay-/disruption-tolerant network) that does not require the fixed network infrastructure is considered as a promising technology. In recent years, DTN has been widely studied and some Android applications have already been released. Therefore, in this paper, we suppose to employ DTN for communication means to transmit medical images.

We assume the following environments taking into account the actual disaster environment by volcano in Indonesia. Because a disaster area is a rural area, there are no special doctors (i.e., ophthalmologists) who can treat appropriately the eye injury. Therefore, local doctors try to ask the main hospital in the large city how to treat each eye injury. In this scenario, we assume that ambulance cars become a carrier of data between a disaster site and the hospital. The image taken by a doctor can be transmitted to the ambulance car over DTN consisting of rescuers and so on. After that, all the images stored in the ambulance car are brought to the main hospital. The responses (e.g., appropriate treatment instruction) from the hospital are informed to the local doctor through ordinary phone or facsimile.

A Message Priority Routing for DTN [1] assumes that an earthquake has occurred in a city and roads have been damaged in the disaster. By the stochastic-based routing protocol, it was shown that the proposed message priority routing protocol was able to deliver more messages with lower overhead ratio and latency than existing DTN protocols. Priority Scheduling for Participatory DTNs [2] implements a priority scheduling mechanism on the top of an existing DTN protocol over a variety of real mobility traces and the value (or priorities) of messages attached by end users is used to decide the order in which the messages will be forwarded. With a prioritization scheme, the scheduling mechanism guarantees that enough resources are available to forward more important messages. Both metrics such as *satisfaction gain* (the difference between the average value of delivered messages with and without priority) and *delivery gain* (the difference between the average number of delivered messages with and without priority) has been computed in wide network measurements. InfoBox [3] provides a service which delivers word-of-mouth messages (including photos) posted on a spot in a sightseeing area to mobile users who are going to visit the spot, taking into account

priority and deadline which are determined based on the content of each message and the time before users requiring the message arrive at the spot, respectively. InfoBox will discard messages that cannot meet the delivery deadline for efficient bandwidth utilization. Medprop [4] provides a DTN-based system that provides medical services to rural area taking into account priority among data. This system is useful not only in the medical scenario but also in many other applications where the delivery rate of critical data is a major concern. One may think that Medprop can be used to send medical images on a volcano disaster in the rural area. In an image content delivery system, however, a large delay may be introduced even when we choose appropriate routing protocol for a particular scenario.

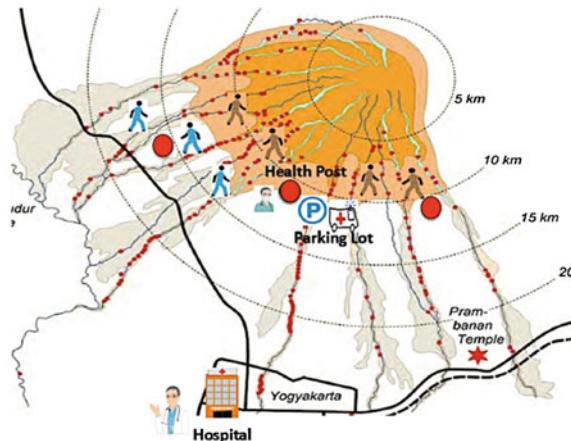
One of problems in image delivery is how to recognize critical images as high priority data especially for eye injury image. To solve this problem, we introduce two methods: Color Marking based image segmentation and priority forwarding. By dividing a large medical image into small pieces, we try to improve the throughput and reliability of DTN. After the system divides the image, the important pieces of the image are identified and a high priority is assigned to them. In addition, each intermediate node in DTN forwards the pieces having higher priority prior to other pieces. As a result, important pieces can be delivered to the hospital as fast as possible. We show the effectiveness of the proposed method through a computer simulation to evaluate the time until all of the high priority pieces in each image are received at the hospital and compare our method with a conventional method without prioritization of pieces for two representative DTN routing protocols (Epidemic and Spray-and-Wait).

2 Target Medical Image Delivery Service

We model a DTN as a set of mobile nodes. A contact occurs when two nodes meet within their wireless transmission range, creating an opportunity to transmit data between nodes. In a realistic scenario of Merapi Mount disaster, the rescue services are provided by an emergency medical response team that consists of rescuers and ambulance drivers. They are treated as mobile nodes in our DTN model.

As shown in Fig. 1, we assume that multiple health posts (where medical doctor treats patients) and one or more ambulance parking lots are placed in the disaster area. Rescuers move between the ambulance car parking lot and one of the health posts selected at random. Thus, we model the mobility of rescuers as a variant of Random Way Point (RWP) mobility model. Similarly to the ordinary RWP, nodes corresponding to rescuers wait for a pause time and then moves to a randomly chosen location at a speed chosen from the range (V_{\min} , V_{\max}) [5]. They walk from a health post to a parking lot in the disaster zone. Meanwhile, ambulance drivers move between a parking lot and a hospital. Based on the geographical features on a specific Merapi disaster area, we assume that the distance between the disaster area (ambulance parking lots) and a hospital is around 20–30 km. We also assume that the parking lots are located about 5 km from the farthest health post.

Fig. 1 Example of medical response to deliver victims to hospital in Merapi disaster



In this situation, we consider how to deliver images quickly from a source location (i.e., health post) to a destination location (i.e., hospital) by forwarding messages over DTN among the nodes.

We consider an evacuation process within 24 h after eruption to find 50–70 victims. In this setting, we assume that there are 5–10 % victims that will potentially cause acute eye injury. After a doctor has found a victim with symptom eye injury, the doctor immediately takes a picture with his/her smartphone's camera. To reduce the total amount of data size to be transmitted, we employ an approach that a doctor manually makes a marking on injury part based using a color (corresponding to priority level or resolution) in the patient's picture so that the injury part in the picture is picked for prior transmission over DTN.

We assume that Android smartphones, which are capable of peer-to-peer communication through WiFi Direct, are used in the disaster area for multi-hop communication between nodes. In this context, we assume that each smartphone acts as a node of DTN, which ensures network connectivity without relying on communication infrastructure (e.g., 3G cell towers) by following a store-carry-and-forward manner for message delivery. We focus on a medical image delivery service over DTN, where medical images are generated by doctors in disaster area, packed to bundle messages, and carried and forwarded by the rescuers and ambulance drivers to deliver the images to the city hospital for diagnosing the condition of eye injury.

We propose a weighted priority forwarding routing scheme with the image prioritization method over the existing DTN routing protocols such as Epidemic and Spray-and-Wait. The goal of the proposed scheme is to minimize the delivery delay and to find the good ratio of delivering medical images at the hospital per unit of time.

3 Priority Assignment to Images

Below, we describe the proposed scheme for medical image delivery. In the scheme, an image captured by a doctor's smartphone is analyzed by an existing algorithm [6, 7] to detect the regions of eyes (i.e., the regions of left eye and right eyes), and these regions are extracted from the image. This process is necessary for reducing the data size of the whole image. Then, the image is partitioned into sub-blocks [5], by splitting the whole image into pieces (or chunks) (see Fig. 2, left). In our method, a region is partitioned into 25 (5×5) equal-sized sub-blocks as shown in Fig. 2 (middle). Then, priorities are assigned to the subset of the whole pieces in the image, which a doctor thinks important (i.e., injured) by manually specifying the pieces with an interface provided by a smartphone application we develop. For example, high priority is assigned to pieces containing serious eye injury such as eye blood, blur, or acute irritation (red eye). We suppose that about 13 pieces of the 25 are assigned a priorities by marking specific color. Here, red, green, and blue colors are marked for very-high, high, and medium-priority, respectively, and a gray color for low-priority.

Then, the remaining pieces without marking are removed from the storage. To diagnose medical images precisely, high-resolution image is needed. Thus, the system gives higher resolution to pieces with higher priority. For example in Fig. 2 (right), red color pieces are stored with the highest resolution (i.e., largest data size) than the pieces with other colors.

4 Message Selection for Priority Forwarding

In the proposed priority forwarding scheme, when a node meets other node, it transmits pieces in the decreasing order of their priority level. We will explain more about using an example in Fig. 3. In the example, we suppose that node which has

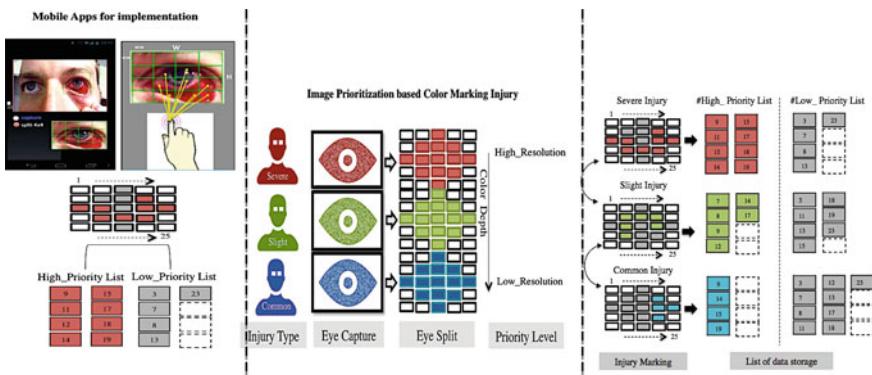


Fig. 2 Image partitioning and image prioritization method

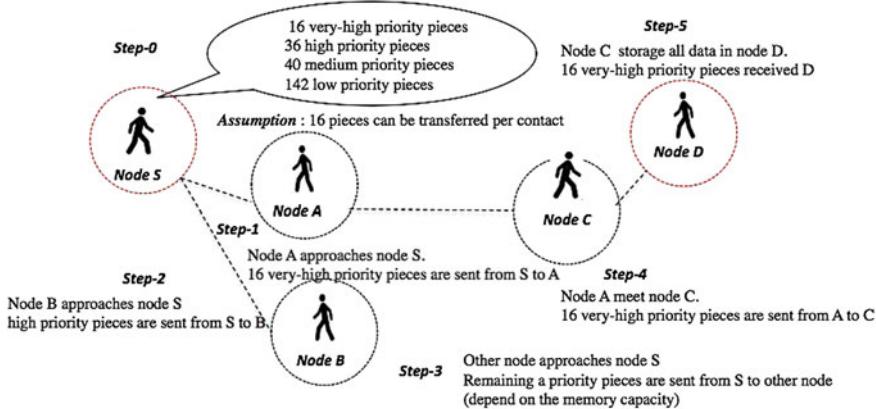


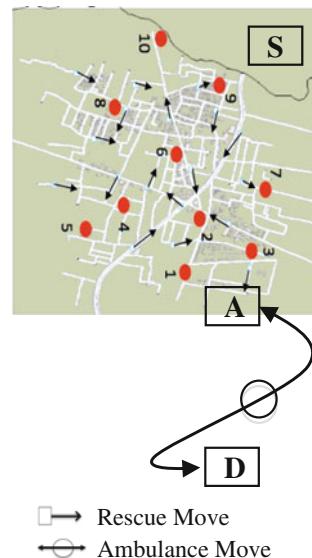
Fig. 3 Priority based routing

16, 36, 40, and 142 pieces with very-high, high, medium, and low priority, respectively wants to deliver the pieces to node D. Suppose that node S meets nodes A and B in this order and that 16 pieces can be transmitted during a contact. When node S meets node A, it transmits 16 pieces with very-high priority to node A. After that, when nodes S meets node B, it transmits 16 high priority pieces to node B. Similarly, node S transmits the remaining pieces whenever it meets other node. Here, note that when higher priority pieces are received at node S (e.g., by taking new pictures, receiving the pieces from other node, etc.), the highest priority pieces are always transmitted first. Then, when node A with 16 very-high priority pieces meets other node, say node C, the pieces are transmitted to C, similarly. Like this way, pieces are transmitted through multiple paths depending on their priority level to the destination node D.

5 Simulation Experiments

We set up a simulation environment with a realistic eruption disaster map and realistic node mobility in an emergency medical response and compared performance of our proposed medical image delivery method with a conventional non-priority image delivery method through simulations.

In order to create realistic node mobility to simulate emergency medical response, we used the Multi-Agent Module of Scenargie Simulator [8]. We configured a simulation field of 2×2 Km using OpenStreetMap corresponding to an actual geographical area near Mount Merapi, Indonesia as shown in Fig. 4. Other parameters used for simulation are shown in Table 1. In our simulation, mobile nodes in a disaster area are equipped with smartphones or similar devices with Wi-Fi Direct at effective transmission rate of 5 Mbps with 100 m effective radio

Fig. 4 Simulation field

transmission range. We assume a small-scale scenario where 21 mobile nodes (20 rescuers and 1 ambulance) move in the simulation field (Fig. 4). As shown in Fig. 4, there are 10 red circles representing PoIs (Points of Interest) which correspond to health posts. In each health post, a doctor treats patients. In Fig. 4, a hospital (D) with ophthalmologists is located 28 km away from the disaster area parking lot (A) through only a direct path A-D.

Mobility patterns of rescuers and ambulance are specified as follows. Each rescuer moves to one of PoIs (red circles in Fig. 4) to find victims, following the shortest path from the current location. After visiting a PoI, the rescuer moves to the

Table 1 Simulation parameters

Parameter	Value
Simulation time	3600 s
Network interface	Wireless link
Interface type	Simple broadcast
Transmission rate	5 Mbps
Transmission range	100 m
Mobility of rescuer	GIS based RWP
Buffer size	10 MB
Routing protocol	Epidemic and SaW
# of nodes :	32
# of rescuers, drivers, doctors	21,1,10
Speed of pedestrian nodes	0.5-1.5 m/s
Speed of Vehicle nodes	2.5-15 m/s
Pieces size (red, green, blue, gray)	35, 25, 10, 5 kB

parking lot (A) to carry the critical patient who needs the treatment at a city hospital. After that, the rescuer selects other PoI, and repeats the above behavior. Meanwhile, the ambulance shuttles between the hospital (D) and the parking lot (A) in the disaster area. Once the ambulance reaches a parking lot, it stops for 5 min (to carry a patient in ambulance), and back to the hospital again. At each health post, one doctor treats patients (there are 10 doctors in total in the disaster area). Each generated image is divided to 25 pieces and 13 of the pieces are marked with a color (red, green, blue, or gray) and delivered to the parking lot (A) through DTN by rescuers and then delivered to the hospital (D) by the ambulance. Here, in one hour simulation time, 10 images (130 pieces with priority) are generated where 80 pieces are high-priority (40_{red} , 24_{green} , 16_{blue}) and 50 pieces are low-priority (50_{gray}). The total of image size is 3005 kB. In simulations, we compare the performance of the proposed method to a conventional method without prioritization of pieces by using Epidemic and Spray-and-Wait routing protocols.

6 Performance Evaluation

We measured the message delivery rate and the CDF (cumulative distribution function) of delivery delay at three points of simulation time ($T_1 = 1200$ s, $T_2 = 2400$ s, and $T_3 = 3600$ s), which are shown in Figs. 5 and 6, respectively. Note that, in the experiment, we repeated the simulation 4 times with different random seeds, and the results are averaged. In the figures, high priority and low priority correspond to the performance of the proposed method and non-priority corresponds to the conventional method. In Fig. 5, there are three points where the delivery rate increases. These points correspond to the time when the ambulance arrives at the hospital. When comparing Epidemic and Spray-and-Wait, the latter shows a slightly better performance. The non-priority achieves about 70 % delivery rate at the end of simulation time. On the other hand, the high-priority (our method) achieves 90 % delivery rate, although the low-priority's delivery rate becomes lower than the non-priority. It is shown that the proposed method with message

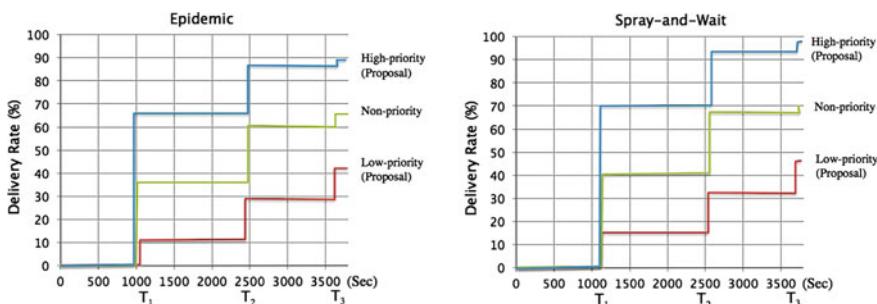


Fig. 5 Message delivery ratio

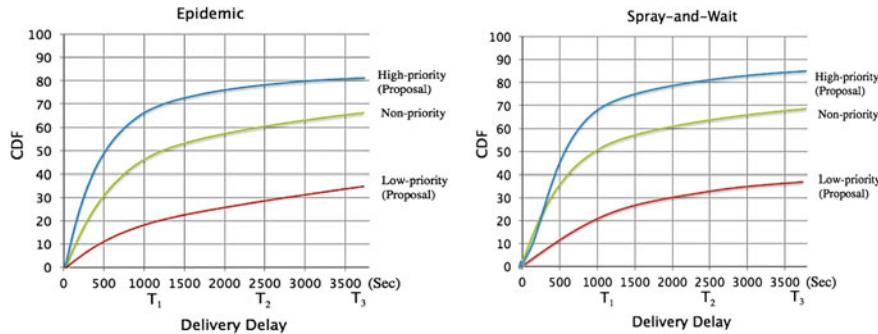


Fig. 6 CDF of message delivery delay

prioritization can deliver high-priority pieces faster than the conventional method without prioritization.

Figure 6 shows the CDF of message delivery delay. At simulation time of 1200 s, more than 70 % of high priority pieces arrive at the hospital, while about 50 % (Epidemic) and 55 % (Spray-and-Wait) of non-priority pieces. For high-priority pieces, Spray-and-Wait slightly outperforms Epidemic, while for non-priority pieces, Epidemic performs better.

We also evaluated our proposed method by introducing performance metric, “number of diagnosable images” referring to the number of images which can be diagnosed by ophthalmologist in hospital. As described in Sect. 2.1, our method manually detects eye injury part in each generated image and put a color (i.e., resolution) for each injury marked. We define that for each image, we say the image is *diagnosable* if all high-priority pieces (i.e., red, green, and blue color pieces) in the image are received at the hospital.

We show how many diagnosable images can be received at the hospital by using the conventional method in Table 2 and the proposed method in Table 3, respectively.

Table 2 Ratio of diagnosable images arriving at hospital by non prioritized method (10 images with 130 non-prioritized pieces where piece size is 25 kB)

Received message (kB)	ID of received image	ID of received pieces	# received pieces	# diagnosable image
T1 = 1250	Sa-Sh	SaP1-13, SbP1-13, ScP1-9, SdP1-5, SeP1-4, SfP1-3, SgP1-2, ShP1	Pt = 30, Pn-42	2
T2 = 1600	Sa-Si	SaP1-13, SbP1-13, ScP1-13, SdP1-5, SeP1-5, SfP1-4, SgP1-3, ShP1-2, SiP1	Pt = 42, Pn-54	3
T3 = 2110	Sa-Sj	SaP1-13, SbP1-13, ScP1-13, SdP1-13, SeP1-13, SfP1-5, SgP1-5, ShP1-4, SiP1-3, SjP1-2	Pt = 48, Pn-62	5

Table 3 Ratio of diagnosable images arriving at hospital by prioritized forwarding method (10 images with 80 high-priority and 50 low-priority pieces where piece size are 35 KB and 5 KB)

Received message (KB)	ID of received image	ID of received pieces	# received pieces	# diagnosable image
T1 = 1400	Sa-Sh	SaH1-8, SbH1-8, ScH1-8, SdH1-8, SeH1-4, SfH1-3, SgH1-2, ShH1, SaL1-3, SbL1-4, ScL1-5, SdL1, SeL1-2, SfL1-3, SgL1-4, ShL1-5	H = 32, L = 20	4
T2 = 1850	Sa-Si	SaH1-8, SbH1-8, ScH1-8, SdH1-8, SeH1-8, SfH1-5, SgH1-3, ShH1-2, SiH1, SaL1, SbL1-2, ScL1-3, SdL1-3, SeL1-4, SfL1-4, SgL1-4, ShL1-5, SiL1-5	H = 40, L = 25	5
T3 = 2400	Sa-Sj	SaH1-8, SbH1-8, ScH1-8, SdH1-8, SeH1-8, SfH1-8, SgH1-8, ShH1-5, SiH1, SjH1, SaL1, SbL1-2, ScL1-3, SdL1-3, SeL1-3, SfL1-4, SgL1-4, ShL1-4, SiL1-5, SjL1-5	H = 56, L = 35	7

As we see in Tables 2 and 3, the number of received diagnosable images at the end of simulation time is 5 for the conventional method and 7 for the proposed method. This shows that our priority-based scheme can improve reception rate of diagnosable images, and give effective medical treatment to more patients with eye injury in volcano eruption than the conventional method without the prioritization scheme.

7 Conclusion and Future Work

We designed a medical image delivery service over DTN in emergency situations, based on color marking prioritization of pieces in each injury image with color resolution value. The proposed priority forwarding routing shows better performance compared to non-priority approach used in the Epidemic and Spray-and-Wait routing in terms of message delivery rate and message delivery delay. Generalization of the proposed method and optimization of the number of medical images to be delivered to destinations are part of future work. We also plan to develop the proposed method as an Android application and conduct experiments in a real environment

Acknowledgments This work is partly supported by JSPS KAKENHI Grant Number 26220001 and Indonesia Government by DGHE-DIKTI Scholarship.

References

1. Joe, I., Kim, S.-B.: A message priority routing protocol for delay tolerant networks (DTN) in disaster areas. In: *Future Generation Information Technology in Computer Science*, pp. 727–737 (2011)
2. Mashhadi, A.J., Capra, L.: Priority scheduling for participatory delay tolerant networks. In: Proceedings of 2011 IEEE International Symposium on World of Wireless, Mobile and Multimedia Networks (WoWMoM 2011), pp. 1–3 (2011)
3. Ishimaru, Y., Sun, W., Yasumoto, K., Ito, M.: DTN-based delivery of word-of-mouth information with priority and deadline. In: Proceedings of the 5th International Conference on Mobile Computing and Ubiquitous Networking (ICMU), pp. 179–185 (2010)
4. Laptev, N.P., Chai, S., Huang, Z.: Opportunistic medical data delivery in challenged environment. http://nikolaylaptev.com/master/research/Opportunistic_medical_data_delivery.pdf. Last accessed on March 30, 2015
5. Eze, T.O., Ghasseman, M.: Heterogeneous mobility models scenario: performance analysis of disaster area for mobile ad hoc networks. In: Proceedings of London Communications Symposium (LCS), University College London (2010)
6. Wahidabu, Fa: A new queuing policy for delay tolerant networks. *Int. J. Comput. Appl.* **1** (20), 56–60 (2011)
7. Takahashi, A., Nishiyama, H., Kato, N.: Fairness issue in message delivery in delay- and disruption-tolerant networks for disaster areas. In: Proceedings of International Conference on Computing, Networking and Communications (ICNC), pp. 890–894 (2013)
8. Space Time Engineering: Scenargie® High Quality System Simulation Framework: <https://www.spacetime-eng.com/>. Last accessed on 30 Mar 2015

A Framework of Personal Data Analytics for Well-Being Oriented Life Support

Seiji Kasuya, Xiaokang Zhou, Shoji Nishimura and Qun Jin

Abstract Nowadays, we are living in a well-suited social environment with a variety of lifestyles and values. Life support has become important in such a diversified society. Along with continuously collecting the tremendous amount of personal big data generated in the social environment, it is possible for us to provide the life support based on personal data analytics. Moreover, analyzing such a kind of data can facilitate deep understanding of individual life. In this study, we focus on personal data analytics to support well-being oriented life. Three categories of personal data are classified from the collection of individuals' daily life data, and a framework of well-being oriented personal data analysis is proposed, which can provide people with suggestions and advices to improve their living life.

Keywords Personal data · Well-Being oriented life support · Big data analysis

S. Kasuya (✉) · X. Zhou · S. Nishimura · Q. Jin
Graduate School of Human Sciences, Waseda University, Tokyo, Japan
e-mail: seijino8@fuji.waseda.jp

X. Zhou
e-mail: xkzhou@ruriwaseda.jp

S. Nishimura
e-mail: kickaha@waseda.jp

Q. Jin
e-mail: jin@waseda.jp

1 Introduction

In recent years, with the high development of social computing technology, larger amounts of individual related data are generated and collected from today's digital society. Application and utilization of this kind of big data has become increasingly important, ranging from personal education to public health. In this study, we focus on utilization of personal data for well-being oriented or healthy life support. Health is generally defined as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" [1]. With the help of lifelogging and personal information management with wearable devices, more and more people are seeking to live in the suitable living environment with various senses of values. Thus, to pursue a better healthy life in the integrated cyber-physical world, it is essential to understand an individual's well-being status deeply based on organization and analysis of personal data. In this study, a framework of personal data analytics for well-being oriented life support is proposed, based on the data collected from individuals' daily activities.

The rest of this paper is organized as follows. We give a brief overview on related work in Sect. 2. In Sect. 3, after introducing the definition of well-being, we describe the basic framework of well-being oriented life support. In Sect. 4, we discuss extraction and collection of heterogeneous data in the daily life cycle, and demonstrate personal analysis toward well-being oriented life support. Finally, we conclude this study and give some promising perspectives on future work in Sect. 5.

2 Related Work

Lots of research works focus on life logs, which record our daily life as digital data. Gemmell et al. [2] introduced the design and implementation of a system named MyLifeBits, which aimed to store all of one user's digital data based on four principles. To analyze the individual's personal life information from social media data, Pan and Matsuo [3] designed several rules to discover behavior patterns based on measuring of the activity record. The observed results can be utilized to improve the management of users' personal lives in terms of "efficiency" and "mood". Teraoka [4] introduced an organizing structure and a zooming user interface in an interactive system, which enabled users to recall the collected personal data from several viewpoints, and further helped them find various related information. Bentley et al. [5] built a health mashups system as the first individually focused platform to discover the trends over time from the multiple aspects of well-being data, and discussed the behavior changes among the participants within the mobile-based environment. The observation results can not only promote the development of well-being related systems, but also benefit users' general well-being in their daily lives.

3 Well-Being and Personal Life Support

3.1 *Definition of Well-Being*

Human well-being refers to the life support of health for the service of eco-system all over the world. It is noted that human well-being concerns providing of personal safety and secure life, as well as the minimum of supplies for a good life. To this purpose, it is necessary to keep a good life living and be able to choose the social connections freely [1]. Generally, well-being is defined to keep a good status for both health and social connections. And the elements of well-being include self-acceptance, positive relations with others, autonomy, environmental control, life goals, and personal growth [6].

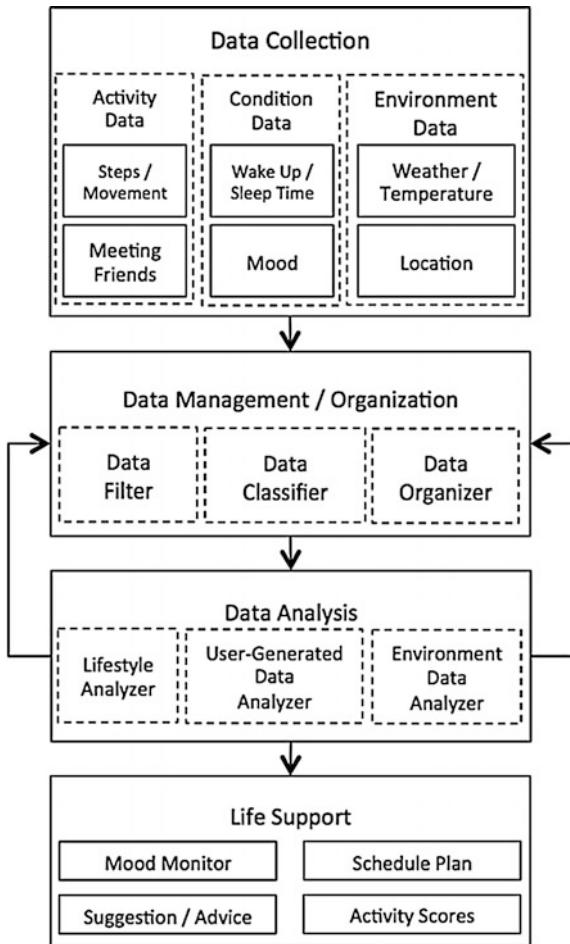
In this study, we focus more on well-being from an individual's personal experience in the daily life. Especially, it would associate with several small successful experiences which can result in the personal satisfaction, the sense of fulfillment, and the sense of achievement. In other words, through cumulating a series of piece of successful experience every day, it is considered to be able to feel a sense of satisfaction, fulfillment, and achievement. Following this way, extraction and analysis of the daily actions related to successful experience can facilitate understanding and sharing of well-being oriented human life, and provide people with adequate and sustainable life support for enhancement of social well-being and health.

3.2 *Well-Being Oriented Life Support*

Following the discussion above, in this paper, the proposed life support aims to let more and more people's life become well-being, which can continuously improve the individuals' quality of life (QoL). Differing with other research works, in addition to multi-faceted understanding of both mental and physical health, our study tries to understand the well-being oriented life from the accumulation of series of small successful experiences. That is, based on the analysis of individuals' daily actions, it becomes possible to provide adaptive support to everyone in accordance with their diversified lifestyles. Furthermore, based on the extractions of features of human-beings and the life environments respectively, both the mental and physical health can be improved according to the comprehensive analysis of their associations.

As shown in Fig. 1, four major modules, i.e., Data Collection, Data Management/Organization, Data Analysis, and Life Support, are considered to provide well-being oriented life support. Firstly, three basic kinds of life data are collected from our daily life, namely, the activity data, which includes an individual's behavioral habits to some common purposes, the condition data, which includes an individual's daily routines and moods, and the environment data, which

Fig. 1 Framework of well-being oriented life support



includes the weather/temperature information and the dynamical location data. Then, the collected data need to be pre-processed and prepared for further analysis in the Data Management/Organization module, which includes Data Filter, Data Classifier, and Data Organizer. Based on these, we analyze the individuals' daily data in three different aspects. That is, the Lifestyle Analyzer is used to analyze the different users' lifestyles, aiming to identify the diversified features in users' daily lives. The User-generated Data Analyzer is employed to extract users' behavioral features, which can provide users with the personalized recommendations. And the Environment Data Analyzer is used to analyze the corresponding data from different environments, which can provide users with timely support according to the dynamical detection of the changed environments. Finally, in the Life Support module, the integrated mechanisms can be developed to provide a specific user with

the personalized suggestions, such as the schedule plan, or activity score, to support his/her well-being oriented life.

4 Personal Data Analytics for Well-Being Oriented Life Support

4.1 Data Collection in Daily Life

Generally, to analyze personal data for well-being oriented life support, the collected data from our daily life can be classified into three basic types [7], that is, the volunteered data, observed data, and inferred data. The volunteered data refers to the data that is created and shared by individuals directly, such as the user profile. The observed data refers to the data that is collected from the record/history of the actions of users, such as the check-in data. The inferred data refers to the data that is extracted and analyzed from the volunteered or observed information, such as some trust-based values. All these kinds of data compose the so-called personal data in our daily life and can be viewed as an important part of big data. The observation, collection and analysis of this kind of personal big data benefit understanding of individuals' diversified life styles, and finally provide them with the corresponding feedback as their well-being oriented support. In other words, the data observed and collected from smart devices, smart phones, and web services can be organized on the cloud, and the integrated data along with the extracted life style can be analyzed to provide different users with the personalized feedback to support well-being oriented life.

Figure 2 demonstrates a conceptual image of collection of heterogeneous data in the daily life cycle. In details, given a specific user, his/her personal data can be collected from the morning to the midnight through a variety of smart devices. All the activity data, condition data, and environment data are detected and selected, including the movements, the weather and room temperature, and the temporal and location data for both work and leisure. In addition, the personal status can also be recorded, such as the body weight, mood, and sleeping time. All these data can be integrated and organized to provide the personalized support for the daily life.

4.2 Framework of Personal Data Analytics

Based on the discussion above, we demonstrate how to analyze the collected personal data, in order to provide users with well-being oriented life support. The framework of personal data analytics is shown in Fig. 3.

As shown in Fig. 3, based on data filtering and classification, three basic data, the activity data, condition data, and environment data, can be extracted as the observed data, while the user profile data can be employed as the volunteered data.

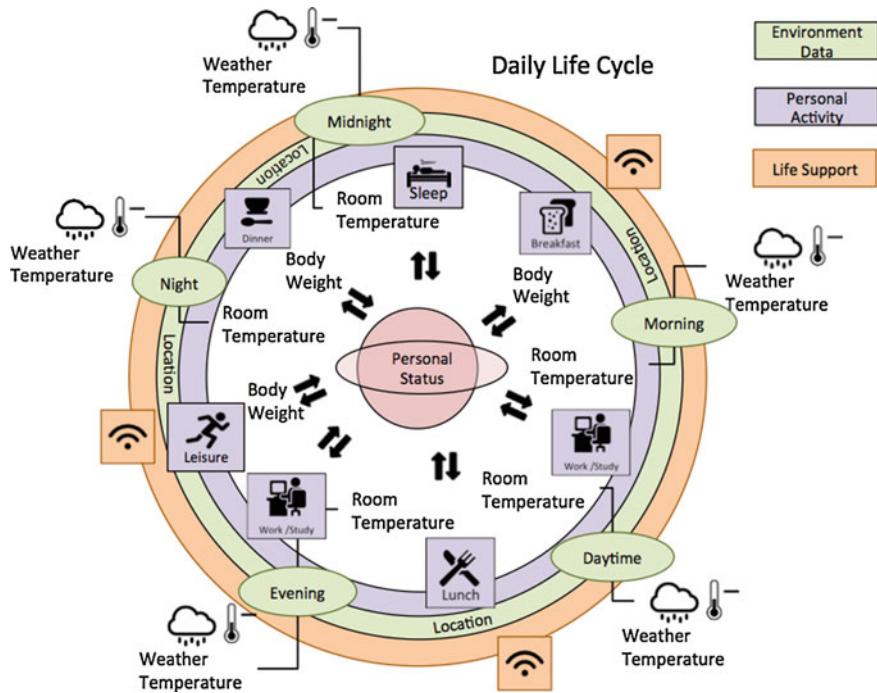


Fig. 2 Conceptual image of date collections in the daily life cycle

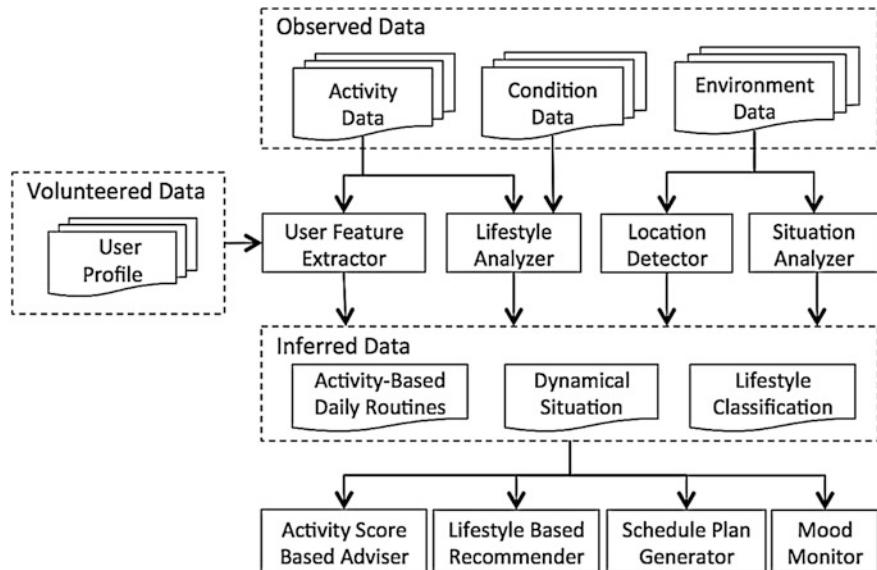


Fig. 3 Life support oriented personal data analysis

Then, four components, User Feature Extractor, Lifestyle Analyzer, Location Detector, and Situation Analyzer, are used to analyze the observed data and volunteered data, and further obtain the inferred data. In details, the activity data and users' profiles can be utilized to analyze and extract users' activity related features to refine their profiles. The activity data and condition data can be used to analyze users' different lifestyles. The environment data can be utilized to automatically detect the current information in terms of location and situation, respectively. In this way, three kinds of inferred data can be obtained, that is, the data related to users' activity-based daily routines, dynamical situation, and classified lifestyles. Based on these, we can go further to utilize these diversified data to provide users with a variety of well-being oriented life support from Activity Score Based Adviser, Lifestyle Based Recommender, Schedule Plan Generator, and Mood Monitor.

5 Conclusion

In this study, we have proposed a framework of personal data analytics to support individuals' well-being oriented life. Firstly, we introduced the definition of well-being, and proposed a framework of well-being oriented life support. Then, based on the classification of collected daily life data, three kinds of data, the volunteered data, observed data, and inferred data, were analyzed and utilized to provide individuals' with the suitable life support.

As for our future work, we will improve the design and implementation of our proposed framework with more functional modules. We will also develop the corresponding algorithms for better support provision. Experiments will be conducted to evaluate the proposed framework and application system.

References

1. World Health Organization: Ecosystems and Human Well-Being: Health Synthesis, <http://www.who.int/globalchange/ecosystems/ecosys.pdf> Accessed 7 March 2015
2. Gemmell, J. et al: MyLifeBits: Fulfilling the Memex Vision, Proceeding of ACM Multimedia, 235–238 (2002)
3. Pan, R., Matsuo, Y.: Discovery behavior patterns from social data for managing personal life. *J. Jpn Soc Artifi Intel* **28**(6), 829–834 (2013)
4. Teraoka, T.: Organization and exploration of heterogeneous personal data collected in daily life. *Hum Centric Comput. Inf. Sci.* **2**(1), 1–5 (2012)
5. Bentley, F. et al: Health mashups: presenting statistical patterns between wellbeing data and context in natural language to promote behavior change. *ACM Trans. Comput. Hum. Interact.* **20**(5), 30 (2013)
6. Ryff, C.D., Keyes, C.L.M.: The Structure of psychological well-being revisited. *J. Pers. Soc. Psychol.* **69**(4), 719–727 (1995)
7. World Economic Forum: Personal data: the emergence of a new asset class, http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf Accessed 7 March 2015

Secure Initial Attach Based on Challenge-Response Method in LTE Conforming to LTE Standards

Jungho Kang

Abstract LTE and LTE-A are globally commercialized technologies. However, when UE is executing Initial Attach processes to access LTE Network, there exists vulnerability that identification parameters like IMSI and RNTI are transmitted as plain texts. Therefore, Security Scheme was proposed in this paper that identification parameters could be safely transmitted in 4 cases where Initial Attach occurs between UE and MME. Proposed Security Scheme not only supports encrypted transmission of identification parameters but also mutual authentication between eNB and MME. Additionally, performance analysis results using OPNET simulator showed the satisfaction of average delay rate which is specified in LTE standards.

Keywords GUTI · IMSI · Initial attach · Mutual authentication · LTE

1 Introduction

Long-Term Evolution (LTE) is 4th generation mobile communication technology. Release 12 is currently underway in LTE standards and LTE network where part of Release 10 technologies are applied based on Release 9 can be classified as LTE advanced technology introduction country. Even though LTE standards which have been drafted since 2004 solved various security vulnerabilities which were found in LTE, some vulnerability have not still been security-updated but they are specified as vulnerabilities in Standards. The most issued vulnerability among various vulnerabilities is the one where the identification parameter values in UE are exposed as plain texts.

J. Kang (✉)

Department of Computing, Soongsil University, Seoul, Republic of Korea
e-mail: kjh7548@naver.com

2 LTE

2.1 LTE Network

The LTE network consists of LTE entities dealing with wireless access network technology and EPC entities dealing with core network technology. UE means user device like Smart-Phone among LTE entities. eNB, serving as the base station provides the user with wireless interface and provides wireless Remote Resource Management (RRM) features such as radio bearer control, wireless admission control, dynamic wireless resource allocation, load balancing and Inter Cell Interference Control (ICIC). MME communicates with HSS for user authentication and user profile download and executes Initial Attach and authentication. In other words, authentication of user personal profile information which is saved in HSS is executed using MME and at this time, eNB is used as a connection instrument.

2.2 LTE Initial Attach for UE and Threats

‘Initial Attach for UE’ process is a case of the first access to the network by the user subscribing the LTE network using UE. For this, eNB is selected in the process of ‘Initial State and after matching synchronization, IMSI which is an identification parameter is transmitted to request MME to access network in ‘ECM Connection Establishment’ process. IMSI refers to the unique ID requested to each user when the net administrator registers the user to service, and this value refers to the unique number of identifications saved in the USIM in the user device. IMSI which is transmitted as plain text is transmitted to MME through thousands of eNBs and has vulnerability that IMSI is leaked to attackers by malicious eNBs. Figure 1 shows the initial attach process.

3 Proposed Security Scheme

Proposed scheme was designed to protect original identification information like IMSI and RNTI which are transmitted as plain texts when UE tries Initial Attach in the network. It is composed of 4 cases in total according to UE’s initial access types and eNB and MME are assumed to be the safe channel along with backbone network. Please see Table 1 for definitions and terms used in this paper.

First protocol is carried out after Initial State after Radio Link Synchronization process is completed in Initial Attach with IMSI Case. First protocol is based on mutual authentication between UE and eNB MME and is designed to protect IMSI which is leaked as plain texts in ECM Connection Establishment process and RNTI which is leaked as plain texts in EPS Session Establishment process.

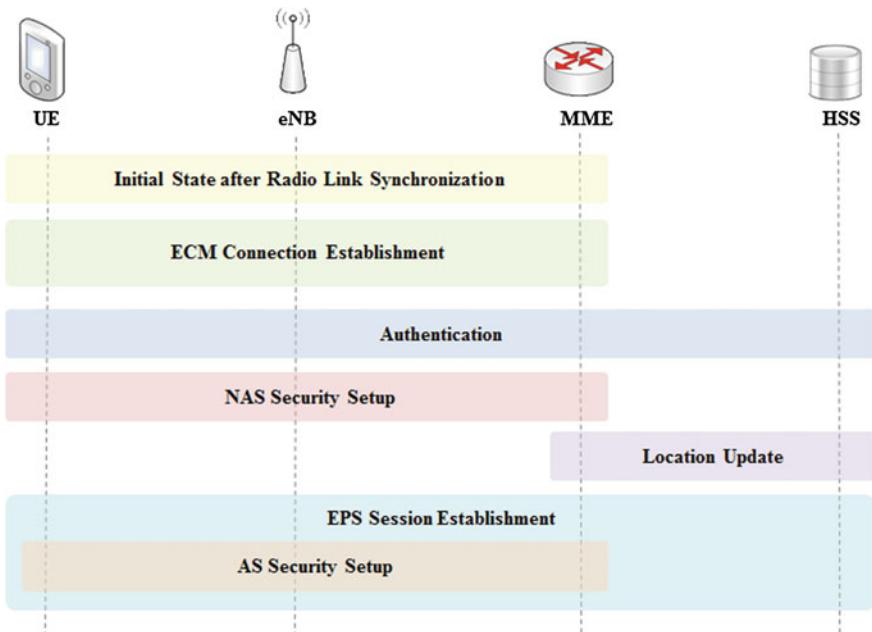


Fig. 1 Initial attach process

Table 1 The terms and symbols used in proposed security scheme

UE	User equipment
eNB	Evolved node B
MME	Mobility management entity
HSS	Home subscriber server
IMSI	International mobile subscriber identity
RNTI	Radio network temporary identities
GUTI	Global unique temporary identifier
PLMN ID	Public land mobile network ID (MCC+MNC)
MCC	Mobile country code
MNC	Mobile network code
RN	Random number
$h()$	Hash function
F	4n bits String by $h()$
C	Challenge bits
SSK	Secret sharing key
S-box	Substitution-box

UE transmits Random Number and UE Network Capability which are generated for Attach Request to MME. MME which receives Attach Request generates Random Numbers and transmit to UE, and UE and MME execute series of

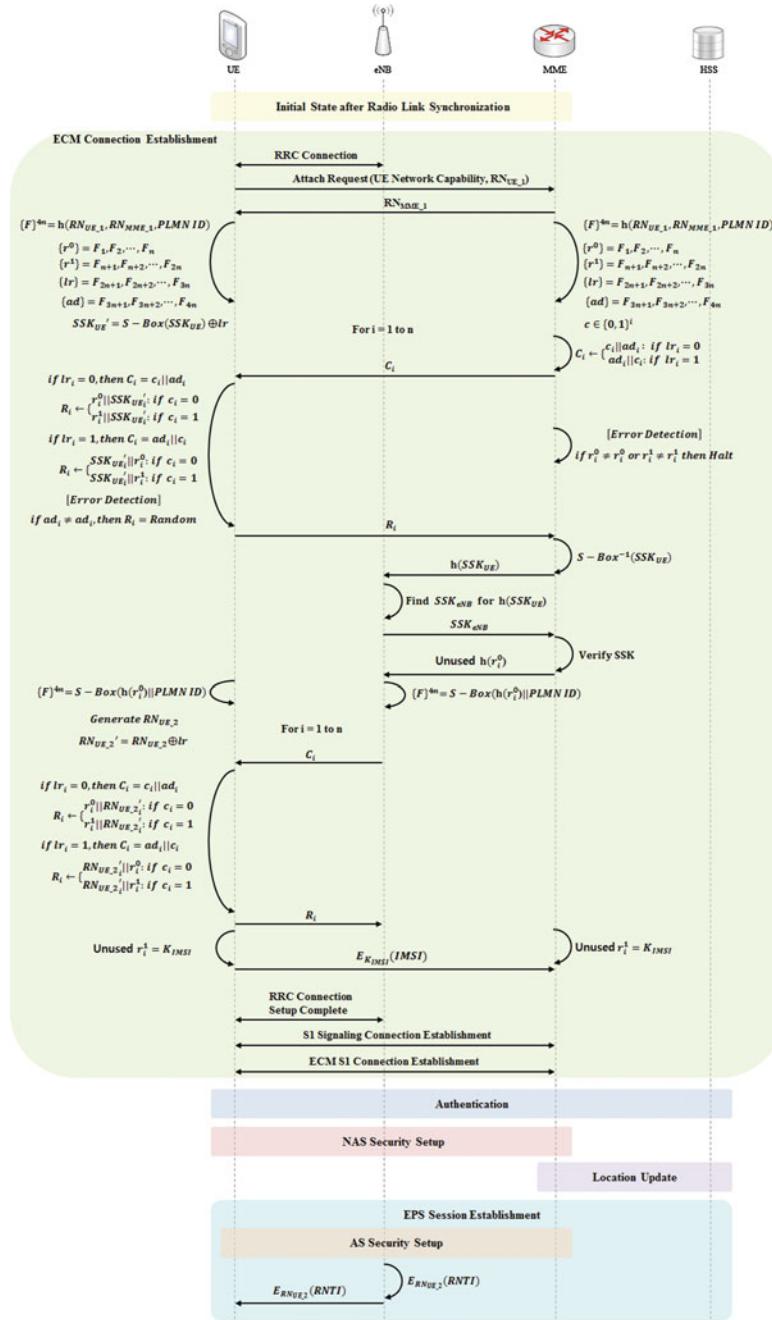


Fig. 2 Proposed protocol: initial attach with IMSI

Table 2 Performance analysis: original process

Device	Original initial attach			Proposed initial attach		
	Time (s)	Delay (s)	Traffic received (bits/s)	Time (s)	Delay (s)	Traffic received (bits/s)
1	–	0.004700	1.234	0.0113	0.007715	1.159
50	–	0.007260	1.249	0.0105	0.010539	1.331
100	–	0.010292	1.039	0.0114	0.014576	1.108

calculation to transmit IMSI safely. UE and MME inputs transmitted and received Random Numbers and PLMN ID to Hash Function which has Secret Sharing according to MNC and generates 4n bits F string.

Generated F strings are divided into 4 sequences with each n bits. MME generates random number sequences which are used as challenge bits, and UE generates SSKUE' and calculates lr sequence and exclusive OR and generates SSKUE'.

MME generates challenge bits Ci using lri, adi and ci. Ci is composed of Ci = cilladi when lri is 0, and of Ci = adillci when lri is 1. MME transmits Ci to UE and verifies UE through response values and UE verifies MME through Ci (Fig. 2).

4 Performance Analysis

The performance analysis environment is shown in Table 2. It was carried out on the LTE network components, UE, eNB and MME.

Since the existing initial attach does not perform encryption, time (s) was excluded. It was found that the delay according to the encryption process in the proposed algorithm decreased a little (somewhat) and since there was no packet switching other than the initial access process, it was analyzed that it did not have an impact on the transmission rate.

5 Conclusion

In this paper, security scheme for LTE Initial Attach Conforming to LTE standards was proposed in order to solve vulnerability where identification parameters are transmitted as plain texts before mutual authentication. The proposed security scheme generated a key through challenge-response method to encrypt and transmit the unique ID and support error detection and verification. Through security and performance analysis the safety and performance of the proposed encryption algorithm were found to be efficient.

A Design and Development of Secure-Coding Check System Based on CVE and CWE

Hyungjoo Kim and Moon-seog Jun

Abstract Recently, software has been utilized in various environments including computer, smartphone and medical devices because the application fields of IT products have been diversified. Moreover, software has evolved in a way to modify and redistribute source code freely by opening source code in recent years. However, open source software is being developed through those developers having no prior security related knowledge. Furthermore, it is being distributed without any verification. Hereupon, there are various security vulnerabilities that are exploited for an attack. Therefore, this paper examined security vulnerabilities from design phase to distribution phase of software and also proposed a system that can check whether software is securely coded. Moreover, this paper analyzed the equivalency of performance to the existing products as a result of the performance evaluation through Juliet code.

Keywords Software vulnerability • Secure-coding • CVE • CWE • Hybrid analysis • OWASP

1 Introduction

The development environment for application integration and outsourcing through high performance needs and price policy has emerged as a result of the change in the IT market. That is to say, the development environment has evolved in the

H. Kim · M. Jun (✉)
Soongsil University, Seoul, Republic of Korea
e-mail: mjun@ssu.ac.kr

H. Kim
e-mail: hyungjoo.kim@ssu.ac.kr

direction toward application integration and outsourcing to create new services by integrating and modifying the existing services in whole or in part rather than developing a project from scratch in terms of reuse or price policy of software. According to Gartner, 33 % more corporations are expected to use application integration in 2016 compared to 2013. The outsourcing ratio in 2014 was 35 %. It is expected to increase to 69 % by 2019. However, many software products are releases with several vulnerabilities derived even from the development phase due to the lack of security awareness and problem solving tools among administrators and developers.

In addition, the open source software market has grown at an annual average rate of 22.4 %. Thus, it is expected that secure coding guide will be needed for non-security experts in the entire process of software development.

Therefore, this paper proposes a system to examine secure-coding in the entire process ranging from design phase to distribution phase for software.

2 Open Source Software

This refers to free software that can be modified and redistributed without any restriction by opening source code. However, it is required to comply with the license regulation for copyright owners. Also, the term “Open” in open software does not necessarily mean “free of charge”. Rather, it refers to unrestricted use of source code and software. The global open source software market has grown at an annual average rate of 22.4 %. This indicates that the open source software market is growing exponentially as compared with the annual average growth rate of 7.6 % in the commercial software market.

Anyone can modify and redistribute source code. Also, anyone can produce source code. Thus, it can be more easily exposed to security vulnerability than open source software. It can become a target of an attacker.

The most prominent attack case for open source software is the attack case targeting the open source called ‘Heart Bleed’. The term ‘Heart Bleed’ that was found in 2014 derived from the fact that it would affect ‘Heart Beat’, the essential protocol of open SSL. Hackers were able to seize all the personal information by penetrating into the memory of web server having installed open SSL by utilizing the aforementioned bug. As a result thereof, a large amount of maintenance and repair related expenses were incurred due to a large-scale patchwork. Furthermore, it caused private information infringement.

Through the most widely used attack case of open SSL created by the experts, it can be said that security vulnerability of the current open source software is like a hidden bomb.

3 Secure-Coding Check System

As shown in Figs. 1 and 2, the support area of the secure-coding checking system proposed in this paper is different from the support are of the existing secure-coding checking system. It has the support area that can be utilized in the entire process including not only development and operation phase but also pre-distribution/operation phase of open source software (design, integration, test and distribution).

The proposed system is configured to be connected with form management system and CMS in order to support software integration, outsourcing and open source modifiers. To support the form management function, developers access and re-run security inspection when generating and modifying source code. If it is safe, then it will support the redistribution function. In the case of CMS, it was

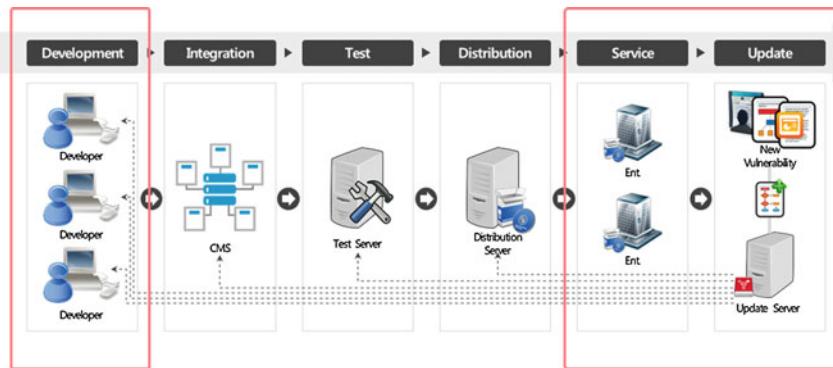


Fig. 1 Support area of the existing secure-coding check systems

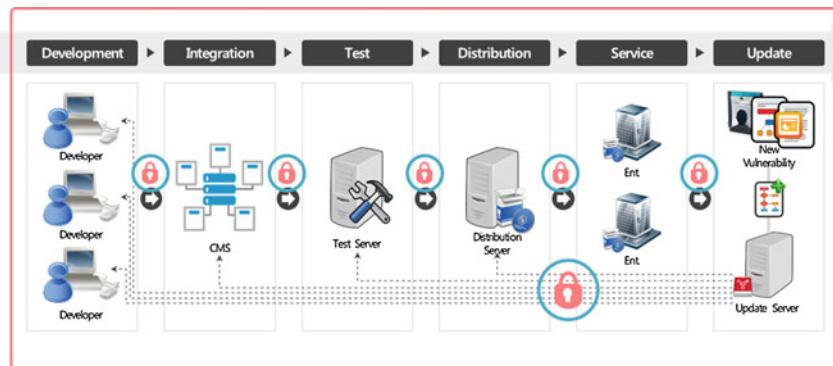


Fig. 2 Support area of the proposed secure-coding check system

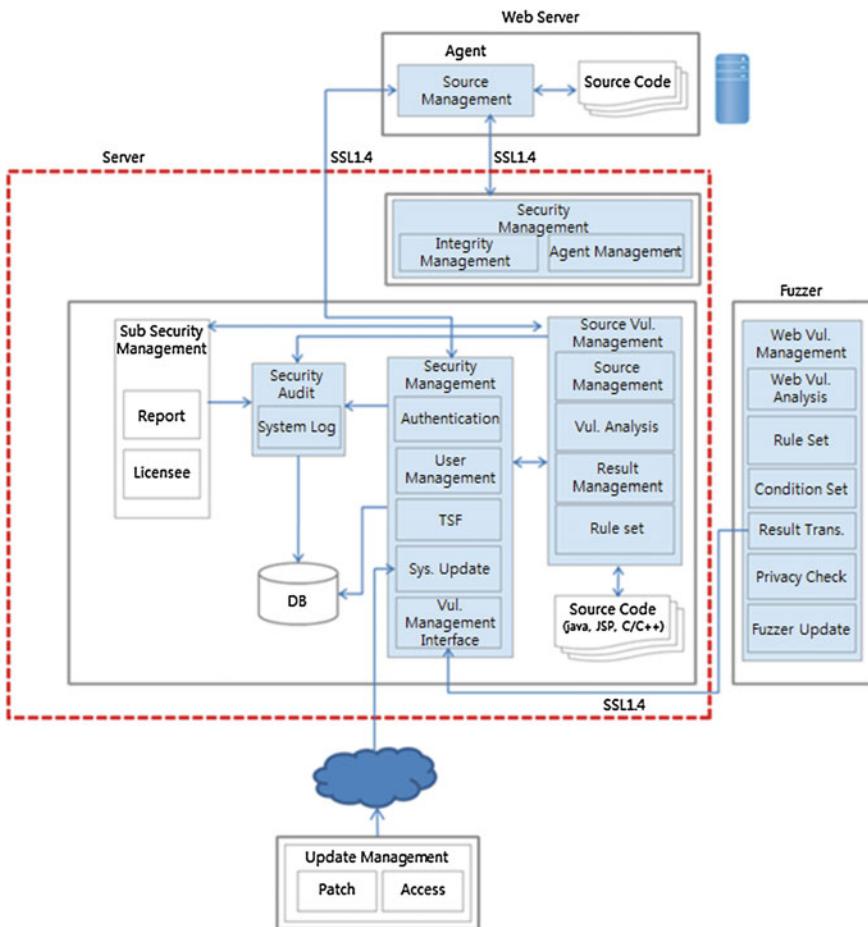


Fig. 3 System architecture 1

configured to inspect and manage security vulnerabilities in the content unit of CMS at the source code transmission module.

The proposed system architecture is as shown in Figs. 3 and 4. The source code analysis process is as shown in Fig. 5. The ruleset DB was configured based on CWE and CVE. Also, it was developed to support MDA for open source software developers.

The use of MDA supporting tools is as shown in Fig. 6. The system was configured to automatically convert vulnerability related source code through the source code automatic conversion rule mapping after automatically modeling the source code through reverse engineering after analyzing the source code analyzed as vulnerability based on compile.

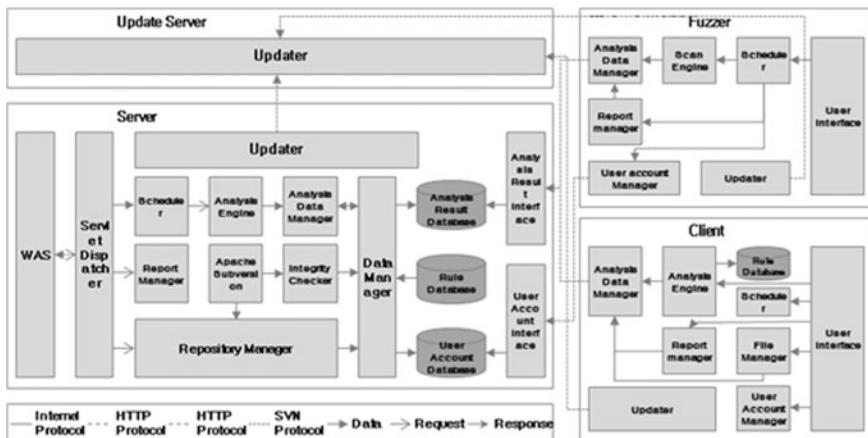


Fig. 4 System architecture 2

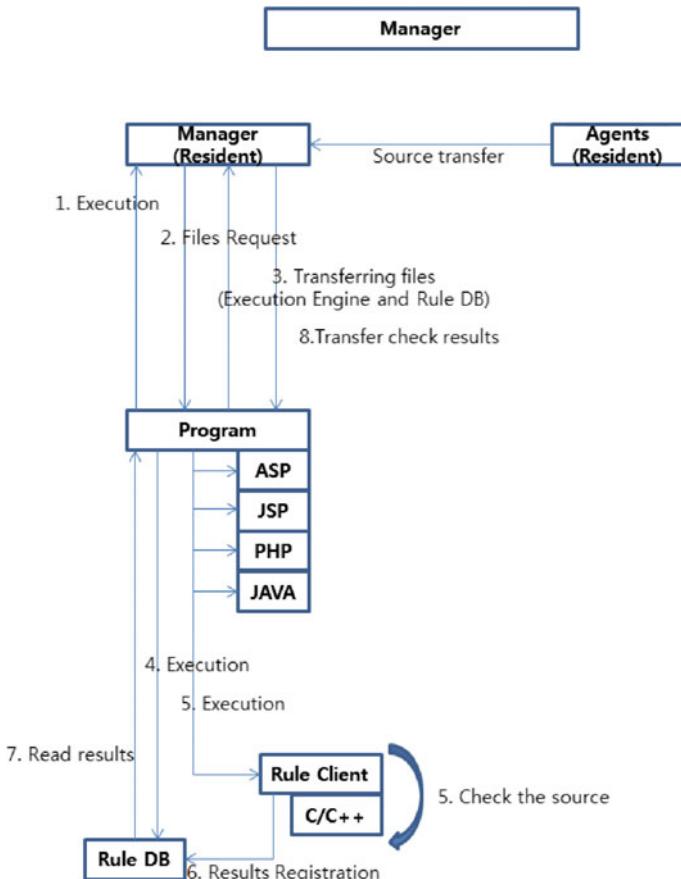


Fig. 5 Source code analysis process

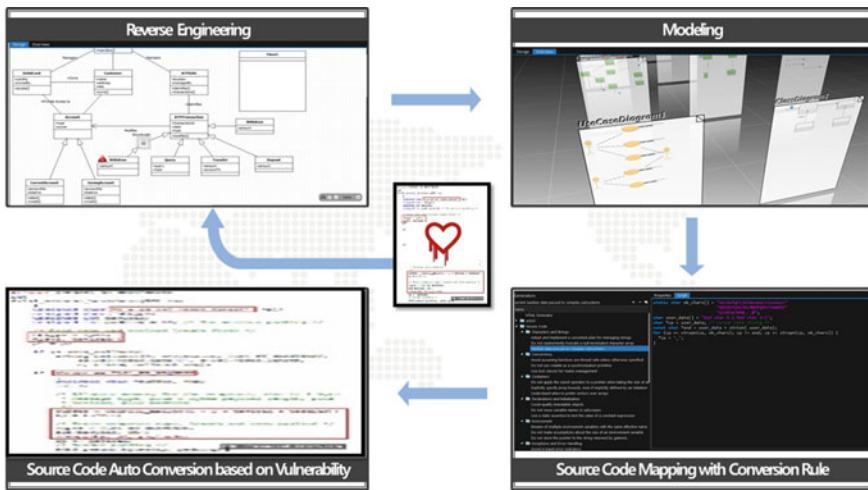


Fig. 6 Source code conversion via MDA

Lastly, the system uses the method to standardize vulnerabilities based on CWE and CVE and save it at RuleDB. If it is data present at the server after conducting the similarity analysis on new vulnerabilities with RuleDB, then it will be required to collect data by standardizing the patterns again and save them at RuleDB. Each system is being updated on a consistent basis through live update system. Thus, it is possible to respond proactively to new vulnerabilities.

4 Security and Performance Analysis

For the performance evaluation of the proposed system, the most widely used open source in the source forge was selected based on CWE/SANS Top25, OWSP 10 and Juliet code. After then, the vulnerability detection test was conducted. Table 1 shows that the open source software in the proposed system detects vulnerabilities based on CWE/SANS Top25 and OWSP 10. Table 2 is the result of returning Juliet code.

Table 1 Performance analysis based on CWE and OWSP

Vulnerability	CWE ID	Result
State directory path manipulation	23	Secure
XSS	79	Secure
Absolute directory path manipulation	36	Secure
...

Table 2 Performance analysis based on Juliet code

Vulnerability	Verify code	Result
SQL injection	char_connect_socket_mysql_query	Secure
	char_connect_socket_OCIStmtExecute	Secure
XSS	char_connect_socket_cgiFormString	Secure
...

5 Conclusion

This paper proposed a system to support for secure-coding in the entire process in which non-experts manage secure-coding process systematically and develop software accordingly. As a result of the performance evaluation, its performance was found to be as good as the performance of existing products. It was analyzed that there were even more areas to support. Moreover, it is expected that open source software developed through non-security experts will be efficiently utilized in the market by improving the source code security based on MDA.

Acknowledgments This work was supported by the ICT R&D program of MSIP/IITP. [R0112-14-1061, The analysis technology of a vulnerability on an open-source software, and the development of Platform].

Lightweight Encryption Technique for Content Security in Open IPTV Environment

Jaewoo Kim, Younggu Lee and Moonseog Jun

Abstract With the advent of open IPTV environment, a variety of IPTV devices have been developed. Open IPTV platform has a high degree of content utilization because it allows users to use various applications and services through diverse devices. Recently, many users watch IPTV using such mobile devices as smartphone and tablet PC in the aforementioned open IPTV environment. On that account, it is required to develop a lightweight encryption technique in consideration of mobile devices when applying encryption/decryption for content protection. This thesis proposes the lightweight encryption/decryption technique that complemented the specific-hardware centric technique in the existing content distribution system. The performance evaluation confirmed that the proposed technique had a higher rate of encryption/decryption and also distributed contents more securely than the existing technique.

Keywords Open IPTV · Content security · Lightweight encryption · Streaming

1 Introduction

The recent digital environment is being transformed into an environment in which there is an increasing demand for various converged services with the emergence of digital media, the evolution of bidirectional broadcasting business and the media and telecommunication convergence (diversification of contents and media).

J. Kim · Y. Lee
Soongsil University, Dongjak-gu, Republic of Korea
e-mail: saypeace@ssu.ac.kr

Y. Lee
e-mail: ad3927@ssu.ac.kr

M. Jun (✉)
Soongsil University, IT Building 402, Sangdo-Dong, Dongjak-gu, Seoul, Republic of Korea
e-mail: mjun@ssu.ac.kr

Open IPTV is certainly at the center of the aforementioned change. Hereupon, it strives to become a win-win ecosystem structure that can provide desired contents to businesses, organizations and individuals through IPTV and share profits resulting from distribution with platform providers [1].

Open IPTV platform has a high degree of content utilization because it makes it possible to use various applications and services through diverse devices. However, there is a growing demand for the protection of applications and contents due to various security threats since leakage path of personal information is becoming increasingly diverse and malicious virus creation at platform level is on the rise.

Therefore, it is imperative for open IPTV platform to take into consideration possible measures for data falsification, data leakage resulting from illegal wire-tapping, wrongful use of applications and contents and infringement incidents as a security requirement [2].

In particular, those service applications in an open platform are more vulnerable to security than in the existing closed-type platform. Thus, it is required to implement additional security measures as to the applications that are applied to such device modules as CAS (conditional access system) and DRM (digital rights management).

Moreover, it is also required to develop a lightweight security module that takes mobile devices into consideration because the service environment is moving from the existing PC online to mobile space.

This thesis proposes a technique to transmit contents more efficiently and securely by complementing the specific-hardware centric technique in the existing content distribution system and also leveraging the lightweight encryption/decryption algorithm.

2 Related Works

2.1 Content Encryption Algorithm

Broadcasting contents (video/audio/data) are transmitted mainly through transport stream (TS) packet of MPEG-2 system standard. The encryption for the protection of broadcasting contents is usually applied at the level of MPEG-2 TS [3].

Typically, the encryption at the level of MPEG-2 TS is applied to the broadcasting contents transmitted through MPEG-2 TS packet. The most frequently used encryption algorithm is DVB common scrambling algorithm. Recently, there is the tendency that AES (advanced encryption standard) is used in IPTV. In addition, there exist some other techniques of encryption and authentication using AES at the level of IP security (IPsec), ISMA encryption and authentication (ISMACryp) and secure RTP (SRTP), which include OMA BCAST service and content protection

technology, 3GPP MBMS security technology and DVB-H IPDC service purchase and protection technology [4, 5].

3 Lightweight Encryption Technique for Content Security

3.1 Key Generation and Frame Partition

First, to encrypt content, secret key shall be first created. The equation to obtain the length of secret key is as shown in (Eq. 1).

$$Key Length = (Frame_{width}) \times (Frame_{height}) \times RGB_{length} \quad (1)$$

Secret key is created through hash operation of unique key with SHA-512. The hash value obtained for each operation should be appended through conducting hash operation repeatedly as follows until the predetermined key length is satisfied.

$$\begin{aligned} K_1 &= H(Unique\ Key), K_2 = H(K_1), \dots, K_n = H(K_{n-1}) \\ Key_{Secret} &= K_1 \parallel K_2 \parallel \dots \parallel K_n \end{aligned} \quad (2)$$

After then, Each frame of contents is divided in the unit of pixel using scrambling technique. The original RGB color information of frames divided in the unit of pixel is saved in an array. When saving value of RGB color information in an array, method of listing value are the 16 ways as shown in Fig. 1. According to the first digit of the secret key, one way is selected. This saved value of RGB color information is encrypted using the proposed technique.

3.2 Encryption Process

The created secret key should conduct XOR operation with the color value in each pixel of content frame. Then, the contents should be encrypted with a modified color rather than the original color. In regard to the encryption process, the secret key should be divided in 6 digits as shown in Fig. 2 and then it should be applied to the encryption operation.

$$Modified_{frame_cv} = (Original_{frame_cv} \oplus Key_{Secret}) \quad (3)$$

If the first 6-digits in the secret key is “B04125”, then one will obtain the encrypted color value of “FFD012” by conducting XOR operation with RGB color value of “4F9137” from the content frame pixel (0, 0).

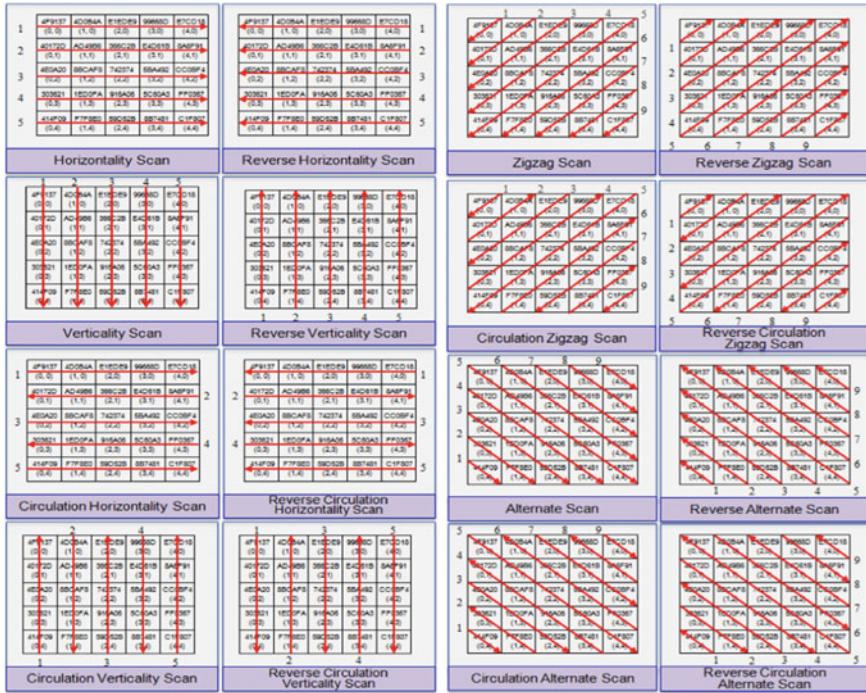


Fig. 1 Method of listing RGB color information

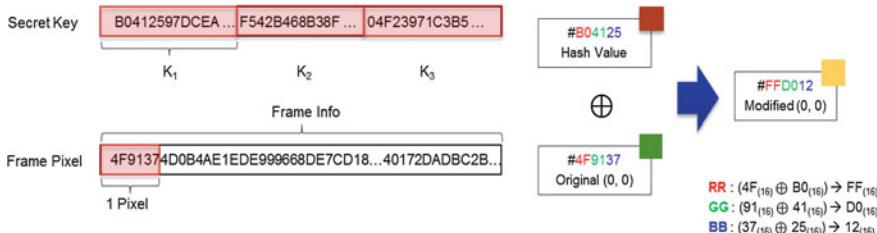


Fig. 2 Content encryption process

3.3 Decryption Process

A receiver is assumed to have received the encrypted contents and the unique key thereof securely. Thus, this receiver will conduct the following decryption process.

First, the receiver repeats to conduct hash operation on the received unique key in order to obtain the same secret key as the key used in the encryption. After then, the receiver divides the frame of encrypted contents in the unit of pixel and lists the color value of each pixel.

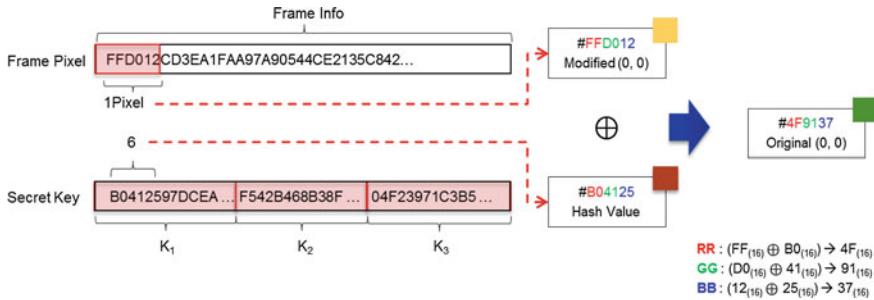


Fig. 3 Content decryption process

Then, the receiver should conduct decryption process by conducting XOR operation on the encrypted frame and the secret key as shown in Fig. 3.

$$Original_{frame_cv} = (Modified_{frame_cv} \oplus Key_{Secret}) \quad (4)$$

If the color value of the first encrypted pixel is “FFD012”, then one will be able to obtain the original color value of “4F9137” by conducting XOR operation with the first 6-digit letter of secret key that is “B04125”.

After then, rearranges the pixels in an array with the selected listing method.

4 Performance Analysis

To evaluate the performance of the proposed technique, this thesis conducted a comparative analysis on the three technologies used in the existing content encryption. These three technologies for the comparison were ATSC/CAS, ATIS/IDSA and TTA/CAS.

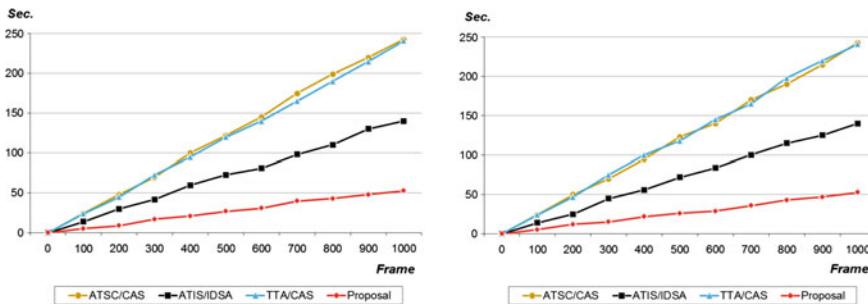
4.1 Analysis of Time for Encryption and Decryption

Table 1 shows the measurement of time to take for performing encryption/decryption per frame.

The measurement result indicates that the proposed technique’s processing rate was 4 times faster than the existing encryption algorithms. Figure 4 indicates that the performance time of encryption/decryption has a greater difference with a higher number of frames.

Table 1 Comparison with the existing content encryption techniques

	ATSC/CAS	ATIS/IDSA	TTA/CAS	Proposal
Algorithm	TDES	AES	SEED	SHA-512
Key size	168 bit	128 bit	128 bit	512 bit
Block size	64 bit	128 bit	128 bit	–
Operation mode	CBC	CBC	CBC	Hash
Encryption time	0.243 s	0.140 s	0.241 s	0.053 s
Decryption time	0.244 s	0.142 s	0.243 s	0.055 s
FPS	4.3	7.0	4.3	18.5

**Fig. 4** Encryption/decryption time

4.2 Security Analysis

Secret key is created through repeated operation of SHA-512 with the unique key of contents. On this account, it is impossible to deduct a key. Therefore, it is possible to guarantee the security of secret key under the assumption that the unique key of contents is transmitted using the algorithm of open key encryption after user authentication. It is not possible to recover an encrypted video without secret key; thus, one can prevent the threat associated with illegal wiretapping.

5 Conclusion

Open IPTV aims to become a service based on an open platform that can be applied to various devices. In particular, it is necessary to have lightweight security technologies customized for mobile environment because the number of mobile device users is on the increase.

This thesis proposed the lightweight encryption technique that can replace the existing heavy encryption algorithms. The proposed technique was designed to process encryption/decryption rapidly in real-time by utilizing only hash function

and XOR operation. As a result of the performance evaluation, it was possible to confirm that encryption/decryption could be processed at a rate that was at least 4 times faster than the existing encryption technologies. The proposed technique allows for rapid encryption/decryption. Thus, it is suitable for IPTV platform that should provide streaming service at a fast rate. Also, its operation is relatively light; thus, it is equally suitable for the use of mobile devices.

References

1. Cedervall, M., Horn, U., Hu, Y., Lvars, I.M., Nasstrom, T.: Open IPTV forum: toward an open IPTV standard. *Ericsson Rev.* **3**
2. Hwang, S.O.: Content and service protection for IPTV. *IEEE Trans. Broadcast.* **55**(2), 425–436 (2009)
3. Takahashi, A., Hands, D., Barriac, V.: Standardization activities in the ITU for a QoE assessment of IPTV. *IEEE Commun. Mag.* **46**, 78–84 (2008)
4. Qiao, L., Nahrstedt, K.: Comparison of MPEG encryption algorithms. *Comput. Graph.* **22**, 437–448 (1998)
5. Bhargava, B., Shi, C., Wang, S.: MPEG video encryption algorithms. *Multimedia Tools Appl.* **24**, 57–79 (2004)

Lightweight Mutual Authentication Routing Protocols Design Based on LEACH in Sensor Network Environment

Jaeseung Lee, Jae-pyo Park, Eunhwan Kim and Moon-seog Jun

Abstract Wireless sensor network technology is a technology that distributes nodes into various areas which is useful in various fields such as exploration for military purposes, and device management, process management, and monitoring specific areas in industry. However, sensor node in wireless sensor network due to use of compact hardware has limitations in energy, processing power, and memory storage, and to overcome these limitations various routing protocols have been suggested. However in the case of existing routing protocol, it focuses on energy efficiency, it is very vulnerable to security in mutual communication, and there is limitations in implementing previous encryption system to overcome this, due to the lack of sensor processing power and memory. Therefore, this study suggests a mutual authentication method that can respond to various security threats by simultaneously introducing mutual authentication, key generation and updating system in the communication process that also considers energy efficiency.

Keywords Sensor network • Sensor authentication • LEACH security • LEACH

J. Lee
Soongsil University, Dongjak-gu, Republic of Korea
e-mail: ljs0322@ssu.ac.kr

J. Park
Graduate School of Information Science, Soongsil University, Dongjak-gu, Republic of Korea
e-mail: pjerry@ssu.ac.kr

E. Kim
Department of Computer Engineering, Soongsil University Life-Long Education,
Dongjak-gu, Republic of Korea
e-mail: africa401@ssu.ac.kr

M. Jun (✉)
Soongsil University, IT Building 402, Sangdo-Dong, Dongjak-gu, Seoul, Republic of Korea
e-mail: mjun@ssu.ac.kr

1 Introduction

Wireless sensor network is a technology that can be used without environmental limitations in collecting data that distributes nodes into various areas which is useful in various fields such as exploration for military purposes, and device management, process management, and monitoring specific areas in industry. In sensor network it is possible to distribute nodes into desired areas according to purpose, and even in areas difficult to approach the nodes can be distributed randomly without environmental limitations. The distributed nodes communicate and exchanges data with each other, establishes network, and the collected data is sent to the base station. Here, due to the limitations in the characteristics of sensor nodes that makes up sensor network of using compact hardware, various problems can be caused due to the energy limitations of nodes. Therefore, this study suggests a mutual authentication technique to overcome the energy limitations of sensor nodes considering energy efficiency.

2 LEACH Routing Protocol

LEACH protocol is a cluster based routing protocol technology that uses the fact that energy of sensor nodes increase with communication distance, to make it possible for nodes to equally use energy according to communication distance.

Nodes in LEACH routing protocol randomly selects cluster heads and through the selected cluster heads each cluster is created. Nodes in the cluster each collect data, and collected data is sent to the cluster head. Cluster head compiles the data send by the nodes and sends it finally to the base station. In LEACH routing protocol, to distribute the concentrated energy consumption of certain nodes, method of newly selecting cluster heads by round is used to prevent phenomenon of work being concentrated on certain nodes. LEACH routing protocol selects cluster heads from nodes that did not undergo cluster head then with the node where newly selected cluster head remains, cluster is newly created.

3 Proposed Protocol

The notation used throughout the paper is shown in Table 1.

Table 1 Scheme notation

Notation	Meaning
N_v, N_p	Nonce
ID	Node ID
CID	Cluster head ID
K	Sharing key
Sk	Session key
C_i	Random bit

3.1 Cluster Head—Node Authentication

In the primary stage, the node selected as cluster head through probability based algorithm, advertises that it has been selected as cluster head through broadcast. Node i that receives the advertised message encrypts its ID and random number N_v to the cluster head of the most strongly received message to send it to the cluster head. The cluster head that receives the message encrypts its ID CID and random number N_p and sends it to node i . The cluster head and node that shares N_v and N_p uses function $f()$ to store $2 * n$ bits of information, and this is distributed evenly in n bit size.

Cluster head created random number C_i to transmit. Node that received each 1 bit from cluster heads now as a response sends 1 bit of R_0 when C_i is 0, 1 bit of R_1 when C_i is 1 to the cluster head. Cluster head creates $R_i^{c_a}$ based on the c sent to node i , and by comparing node i 's response value with aggregate value checks that data was transmitted from the right node. Cluster head that authenticated node i sends the created value from the received $R_i^{c_a}$ values and n_k in $f()$ function to node i .

Node i that receives $f(n_k, R_i^{c_1}, R_i^{c_2}, \dots, R_i^{c_n})$ from cluster head uses the same method to create $f(n_k, R_i^{c_1}, R_i^{c_2}, \dots, R_i^{c_n})'$ to compare with the value received from cluster head, to verify the cluster head. Cluster head and node uses the n bit left over from $2 * n$ bit to create secret value n_2^i that will be used later and stores it to complete the authentication process (Fig. 1).

3.2 Cluster Head—Base Station Authentication

Cluster head that is selected through probability based algorithm encrypts its CID and N_p to send it to the base station to notify it has been selected as cluster head. Base station that received the message from cluster head encrypts its BID and N_v to send it to the cluster head.

The cluster head and base station that shares random number uses function $f()$ to store $2 * n$ bits of information, and this is distributed evenly in n bit size.

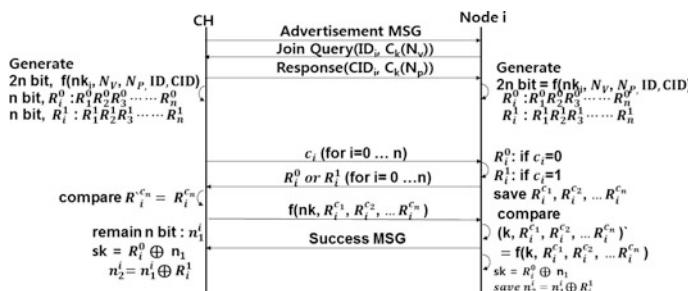


Fig. 1 Cluster head—node authentication

To undergo the authentication process the base station creates random number C_i to send it by bit to node i . Cluster head created random number C_i to transmit. Cluster head that received each 1 bit from base station now as a response sends i bit of R_0 when C_i is 0, or i bit of R_1 when C_i is 1.

Base station creates $R_i^{C_a}$ based on the sent c , and compares the cluster head's response value and aggregated value to confirm that data was transmitted from the right node. Base station that authenticated cluster head uses the received $R_i^{C_a}$ values and n_k in $f()$ function to create value to send to cluster head. Node i that receives $f(nk, R_i^{C_1}, R_i^{C_2}, \dots, R_i^{C_a})$ from cluster head creates $f(nk, R_i^{C_1}, R_i^{C_2}, \dots, R_i^{C_a})'$ using identical method to compare it to the value received from the cluster head to verify the cluster head. Cluster head encrypts n bit that was left over from $2 * n$ bit to send to base station, and base station completes the authentication process by storing each of the $n_2^1, n_2^2, \dots, n_2^i$ bit received from each cluster heads (Fig. 2).

3.3 Key Updating and Routing Protocols

When the new round starts, through probability based algorithm, new cluster head is selected. The selected cluster head advertises that it has been selected as the cluster head to the nodes by broadcast, and through responses from nodes creates cluster.

When new cluster head and node i registration and response is completed. Base station and cluster heads repeats 3.2's Distance Bounding BS-CH process. Cluster heads sends the ID's of the nodes currently connected to itself to the base station.

Base station derives the n_2^i stored in 3.2 according to node ID sent by cluster head to encrypt it, then it sends it to the new cluster head. Cluster head and node calculates the n_k that will be mutually used in $f()$ and processes the Distance Bounding CH-Node authentication of 3.2. Cluster sends the new values derived from the result of authentication between cluster head and node to the base station, and the base station stores this and when cluster head is newly selected, uses it again (Fig. 3).

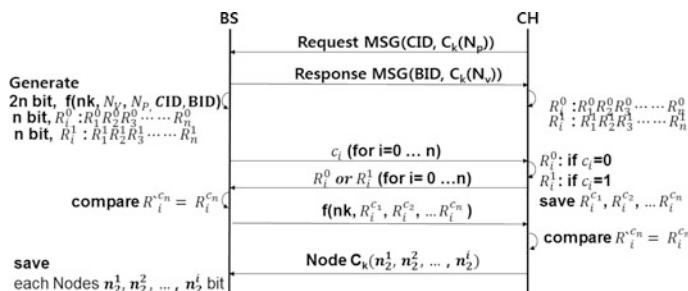


Fig. 2 Cluster head—base station authentication

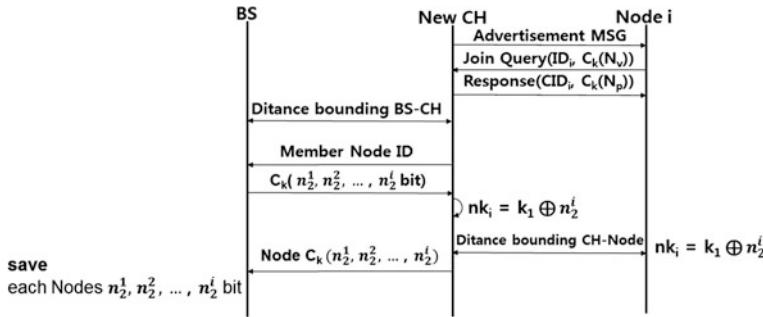


Fig. 3 Key updating and routing protocol

4 Security Evaluation

Sensor network environment inherits the same security threats and vulnerabilities that the previous information network had. Thus in the performance evaluation, the suggested protocol's safety is tested on the security threats the previous information network and additional security threats. Table 2 is a safety comparative analysis result of the previous method and suggested protocol in security threats that could occur in information network.

In this study, when Distance-Bounding protocol is used, each base station, cluster head, and sensor node created random number is used to create R_i^0, R_i^1 , and because it uses R_i^0, R_i^1 that is promised response value of random value c mutual authentication is possible, and because during message reuse attack and message forgery/corruption attack it uses new session key $sk = R_i^0 \otimes n_1$ that is newly created and different from the session key when it was stolen, it is safe against attack through reuse of message.

In the case of sniffing, because the messages that are transmitted between nodes use continuously updated inter node security key $nk_i = k_1 \oplus n_2^i$ to encrypt before sending, thus it is safe. In case of spoofing, they are already mutually authenticated nodes and because even during spoofing attack the security key that is primarily shared cannot be known, it is safe. Also, in side channel attack the transmitted message size is always same regardless of length of data, therefore it is safe.

Table 2 Security evaluation table

	Proposed protocol	Leach	Leach-C	Heed
Mutual authentication	o	x	x	x
Re-use attack	o	x	x	x
Forgery attack	o	x	x	x
Sniffing	o	x	x	x
Spoofing	o	x	x	x
Side channel attack	o	x	x	x

5 Conclusion

Considering the performance limitations of sensor network hardware the study suggested a mutual authentication routing protocol that supports mutual authentication even with low energy consumption by changing LEACH routing protocol and Distance-Bounding protocol that considers energy efficiency that can respond to various security threats using methods such as session key creation.

Method Research for Safe Authentication in Cloud Environment

Junho Song, Jaesoo Kim, ManSik Kim and Moon-seog Jun

Abstract Cloud computing technology provides services in various fields. Although cloud computing has a different from than previous architecture, the applied functions and services are based on previous architecture. This study suggests a multi-session authentication method based on the characteristics of SaaS (Software as a Service) among cloud services. The suggested technology safely creates session where keys are exchanged using cloud virtualization layer, and provides efficiency with lightweight key updating process. It is expected that through the suggested method it will be possible to provide safe environment for cloud based services.

Keywords Cloud · Service · Personal cloud · SaaS · Mutual authentication

1 Introduction

Cloud computing is a computing technology of internet based cloud, and Gartner defines it as “A service structure where using internet technology virtual IT resources are provided as service and fees are charged according to the used

J. Song · M. Kim · M. Jun (✉)
Soongsil University, Seoul, Republic of Korea
e-mail: mjun@ssu.ac.kr

J. Song
e-mail: jhsong@ssu.ac.kr

M. Kim
e-mail: sirano979@gmail.com

J. Kim
Seoul National University of Science and Technology, Seoul, Republic of Korea
e-mail: jskim@seoultech.ac.kr

resources" It is predicted that global cloud market will grow 23.5 % annually to 108 billion in 2017. Early cloud environments focused on security in platform and physical environment and stable operation of services, but as recently the target of services expand and diverse services are created, the range of security requirements and the importance of target is changing. This study suggests user authentication and permission management method for SaaS (Software as a Service) based service which is the most commonly used among Public Cloud Computing Services.

2 Related Research

Because in cloud computing is in virtual environment and the resources are all focused on the cloud center, there is high risk of threats being concentrated, such as becoming targets for hackers and internal data leaks. However because it has characteristics of having separated application layer by creation of virtualization layer, it is analyzed that it is safe in terms of security.

Currently in research for cloud authentication, there is study being done for applying previous authentication system on IDP (Identity Provider), RP (Relying Party), and trusted 3rd party objects. By applying research about how cloud computing's virtualization area can separate from previous application layer to have security, environment can be composed with cloud manager that manage access in the application layer and internal cloud service users as objects of separate layers. This structure provides background where the previously studied inter 3rd party exchange protocol can be improved and applied, and through the use of this authentication system through inter 3rd party key exchange protocol is established [1–16].

3 Interdependent Multi-Session Authentication Technique in Cloud Environment

The suggested technique is based on cloud environment composed of PC (personal computer) where the user receives service, SS (service server) that provides service in virtualization area, and BS (business support server) that manages authentication and permission information. Also the user uses mobile network through MD (mobile device) to acquire 3rd party channel for safe transmission of authentication information.

PC (personal computer), SS (service server), and BS (business support server) all each undergoes 1:1 service object authentication, and through inter 3rd party key agreement mutually authenticates, and conducts lightweight key updating process to protect the session key.

3.1 Key Agreement Process

The basic structure of key agreement protocol is password based key agreement protocol, and to secure safe channel it uses open key encryption protocol of service object authentication. According to the previously set process BS (business support server) receives PC (personal computer) and SS (service server)'s previous information, and for this through credential certification procedure creates permission maintenance session. PC (personal computer) and SS (service server) can decide on validity and respond through evaluation of session information. Detailed procedure follows key agreement process in Fig. 1.

3.2 Lightweight Key Updating Process

The lightweight key updating process continuously updates the user's use permissions by creating pseudo-random number at the time when the PC (personal computer) attempts log-in. Key updating is done through method of creating

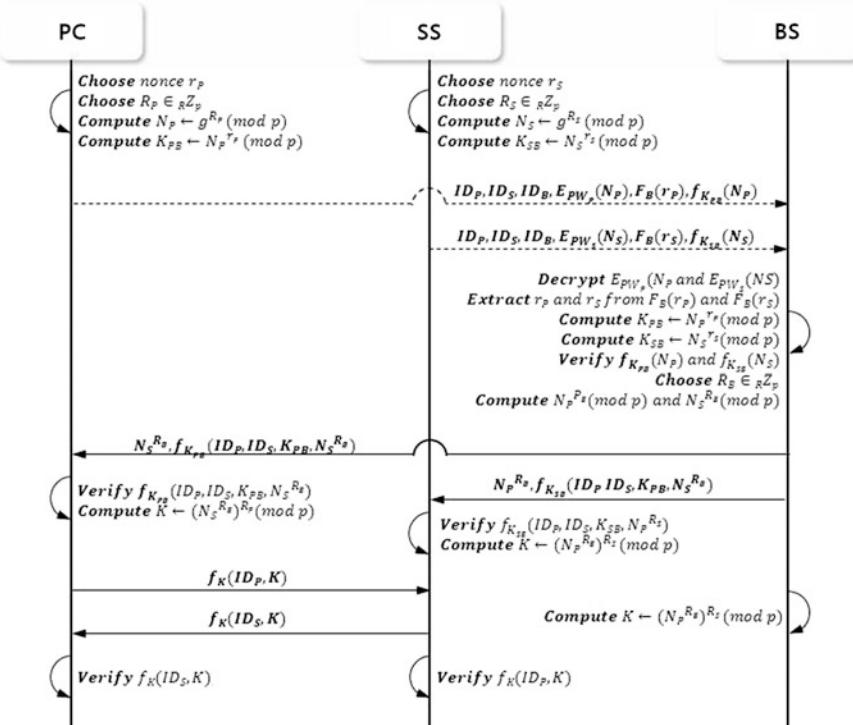


Fig. 1 Previous shared information based key agreement process

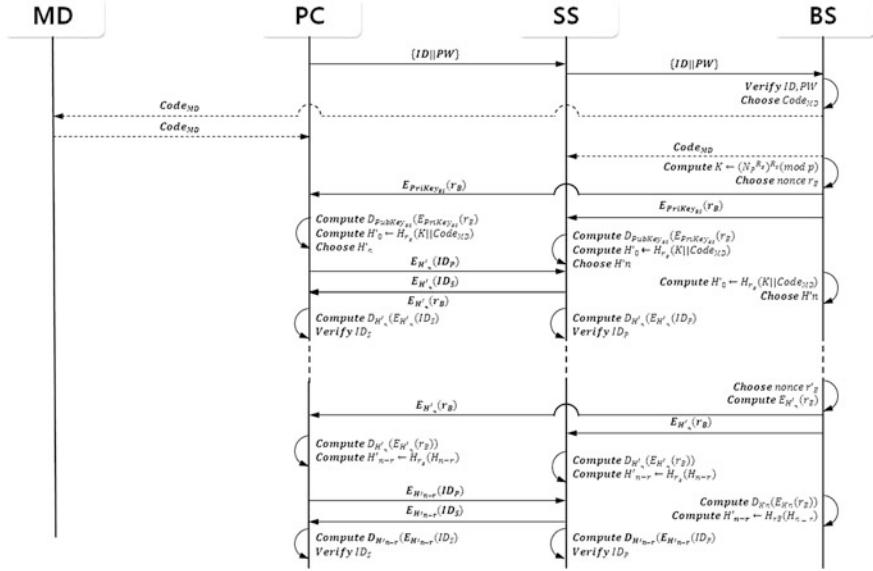


Fig. 2 Lightweight key updating process

temporary authentication key according to procedure and using this as initial key value to create new keys, and detailed procedure follows lightweight key updating process in Fig. 2.

4 Safety and Performance Analysis

4.1 Security Analysis of Key Agreement Protocol

1. Mutual authentication

First PC (personal computer) and SS (service server) uses function F_s to hide random number r_A and r_B that encrypts N_F and N_S in step 1 of procedures and PW_F and PW_S . After because it only knows trap door and PW_F and PW_S , after receiving message from key agreement 1st step, only BS (business support server) can authenticate PC (personal computer) and SS (service server). Second, in step 2 BS (business support server) sends $\{N_S^{R_S}, f_{K_{PB}}(ID_P, ID_S K_{PB} N_S^{R_S})\}$ to PC (personal computer) and $\{N_P^{R_S}, f_{K_{SB}}(ID_P, ID_S K_{SB} N_P^{R_S})\}$ to SS (service server), and as mentioned in key agreement step 2 these messages are used to authenticate BS (business support server). Third, PC (personal computer) and SS (service server) each derives its keys from $N_S^{R_S}$ and $N_P^{R_S}$. Through $f_K(ID_S, K)$ and $f_K(ID_P, K)$ PC (personal computer) and SS (service server) can authenticate each other.

2. Resistance against password guessing attack

First, because there are messages sent by PC (personal computer) to server $E_{PW_F}(N_P), F_B(r_P), f_{K_{PB}}^{(N_P)}$ and messages sent by SS (service server) to server $E_{PW_S}(N_S), F_B(r_S), f_{K_{PB}}^{(N_P)}$, there is no chance of A and B guessing each other's passwords. Second, when message $E_{PW_F}^{(N_P)}, F_B(r_P), f_{K_{PB}}^{(N_P)}$ is leaked to the attacker, because when encrypting N_F using password PW_F , r_P for opening trap door that that only the server knows, $K_{PB} = N_P^{r_P} \pmod{p}$ cannot be checked without N_P and r_P , thus from the leaked message no information can be obtained. Therefore 3rd party cannot attempt any type of attack. Third, even in the case where password PW_F is guessed and N_P decrypted by $E_{PW_F}(N_P)$ is exposed, to receive authentication from BS (business support server) $K_{PB} = N_P^{r_P} \pmod{p}$ must be known. According to the complexity of discrete logarithms, safety is acquired.

4.2 Security of Lightweight Key Updating

Lightweight key updating method provides security using safe channel through MD (mobile device) and unidirectional character of hash function. Hash function has characteristics such as collision resistance, inverse resistance and 2nd inverse resistance, and through key induction function the safety can be proven. It provides safety using stretching method that increases attack requirements by repeatedly applying improved key lengths such as SHA-256 and complex operations, and by using MD (mobile device) code value and Salt in the procedure to prepare for off-line attack.

4.3 Efficiency of Lightweight Key Updating

The quantified data was organized on Fig. 3 based on the range of execution time identical to the primary authentication process. The small line represents the

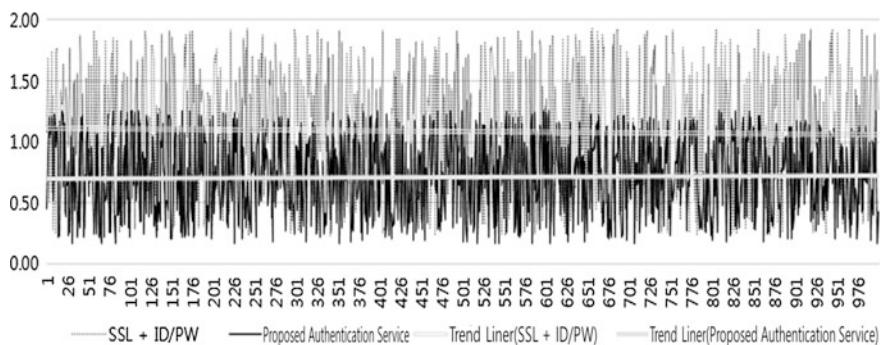


Fig. 3 Key updating process execution time

execution time of applying the procedure of suggested technique, and the dotted line represents the execution time of applying the previous procedure. About the procedure of suggested technique the average value through linear trend analysis was expressed as a thick white line, About the previous procedure average value through linear trend analysis was express as a double white line.

5 Conclusion

This study suggested authentication process and multi-session authentication technique to provide safe cloud service authentication. Using the characteristic of platform provider and service provider being separated due to virtualization among cloud computing environmental characteristics, inter 3rd party protocol session key was created, and technique to encrypt session and authenticate user using session key was suggested. This suggests a plan that can improve previous authentication methods through using the structural characteristics of cloud. Based on this there needs to be continuous research on methods to improve inter 3rd party key exchange algorithms and actual service application plans.

References

1. Kang, Y.W.: Recent trends of cloud computing services. *NET Term* (2012)
2. Plummer, D.C., Bittman, T.J., Austin, T., Cearley, D.W., Smith, D.M.: *Cloud Computing: Defining and Describing an Emerging Phenomenon*. Gartner Research (2008)
3. IDC. Worldwide and regional public IT cloud services 2013–2017 forecast (2013)
4. Gartner. Forecast overview: public cloud services, worldwide, 2011–2016 (2013)
5. Heo, E.N.: Personal cloud security technology and privacy. *Telecommun. Technol. Assoc. TTA J.* 139 (2012)
6. Yang, H.S.: A study on improvement stability of cloud service using attack information collection. *Korea Soc. Digital Ind. Inf. Manag.* 9(2) (2013)
7. Yang, J.M.: A study on improving the reliability of cloud computing. *Korea Soc. Digital Ind. Inf. Manag.* 8(4) (2012)
8. Im, C.S.: Cloud computing security technology. *Korea Inst. Inf. Secur. Cryptology* **19**(3), 12–15 (2009)
9. Kwon, A.: Changes in the IT ecosystem and countermeasures according to the diffusion cloud services. *Korea Information Technology Service Industry Association*
10. Yoon, H.Y.: Cloud services trends and issues. *Korea Communication Agency* (2011)
11. Kim, K.S.: Domestic and foreign market and policy trends in cloud computing. *National IT Industry Promotion Agency* (2013)
12. IBM. Defining a framework for cloud adoption. *IBM Global Technology Services, thought leadership white paper* (2010)
13. Kim, Ki-Chul: A security evaluation criteria for korean cloud computing service. *Korea Inst. Inf. Secur. Cryptology* **23**(2), 3–17 (2013)
14. CSA. The notorious nine: cloud computing top threats in 2013. *Cloud Security Alliance* (2013)

15. Ko, K.S.: A study on security-enhanced cloud service E&C (evaluation and certification scheme). *J. Secur. Eng.* **9**(6), 481–494 (2012)
16. Plummer, D.C., Bittman, T.J., Austin, T., Cearley, D.W., Smith, D.M.: Cloud computing: defining and describing an emerging phenomenon. Gartner Research (2008)

A Design of Dual Encryption Based on Data Obfuscation and Mutual Authentication in the Smart-Grid Environment

Wonkyu Choi, Jungoh Park, Jaesik Lee and Moon-seog Jun

Abstract On this thesis, cross certification is conducted by using secret key which is shared between each node of Server—DCU—Smart Meter. The process of cross certification is conducted through calculation of hash algorithm of secret key and random numbers. Using key values and random numbers, continuously renewed during certification process, creates secret key between each node, and it is used as double password key to enhance the safety of transferring process of later sent control data, account data, inspection data. Also, when inspection data message is collected, it is possible to infer user's daily pattern, so a protocol is suggested, which duplexes inspection data and time data to metering data and transferring it with the same padding value.

Keywords M2M · Smart-Grid · AMI · Mutual authentication

1 Introduction

Out of numerous technologies using Machine-to-Machine (M2M), a smart grid the most often seen and used is formed by fusion of existing electricity and information network system. Communications among numerous devices brings

W. Choi · M. Jun (✉)
Soongsil University, Seoul, Republic of Korea
e-mail: mjun@ssu.ac.kr

W. Choi
e-mail: gktechrm@ssu.ac.kr

J. Park
School of Electrical Engineering, Dongyang Mirae University, Seoul, Korea
e-mail: jopark13@dongyang.ac.kr

J. Lee
Korea Internet and Security Agency, Seoul, Republic of Korea
e-mail: j30231@kisa.or.kr

out effective benefits such as users' power usage efficiency, cost saving, and preventing wasting energy, and provides Outage Management System (OMS) and Demand Response (DR) enabling bi-directional data transmission between electricity providers and consumers. In this process as the data collected equal to the amount of electricity used by time can invade user privacy, in case the information is leaked, personal information such as users' life pattern and privacy can be invaded. Further, as additional encryption process is not applied to transmitting MD, security threat on personal information exists. Hence, dual encryption technique is set forth in this research to safely transmit metering data, control information, accounting information, and mutual certification between Data Central Unit (DCU) collecting data from the smart meter (SM) of the server and each node in initial set up process [1–11].

2 Suggestions

2.1 Mutual Certification-Based Dual Encryption Technique

The system set forth distributes the private key between Server—DCU, Server—Smart Meter, and DCU—Smart Meter in a reliable server to each node. Mutual certification performs between nodes by using the distributed private key and random value, and links to the session through time stamp values and arithmetic operation by consistently updating the private key used in processing mutual certification. After this, by using an encrypted key in generating sessions, the encrypted key is once again encrypted to transmit control information between Server—Smart Meter, and the double encryption that is once again encrypted between Server—DCU is transmitted to the smart meter. Also for the metering data as what is conducted to transmit control information, the encrypted key between DCU—Smart Meter is once again encrypted, and by once again encrypting the encryption between Server—Smart Meter, the double encryption is transmitted. As the double encryption is applied, relevant information cannot be accessed in the DCU, intermediate node, and even if a packet is revealed by malicious invaders, relevant information cannot be accessed as all the information about the key values between Server—DCU, DCU—Smart Meter, and Server—Smart Meter is required.

2.2 Mutual Certification Protocol

Figure 1 shows mutual certification procedures of each node among Server—DCU—Smart Meter

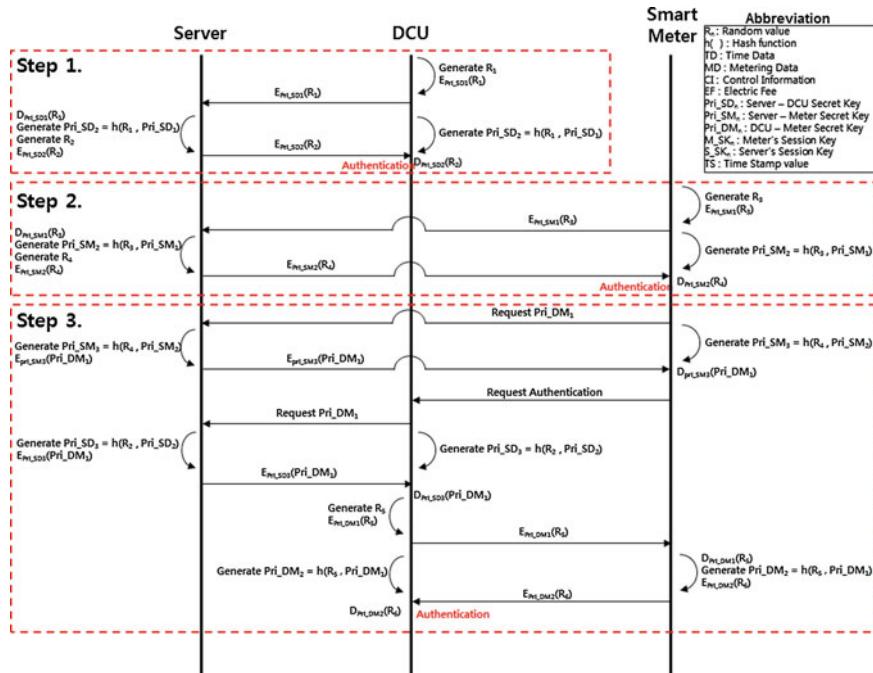


Fig. 1 Mutual certification protocol among Server—DCU—Smart Meter

2.3 Step 1. Mutual Certification Between Server—DCU

- After generating R_1 a random value, the DCU encrypts Pri_SD_1 a private key between Server—DCU to transmit the private key to the server.
- To generate a second private key between Server—DCU after decrypting the encryption transmitted $E_{Pri_SD1}(R_1)$, the server processes hash algorithm for the R_1 obtained from the decryption and the first private key between Server—DCU.
- The server R_2 generates the second random value, and encrypts the second private key between Server—DCU to transmit it to DCU.
- After hash algorithm is used to generate the second private key between Server—DCU in the DCU, the encryption transmitted from the server is decrypted by using the second key as the decryption.

Through the process stated above, the Server and DCU encrypt and transmit a random value, R_1 and R_2 by using the private key of Server—DCU shared beforehand, and through hash algorithm of transmitted values and the private key, mutual certification is completed to check whether or not the correspondent node between Server—DCU is an appropriate node, after checking the values transmitted by repeating the key updating process.

In Step 2, the certification process performed by the mutual certification process between Server—Smart Meter is performed as the same as what is conducted in Step 1.

2.4 Mutual Certification Between DCU—Smart Meter

- For mutual certification with DCU, Smart Meter requests the private key between DCU—Smart Meter to the server.
- The server requested of Pri_{DM_1} transmits the encrypted private key between DCU—Smart Meter by generating Pri_{SM_3} through hash algorithm of Pri_{SM_2} and a random value R_4 lastly used in the process of mutual certification between Server—Smart Meter.
- To generate Pri_{SM_3} a decrypted key, the smart meter receiving $E_{Pri_{SM_3}}(Pri_{DM_1})$ decrypts the encryption transmitted after generating Pri_{SM_3} through hash algorithm of Pri_{SM_2} and R_4 a random value lastly used in the process of mutual certification between Server—Smart Meter.
- The smart meter obtaining a private key between DCU—Smart Meter requests mutual certification with the DCU, and the DCU requested of mutual certification requests a private key between DCU—Smart with the server.
- The server requested of Pri_{DM_1} transmits an encrypted private key between DCU—Smart Meter by generating Pri_{SD_3} through hash algorithm of Pri_{SD_2} and a random value lastly used in the process of mutual certification between Server—DCU.
- To generate Pri_{SD_3} a decrypted key, the DCU receiving $E_{Pri_{SD_3}}(Pri_{DM_1})$ decrypts the encryption transmitted after generating Pri_{SD_3} through hash algorithm of Pri_{SD_2} and R_2 a random value lastly used in the process of mutual certification between Server—DCU.
- The DCU generates R_5 a random value, and encrypts by using a private key between DCU—Smart obtained from the server in order to transmit the encrypted R_5 to Smart Meter.
- The smart meter receiving $E_{Pri_{DM_1}}(R_5)$ decrypts the encryption transmitted by using a private key between DCU—Smart Meter, generating the second private key between DCU—Smart through hash algorithm of the decrypted value R_5 and the private key between DCU—Smart Meter. After encrypting R_6 a new random value generated from encrypting the second private key, it is transmitted to the DCU.
- To generate Pri_{DM_2} a decrypted DCU key receiving $E_{Pri_{DM_2}}(R_6)$, decrypted is the encryption transmitted after generating Pri_{DM_2} through hash algorithm of R_5 a random value and Pri_{DM_1} .

Mutual certification between DCU—Smart Meter is completed through the process stated above, and final mutual certification between Server—DCU—Smart Meter is completed through a private key of each node and a random value.

2.5 Step 4. Session and Data Transmission Key Generation

- To safely transmit MD after completing mutual certification between Server—DCU
- Smart Meter, a session obtained from time stamp value is generated and an encrypted key is used by using a key value obtained from generating the session.

In the process of generating encrypted keys, an encrypted key for safe data transmission by generating a value of the private key used in mutual certification between nodes, and a session key through arithmetic operation of time stamp values.

Figure 2 is a protocol for session generation and safe transmission of MD.

The smart meter completing mutual certification generates M_SK_1 to safely transmit MD collected from sources of electricity demand such as households or buildings.

- M_{SK_1} generates a time stamp value and hash algorithm through arithmetic operations after processing Pri_{DM_2} a private key used in mutual certification between DCU—Smart Meter, Pri_{SM_2} a private key used in mutual certification between Server—Smart Meter, and XOR operations.
- After generating M_{SK_1} , and the third private key Pri_{DM_3} is generated through hash algorithm operations of Pri_{DM_2} and R_6 a random value used in processing mutual certification DCU-Smart Meter, Pri_{DM_3} is encrypted to transmit M_{SK_1} .

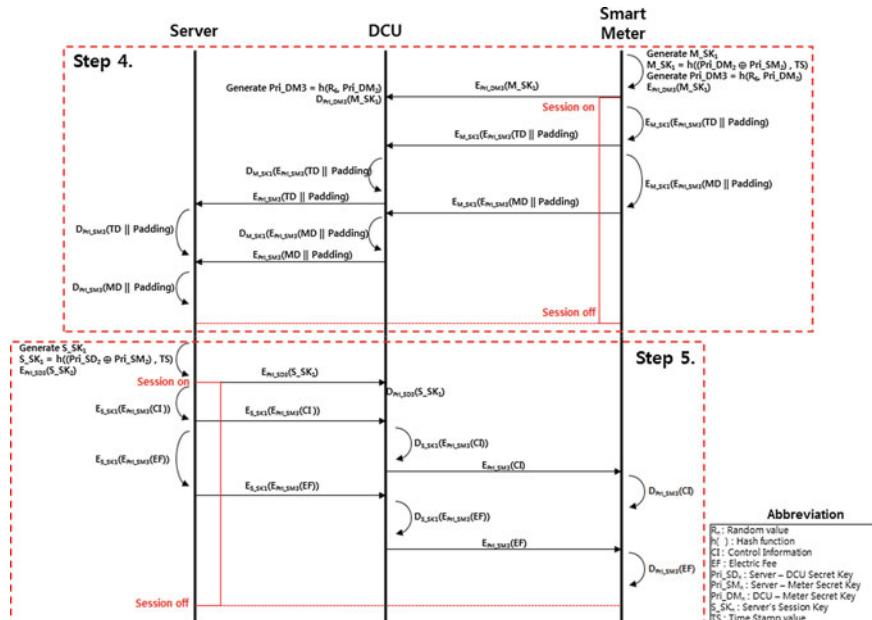


Fig. 2 Session generation and data transmission protocol

- To decrypt the encryption, the DCU receiving $E_{Pri_DM3}(M_SK_1)$ generates Pri_DM_3 to decrypt the encryption received.
- To safely transmit MD, Smart Meter connects the same padding value to each of the data by duplexing TD and MD, and after encrypting a private key between Server—Smart Meter, the encryption is once again encrypted with a session key to transmit the double encryption.
- For two encryptions of TD and MD receiving the encryption, the DCU decrypts the first encryption that can be decrypted, and transmits the second encryption to the server.
- For the encryption of double-transmitted TD and MD, the server obtains TD and MD by using and decrypting a private key used in the process of mutual certification between Server—Smart Meter.

In Step 5 as what is conducted in Step 4, to transmit safe data in three nodes mutual-certified Server—DCU—Smart Meter, control information transmission also generates a session resulting from a time stamp value, and performs double ciphered by using a private key between session key and node. Furthermore, to safely transmit data, by connecting TD and MD to the same padding value, double ciphered transmission was conducted, and in a final server, a matching process of TD and MD is performed through the identical padding value.

3 Conclusion

As a smart grid is in a very incomplete state in terms of security such as certification and encryption in the existing system, stability and efficiency were analyzed by comparing EAP-TLS, EAP-TTLS, and EAP-MD5 on the standard of IEEE_802.1x applied in M2M, the upper concept of smart grid. In terms of efficiency and safety, it showed 2.6 times higher compared to certification methods of the standard of IEEE-802.1x, and lightweight was identified in certificate certification attributes compared to the existing methods.

The protocol set forth for the future will enhance efficiency by grouping and managing many nodes under smart grid environment where device nodes gradually increased, requiring research on more safe, flexible smart grid management system, and extensive research to apply it to any circumstances. Furthermore, counter-measures are required to cope with many different security threats not found yet.

References

1. IEC. IEC62056-62 International Standard, electricity metering—data exchange for meter reading, tariff and load control—part 62: interface classes (2006)

2. IEEE. IEEE Guide for smart grid interoperability of energy technology and information technology operation with the electric power system (EPS), end-use applications, and loads (2011)
3. ITU-T. Smart grid—deliverable on smart grid architecture (2011)
4. Korea Smart Grid Association. Smart grid demonstration only security guidelines (2010)
5. Hwang, M.Y.: A study on SmartGrid system with improved security. A Master's thesis (2011)
6. NIST. DRAFT NISTIR 7628 Smart grid cyber security strategy and requirements Feb (2010)
7. NIST. NIST Framework and roadmap for smart grid interoperability standards, Release 2.0 (2012)
8. Ho, R.J., Seo-Dong Il, Youl, Y.H.: EAP using split password-based authenticated key agreement protocol for IEEE Std 802.1x User Authentication. Journal of Internet Computing and Services (JICS) (2005)
9. Eun, S.K.: Mutual authentication and group key management mechanism for implementation of secure smart grid environments. A Master's thesis (2011)
10. Choi, W.K., Kim, S.H., Kim, B.S., Jun, M.S.: Korea Institute of Communications and Information Sciences. Research on strengthening safety of pattern data collected in Smart-grid AMI Environment through dual method and ISBox (2013)
11. Hwang, M.Y.: A study on SmartGrid system with improved security. A Master's thesis (2011)

A Study on Realtime Detecting Smishing on Cloud Computing Environments

Ayoung Lee, Kyounghun Kim, Heeman Lee and Moonseog Jun

Abstract Learning By smartphone developed, mobile malicious code will produce many new species of fraud technique that Smishing using a mobile malicious code was born. Damage of Smishing Because of the increase, against this possibility is urgent. In this paper, filtering, API analysis, bypass detection, in order to capture the sewing machine's detection is performed. The proposed test method is performed directly in a virtualized environment, so you can also detect unknown malware, and that's sewing machine can reduce the damage caused.

Keywords Cloud computing · Smishing · Smart phone · Malware detection

1 Introduction

The number of users and the usage rate of Smartphone, terminal device that supports Internet communication and information search in a mobile phone. As more people use Smartphone, mobile malicious also increases, especially Smishing malicious code. Smishing is a coined word with SMS (Short Message Service) and (Phishing). It is to steal personal information through text message. Malicious code or malicious application used for Smishing can intercept authentication numbers and make micro payment illegally. Furthermore, it misuses the address list saved in

A. Lee (✉) · H. Lee · M. Jun
Soongsil University, Dongjak-gu, South Korea
e-mail: layinc@daum.net; layinc@ssu.ac.kr

H. Lee
e-mail: heeman930@ssu.ac.kr

M. Jun
e-mail: mjun@ssu.ac.kr

K. Kim
Department of Computer Information, Gangdong University, Gangdong, South Korea
e-mail: iioii.net@gmail.com

a user's phone to others to attempt Smishing and causes harms. In addition, Smishing can steal personal information such as security card, ID photo or accredited certificate, which leads to more serious crimes. As the damages by Smishing have escalated, a number of Smishing blocking plans also came out. However, Smishing technology keeps developing and thus a new countermeasure become necessary [1–8].

In this respect, the present study is aimed to propose a Smishing detection technique by using Cloud virtual environment. Existing Smishing detection methods focus on knowing whether it requires downloading apk file in a suspicious URL or it is an application with no sound source or not. In the technique proposed here, moreover, we try to increase the probability of Smishing detection by using filtering, PI analysis and bypass detection in order to minimize false and incomplete detection.

2 Proposal

2.1 Real-Time Smishing Detection Technique

SMS/MMS is transmitted to a user's Smartphone. The transmitted SMS/MMS detects URL link and confirms the source of an application, location of server and contents. First, when it detects a sourceless application, it sends information to Cloud virtual environment. And then, it confirms the location of the server. It podcasts the server address of C&C (Computer and Communication) and check IP Address to confirm if the server is located (local or overseas). When it is determined that the C&C server is located abroad, it blocks linking with the server. Last, the proposed system confirms information contents. When the contents are confirmed to be apk file, it also transmits them to the virtual environment. Then, Smartphone running under the virtual environment implements verification processes such as filtering, API analysis, bypass detection and screen capture to detect Smishing. The results from the verification processing become the basis of a suspicion report and the report is transmitted to a user inform him or her of the possibility of Smishing.

2.2 Real-Time Smishing Detection Process

The proposed real-time Smishing detection process is summarized in Fig. 1.

1. Filtering

Filtering is a process in which data or rules are made to compare and screen. Filtering process is divided again into application filtering, authorization filtering, port filtering and information filtering. Figure 2 shows the performance procedure of filtering.

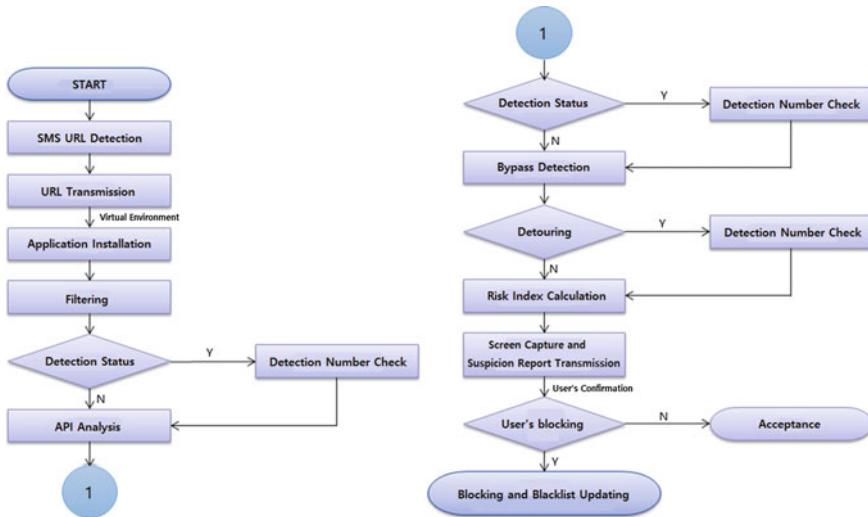


Fig. 1 Real-time Smishing detection process

- Application Filtering: It check an existing blacklist to know if a concerned application is included in the list. If it is, it checks detection status necessary to preparing for a suspicion report.
- Authorization Filtering: It confirms the authorization of the application and performs filtering work. A user can confirm what requirements the application asks for before installing the application. The process checks the authorizations mainly used for a malicious application and compares them with those of the application to be installed in URL transmitted from SMS. Table 1 shows the authorizations often used by a malicious application.
- Port Filtering: It checks the ports often under attack and determines if it is a malicious application. Port 7777 is one often used for Trojan and backdoor. So it shuts down the port in advance.
- Information Filtering: It checks and detects information in transmission. When a mobile device is in communication, the process checks transmitting packet and confirms information scripts through the packets. It is the process to check if a packet contains important personal information or SMS by checking packets such as Wireshark.

2. API Analysis

In the process of API analysis, an application is de-compiled; its java source codes are converted; and they are compared with API blacklist. First, apk file is converted into zip file and then dex file. And then dex file is converted into java file to check API. API on the application is compared with API blacklist to determine whether it is a malicious application. Figure 3 shows the performance procedure of API analysis.

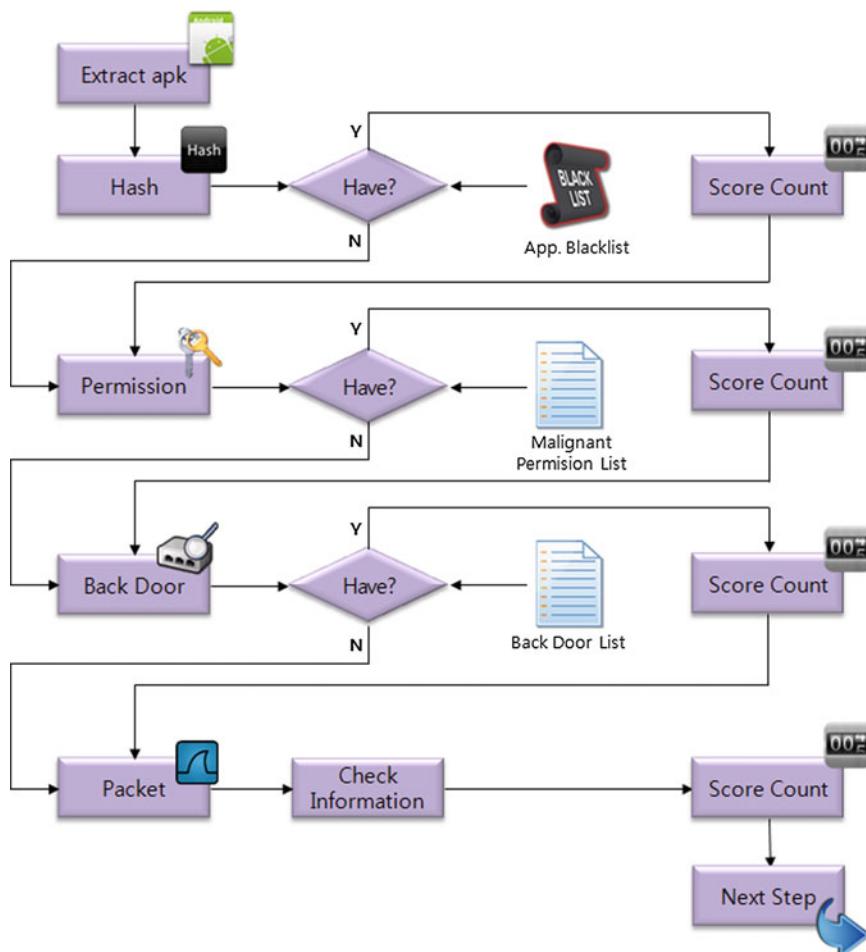


Fig. 2 The performance procedure of filtering

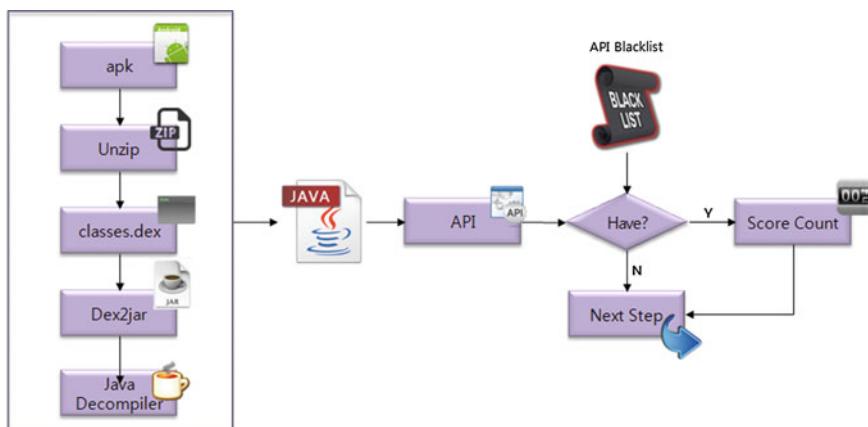
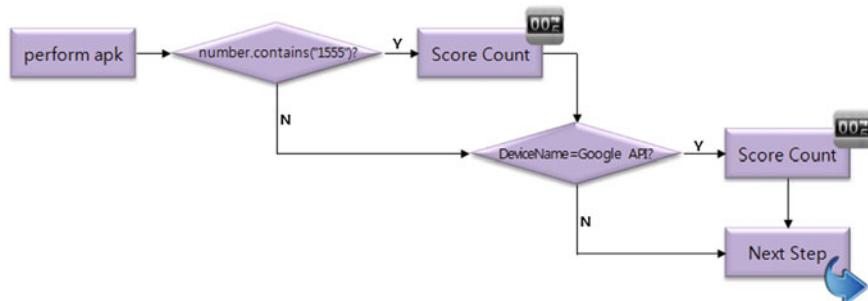
3. Bypass Detection

Bypass detection is to check if an application makes detour such as running emulator instead of starting in the Smartphone when it is started. Figure 4 shows the performance procedure of bypass detection.

If an application calls for emulator, it is determined as suspicious behavior and as making detour. in most cases, telephone number of emulator begins with 1555. Under Android system, there is TelephonyManager.getLine1Number function among basic API. A terminal returns its phone number by using the function. Therefore, it is checked if there exists 1555 by using the java.contains function.

Table 1 Examples of the authorizations often used by a malicious application

Authorization	Explanation
ACCESS_COARSE_LOCATION	Identifying location based on data network suitability
ACCESS_FINE_LOCATION	Identifying location based on WIFI
READ_PHONE_STATE	Reading the status of a terminal
READ_SMS	Reading SMS
SEND_SMS	Transmitting SMS
READ_CONTACTS	Retrieving address book
RECEIVE_BOOT_COMPLETED	Implementing service after completing booting
RECEIVE_SMS	Receiving SMS
WRITE_SETTINGS	Changing environment setting
RECORD_AUDIO	Receiving audio

**Fig. 3** The performance procedure of API analysis**Fig. 4** The performance procedure of bypass detection

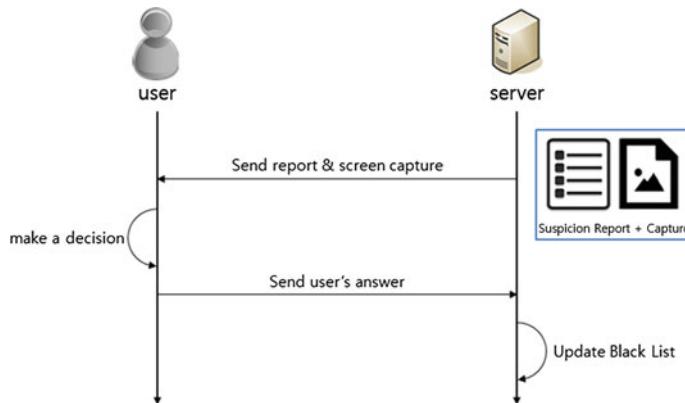


Fig. 5 Process of user's confirmation

4. Screen Capture and Transmission of suspicion Report

In the process of screen capture and suspicion report transmission, a tool is used to test and captured screen and suspicion report are transmitted to a user. Android-tool-monkey is used to run an application and capture screens. In addition, a suspicion report is prepared on the basis of the results from the previous processes in numerical rating and the captured screens and the report are sent to a user (Fig. 5).

When a suspicion report and captured screen(s) are sent to a user, the user determines the application contains Smishing based on the information given. And then the user sends feedback to the server and it is updated in the blacklist.

3 Conclusion

The present study examined a real-time Smishing detection technique using Cloud environment. When a user receives text message, he or she can judge its risk status in real time under virtual environment. To determine that it is Smishing, a user can conduct verification test in his own Smartphone because it can be damaged. Therefore, the present study proposed the test in Cloud environment. Since verification test is carried in Cloud environment, the user can be safe from possible risk and damage. And when the test completes, the system does not determine Smishing and block it, but transmits captured screen and suspicion report to the user by using TSI. Then it is the user that determines the status of Smishing. Since a user decides whether it is Smishing or not with captured screen and suspicion report, it can reduce false and incomplete detection. Currently, commercialized applications or existing studies focus on the detection of malicious code with a blacklist. Therefore, it is undetectable if unknown malicious code is included. However, the detection

technique proposed in this study actually runs a test in virtual environment and determines malicious behaviors by dynamic analysis. Therefore, it can detect unknown malicious code in advance and prevent damages by malicious code from spreading. This research aims to make a framework for learning system on cloud computing environment.

References

1. Amamra, A., Talhi, C., Robert, J.-M., Hamiche, M.: Impact of dataset representation on smartphone malware detection performance. *IFIP Adv. Inf. Commun. Technol.* **401**, 166 (2013)
2. Amamra, A., Talhi, C., Robert, J.-M., Hamiche, M.: Enhancing smartphone malware detection performance by applying machine learning hybrid classifiers. *Commun. Comput. Inf. Sci.* **340**, 131 (2012)
3. Hazarika, B., Aghakhani, N., Mannino, M.: Understanding the concept of deception in mobile commerce: an empirical examination of Smishing in mobile banking. In: *Proceedings of the Americas Conference on Informat*, vol. 20, no. 4 (2014)
4. Joo, C.K., Yoon, J.W.: Discrimination of spam and prevention of Smishing by sending personally identified sms (for financial sector). *Korea Inst. Inf. Secur. Cryptology*, vol. 24, no. 4 (2014)
5. Kang, A., Lee, J.D., Kang, W.M., Barolli, L., Park, J.H.: Security Considerations for Smart Phone Smishing Attacks. *Lecture Notes in Electrical Engineering*, vol. 279, no. 1 (2014)
6. Lee, S.Y., Kang, H.S., Moon, J.S.: A study on smishing block of android platform environment. *Korea Inst. Inf. Secur. Cryptology*. vol. 24, no. 5 (2014)
7. Mylonas, A., Dritsas, S., Tsoumas, B., Gritzalis, D.: On the feasibility of malware attacks in smartphone platforms. *Commun. Comput. Inf. Sci.* **314**, 217 (2012)
8. Park, D.W., Seo, J.M.: A study of information leakage prevention through certified authentication in phishing, vishing, smishing attacks. *J. Korea Soc. Comput. Inf.* **12**(2), 171 (2007)

Application of Data Mining for Crime Analysis

**Aziz Nasridinov, Jeong-Yong Byun, Namkyoung Um
and HyunSoon Shin**

Abstract Data mining can be considered as a powerful tool that enables law enforcement and emergency agencies to discovery meaningful patterns in a large amount of danger situation. In this paper, we propose a data mining framework for predicting crimes. The proposed framework consists of the following modules: test data generation, classification, clustering and data ranking. We performed the experiments using various datasets in order to determine which one is potentially best for performing crime pattern prediction task.

Keywords Data mining framework · Danger situation dataset · Performance evaluation

1 Introduction

Even though law enforcement and emergency agencies put major efforts to stop crime situations, the challenge of dealing with a large amount of data about danger situations has become a problem. As this data grows at an accelerating pace, it is difficult to manually recognize danger patterns, analyze and predict them. Thus,

A. Nasridinov · J.-Y. Byun

School of Computer Engineering, Dongguk University, 123 Dongdaero, Gyeongju,

Gyeongbuk 780-714, South Korea

e-mail: aziz@dongguk.ac.kr

J.-Y. Byun

e-mail: byunyj@dongguk.ac.kr

N. Um · H. Shin (✉)

Things and Emotion Convergence Research Team, ETRI, 218 Gajengno, Yuseong-gu,

Daejeon 305-700, South Korea

e-mail: hsshin@etri.re.kr

N. Um

e-mail: nkum@etri.re.kr

data mining can be considered as a powerful tool that enables law enforcement and emergency agencies to discover meaningful patterns in a large amount of danger situation dataset using various pattern recognition, mathematical and statistical methods [1, 2].

In this paper, we propose a data mining framework for predicting crimes. The proposed framework consists of the following modules: test data generation, classification, clustering and data ranking. In test data generation module, we used Bayes' Theorem so that it comprises with well-known data mining pattern prediction criteria. In classification module, we used k-nearest neighbor (KNN) algorithm for crime prediction. In clustering module, we used k-means algorithm for grouping the crimes into clusters. In data ranking module, we used skyline query method that outputs the data points that are not dominated by other data points. We performed the experiments using various datasets in order to determine which one is potentially best for performing crime pattern prediction task.

The rest of the paper is organized as follows. In Sect. 2, we discuss the literature dedicated to analyze the danger data using data mining techniques. In Sect. 3, we briefly present steps of the proposed method. In Sect. 4, we describe conclusion and future work.

2 Data Mining Framework

The proposed data mining framework consists of the following modules:

Test Data Generation: We use a test data generation method, which carefully designs test dataset so that it comprises with well-known data mining algorithms. Specifically, in this research, we focus on generating danger situations dataset that can be used as input data to the data mining. We used Bayes' theorem to generate danger test data, which describes the relationships that exist within a set of conditional probabilities.

Classification Module: In classification module, we used k-nearest neighbor (KNN) algorithm for crime prediction. KNN [3] is a simple algorithm that classifies new observations based on a similarity measure. More specifically, KNN algorithm contains four steps. The first step determines the k nearest neighbors, and the second step calculates the distance between those neighbors with other data points. Distance functions are used to calculate the distance between two data points. The third step of KNN algorithm defines nearest neighbors based on the k -th minimum distance. The fourth part uses a majority of the category of nearest neighbor to classify the test data.

Clustering Module: In clustering module, we used k-means algorithm for grouping the crimes into clusters. K-Means algorithm is partitioning clustering approach, where each cluster has the center point (i.e. centroid). The goal of k-means algorithm is to cluster n data points into k groups, so that distance between data points in a cluster and center point is minimal [4].

Data Ranking Module: In data ranking module, we used skyline query method that outputs the data points that are not dominated by other data points. A data point dominates another point if it is as good or better in all dimensions and better in at least one dimension.

Figure 1 demonstrates the GUI (Graphic User Interface) of the proposed data mining framework. The GUI has four parts that enables user to choose file, generate test data, select classification/clustering/ranking options, and observe the output. We describe each option in details. First, the user has an option to select dataset from hard disk. On the other hand, he can choose to generate test data. Test data generation enables to choose type of data such as correlated, anti-correlated and uniform dataset. Further, user can specify the size of data and test data. In classification options, the user can choose the number of k , where larger value of k , the higher accuracy is for KNN algorithm. In clustering options, the user can choose the number of clusters k . In both classification and clustering modules, the user can choose distance function for calculating the similarity of neighbors with test data. In

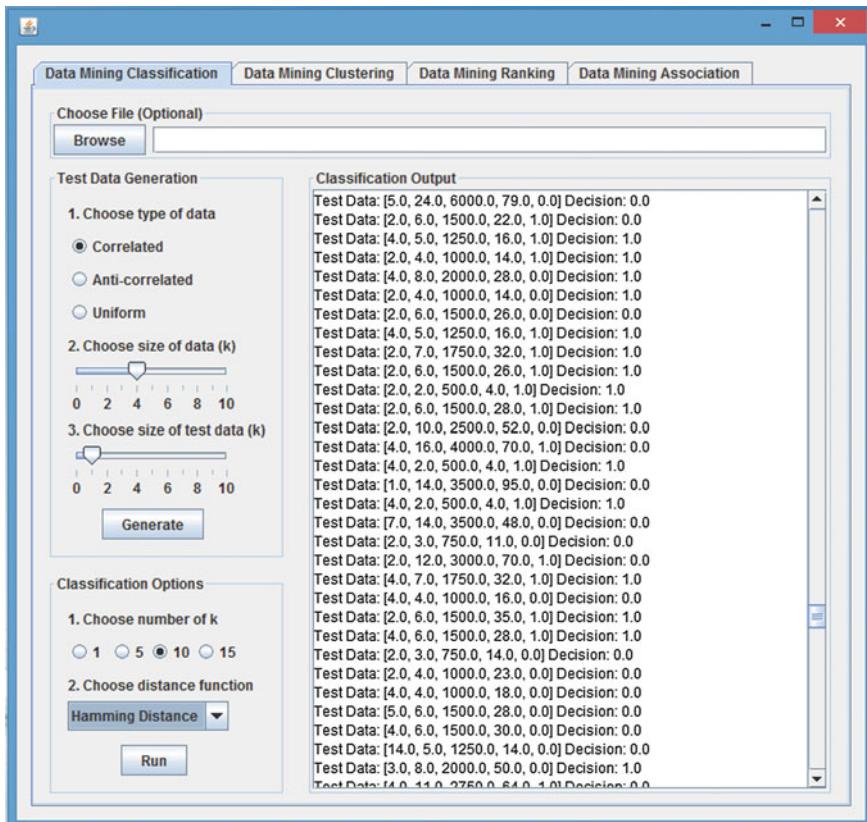


Fig. 1 Data mining framework for crime prediction

data ranking options, the user can choose the number of skyline layers. The higher number of skyline layers, the more ranked data user can see. Finally, output shows how each danger case is classified/clustering/ranked according to the algorithm.

3 Performance Evaluation

In this section, we analyze a variety of data mining techniques using the generated test data. By this, we determine which is best for performing crime pattern prediction task.

Experiment 1: Comparison of the computational time of three methods on correlated dataset. Figure 2 shows the comparison of the computational time of KNN, k-means and Skyline algorithms. In Fig. 4, x axis and y axis are dataset size and the computational time (ms), respectively. Among experimented methods, we can observe that skyline and KNN uses less time (between 2 and 50 ms). This is because, skyline simply compares each element with other elements, and KNN calculates distance between test data point with other data points. For this, skyline and KNN algorithms consume less time. Also, in correlated dataset, skyline includes less data points, which means skyline algorithm performs less comparisons. However, k-means consumes much time comparing skyline and KNN (4–12 times), because k-means algorithm performs several iterations in order to determine correct clusters. In each iteration it performs distance calculation, and thus, consumes more time than other algorithms.

Experiment 2: Comparison of the computational time of three methods on anti-correlated dataset. Figure 3 shows the comparison of the computational time of KNN, k-means and Skyline algorithms. In Fig. 3, x axis and y axis are dataset size and the computational time (ms), respectively. The graph shows that KNN algorithm does not change its performance comparing to previous experiment (Experiment 4). This is because KNN algorithm always computes distance between test data point and other data points. Here, data distribution does not influence its performance. Comparing to previous experiment (Experiment 4), k-means shows

Fig. 2 Comparison of time for correlated dataset

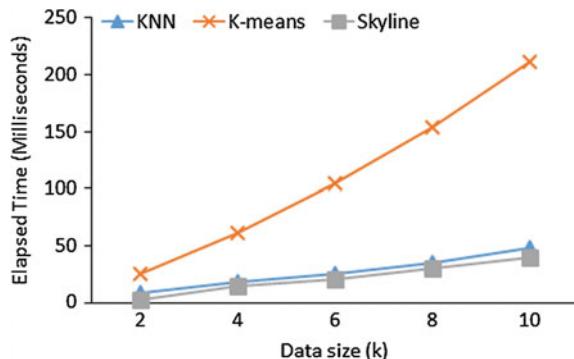
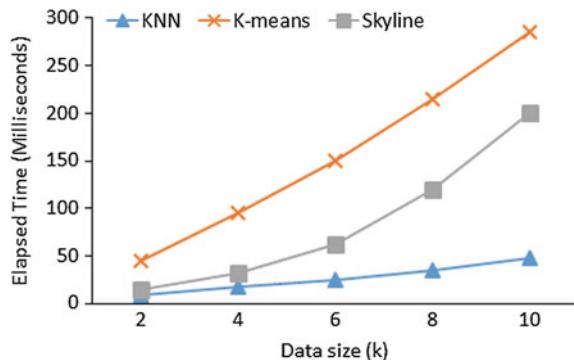


Fig. 3 Comparison of time for anti-correlated dataset



identical results. K-means uses much time as number of elements in dataset increases. Similarly, skyline algorithm consumes more time as data size increase. This happens because distribution of data points on anti-correlated dataset are similar to distribution of skyline points. Thus, skyline includes more points, and performs more comparisons.

Experiment 3: Comparison of the computational time of three methods on uniform dataset. Figure 4 shows the comparison of the computational time of KNN, k-means and Skyline algorithms. In Fig. 4, x axis and y axis hold dataset size and the computational time (ms), respectively. Similar to Experiment 1 and Experiment 2, the performance of KNN does not change when using uniform dataset. Time consuming of k-means is still high, because it performs numerous iterations to determine clusters. Time consuming of skyline lower than previous experiment (Experiment 1). This is because, in uniform dataset, data points are randomly scattered. Thus, when skyline is build, it may include less data points.

Experiment 4: Comparison of the computational memory usage of three methods on correlated, anti-correlated and uniform datasets. Figure 5 shows the comparison of memory usage of KNN, k-means and skyline algorithms. In Fig. 5, x axis and y axis are dataset size and memory usage, respectively. The graph presents memory usage values of each algorithm. Among these methods, skyline consumes

Fig. 4 Comparison of time for uniform dataset

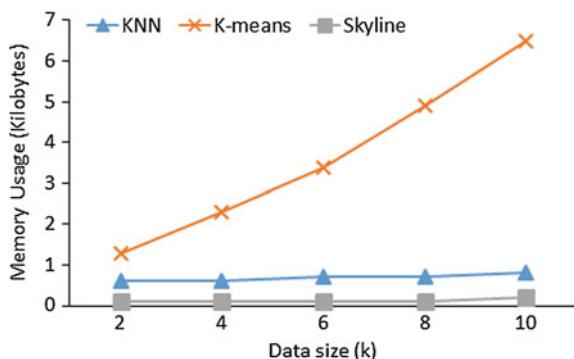
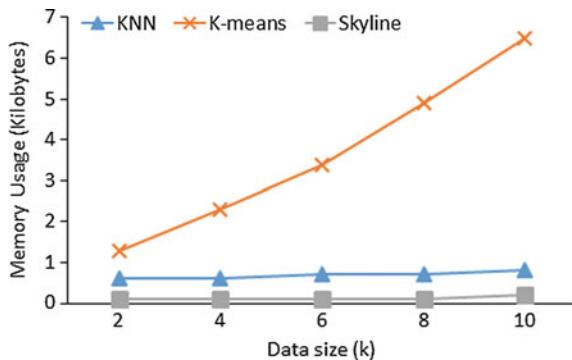


Fig. 5 Comparison of memory usage for correlated, anti-correlated and uniform dataset



the less memory. Because, for building skyline, memory maintains only a single dataset (candidate dataset). On the other hand, KNN uses more memory than skyline. Because memory maintains two or more decision sets at the same time. k-means uses more memory than skyline and KNN because in order to build k-means memory always needs to maintain and write cluster sets.

4 Conclusion

In this paper, we describe the data mining framework. We briefly demonstrated that methods used in this project can help predict crimes. First, test data generation can replace real data set. And also, it can be used for testing data mining techniques. Third, data mining techniques, which we described in this paper, can be effectively used for crime prediction. Finally, our experiment result shows that among KNN, k-means and Skyline algorithms, it is important to use according to dataset.

References

1. De Bruin, J.S., Cox, T.K., Kosters, W.A., Laros, J.F.J., Kok, J.N.: Data mining approaches to criminal career analysis. In: Proceedings of the Sixth International Conference on Data Mining (ICDM), pp. 171–177 (2006)
2. Nasridinov, A., Ihm, S.Y., Park, Y.H.: A decision tree based classification model for crime prediction. In: Proceedings of the 10th International Conference on Secure and Trust Computing (STA), Data Management, and Applications, pp. 531–538 (2013)
3. Cover, T.M., Hart, P.E.: Nearest neighbor pattern classification. *IEEE Trans. Inf. Theor.* **13**(1), 21–27 (1967)
4. Kumar, Y., Sahoo, G.: A new initialization method to originate initial cluster centers for K-Means algorithm. *Int. J. Adv. Sci. Technol.* **62**, 43–54 (2014)

Graph-Based Motor Primitive Generation Method of UAVs Based on Demonstration-Based Learning

Yunsick Sung, Jeonghoon Kwak and Jonghyuk Park

Abstract Motor primitive generation approach is essential technique to improve the quality of UAVs. In this paper, demonstration-based learning is applied to generate motor primitives. By collecting and merging command signals of UAVs, one graph-based motor primitive was generated.

Keywords UAV · AR.Drone 2.0 · Motor primitive · Demonstration learning

1 Introduction

Recently, there are diverse kinds of research of UAVs. One of UAV domains is the surveillance of buildings, cars, and people. Given that motor primitives determine which UAVs can do, the way to define motor primitives is core technique. Motor primitives can be generated by demonstrations-based learning, which has the advantages of learning speed and learning cost. To apply demonstrations-based learning to UAVs, algorithms that consider the unique features of UAVs are required.

This paper proposes a motor primitive generation method for UAVs by applying demonstration-based learning. In experiments, the proposed method was applied for

Y. Sung (✉)

Computer Engineering Division, Keimyung University, Daegu, South Korea
e-mail: yunsick@kmu.ac.kr

J. Kwak

Department of Computer Engineering, Keimyung University, Daegu, South Korea
e-mail: jeonghoon@kmu.ac.kr

J. Park

Department of Computer Engineering, Seoul National University of Science and Technology,
Seoul, South Korea
e-mail: jhpark1@seoultech.ac.kr

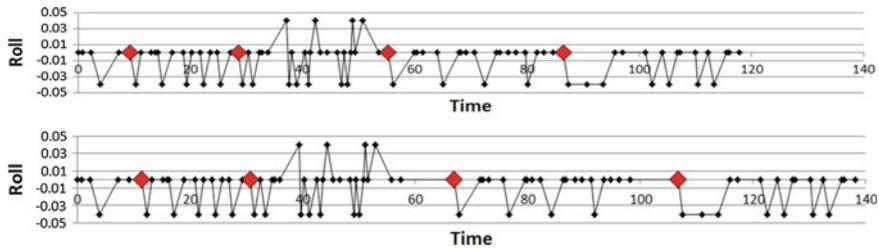


Fig. 1 Red mark means the time when UAVs are in front of any pin points. *Upper figure* shows the collected rolls with the corresponding time and *lower figure* shows the adjusted time of rolls the parking cars' surveillance system by UAVs. By merging collected control data, one graph-based motor primitive is generated.

2 The Generation Process of Motor Primitives

The process of generating motor primitives is divided into three stages. During operation collection stage, the values of pitches and rolls are collected. In addition, the times when UAVs face to an intermediate pin point to adjust the locations of UAVs are obtained. During time adjustment stage, the time of the collected pitches and rolls are adjusted. One learning section is defined by the moved path from any one intermediate pin point to one of other intermediate pin points. Therefore, all intermediate pin points' times at the same learning section are set to the maximum of those times. Motor primitive generation stage merges all motor primitives into one graph-based motor primitive by connecting the pin point time of each learning section in order. Figure 1 shows the collected and adjusted rolls.

Any part of motor primitives that will start at the time when a previously executed part of a motor primitive is finished can be executed. As result, motor primitives can be executed changing its form diversely.

This paper proposed an approach to generate graph-based motor primitives. By utilizing the pin point time of motor primitives, the motor primitives were integrated into one graph-based motor primitive. In the future, the approach to execute generated motor primitives is going to be researched.

Acknowledgments This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT and Future Planning (NRF-2014R1A1A1005955).

Author Index

A

Ahn, Kyeongrim, 269
Al-Absi, Ahmed Abdulhakim, 9
AlShemeili, Ahmed, 259
Anze, Shogo, 25
Arakawa, Yutaka, 431
Ashar, Muhammad, 431
Aung, Khin M.M., 405

B

Bae, Ihn-Han, 59
Baek, Joonsang, 259
Byun, Jeong-Yong, 503

C

Chang, Chun-Hao, 389
Chen, Jin-Yi, 389
Cheng, Jingde, 25
Cho, Chae Ho, 1
Cho, Hanbyeog, 153
Cho, Sungkap, 357
Cho, Wanhyun, 279
Cho, Young-Hwa, 161, 169, 177, 185, 193, 203
Choi, Byeong Cheol, 349
Choi, Gyuyeun, 95
Choi, Jae-Young, 161, 169, 177
Choi, Jin-Kyu, 153
Choi, Wonkyu, 487
Chung, Ji-Hwan, 253
Chung, Kwang Sik, 1
Chung, Tae-Sun, 253
Cui, Xingmin, 145

D

Dai, Yongchuan, 301

E

Elhadef, Mourad, 17

G

Go, Byung Gill, 367
Goto, Yuichi, 25
Gu, Mi Sug, 107

H

Han, Seonmi, 1
Hao, Fei, 233
Harada, Tsubasa, 25
Heo, Daeyoung, 95
Hui, Lucas C.K., 145
Hustak, Tomas, 51
Hwang, Jeong Hee, 107
Hwang, Sang-Won, 225
Hwang, Suntae, 95
Hwang, Young-Ho, 317, 357

I

Im, Jong-Hyuk, 71

J

Jang, Byungtae, 137
Jang, Won Seuk, 349
Jeon, Sangduck, 219
Jeon, Sung-Yun, 71
Jeon, Yong-Hee, 115
Jeong, Changhyun, 123
Jeong, Jong Seob, 349
Jeong, Young-Sik, 233
Ji, Seon Mi, 349
Jin, Chao, 415
Jin, Qun, 443
Jo, Changmin, 79
Jo, Young-Hoo, 71
Jun, Moon-seog, 457, 473, 479, 487
Jun, Moonseog, 465, 495
Jung, Daeyong, 41
Jung, Dohyun, 123
Jung, Seung-Won, 333

Jung, Wonjin, 383

K

Kang, Dae-Ki, 9

Kang, Jungho, 451

Kasuya, Seiji, 443

Khan, Adil Mehmood, 367

Khattak, Asad Masood, 367

Kim, Byeongwoo, 123, 211, 219, 241, 247

Kim, Changhee, 333

Kim, Eunhwan, 473

Kim, Gyoungeun, 211

Kim, Hyeon Gyu, 375, 397

Kim, Hyungjoo, 457

Kim, HyunYoung, 203

Kim, Jae-Soo, 225, 429

Kim, Jaewoo, 465

Kim, Jeong Ah, 161, 203

Kim, Jinyong, 123

Kim, JongBae, 325

Kim, Jung-Ah, 169, 177

Kim, Kyong-Ho, 153

Kim, Kyounghun, 495

Kim, ManSik, 479

Kim, Myong-Jong, 9

Kim, Sangkyoon, 279

Kim, Sun-Tae, 161, 169, 177

Kim, Sung Min, 349

Kim, Sungsuk, 1

Kim, Suntae, 185, 193, 203

Ko, Jae Jin, 341

Ko, Jong-Won, 161, 169, 177

Kreesuradej, Worapoj, 87

Krejcar, Ondrej, 33, 51

Kwak, Jeonghoon, 509

Kwon, Lee-Nam, 225

L

Lee, Ayoung, 495

Lee, Byoung-Dai, 225

Lee, Daewon, 41

Lee, Deok Gyu, 211, 219

Lee, Heeman, 495

Lee, HwaMin, 41

Lee, Illo, 185

Lee, Im-Yeong, 285

Lee, Jae-Kon, 59

Lee, Jae Kyu, 341

Lee, Jaeseung, 473

Lee, Jaesik, 487

Lee, Kwangil, 129, 137

Lee, Malrey, 193

Lee, Mun-Kyu, 71

Lee, Sang Yub, 341

Lee, Younggu, 465

Lee, Younggyo, 269

Lim, Sangwon, 269

Lin, Ming, 241

Lin, Yih-Kai, 389

M

Moon, Yoo-Jin, 317, 357

Mun, Hyeonggeun, 247

N

Nam, Young-Kwang, 225

Nasridinov, Aziz, 503

Nishimura, Shoji, 443

O

Oh, Hyun-Seo, 153

P

Park, Chan Yuk, 349

Park, Doo-Soon, 41, 233

Park, Duck Keun, 341

Park, Jae-pyo, 473

Park, Jong Hyuk, 233

Park, Jonghyuk, 509

Park, Jungoh, 487

Park, Soonyoung, 279

Park, Sooyong, 185

Park, Sung-Wook, 285

Park, Sung Yun, 349

Peng, Junjie, 301

Pumjun, Nophadon, 87

R

Rao, Yi, 301

Ro, Soonghwan, 137

Ryu, Heung-Gyo, 153

S

Sangkul, Konthorn, 423

Shah, Babar, 367

Shin, HyunSoon, 503

Smanchat, Sucha, 423

Song, Junho, 479

Song, Moonsub, 129, 137

Sung, Jin Ho, 349

Sung, Yunsick, 509

T

Tamai, Morihiko, 431

Tham, Jo Yew, 423

Tsang, Yu Hin, 145

U

Um, Namkyoung, [503](#)

W

Wagatsuma, Kazunori, [25](#)

Won, Chee Sun, [333](#)

X

Xi, Weiyu, [415](#)

Xu, Quanqing, [405](#), [415](#)

Y

Yang, Cheng-Hsing, [389](#)

Yasumoto, Keiichi, [431](#)

Yeun, Chan Yeob, [259](#)

Yiu, S.M., [145](#)

Yong, Khai Leong, [405](#), [415](#)

Yoon, Jaewoo, [241](#)

Yun, Seokmin, [333](#)

Z

Zhi, Xiaofei, [301](#)

Zhou, Xiaokang, [443](#)

Zhu, Yongqing, [405](#)

Zmitko, Martin, [33](#)