Lecture Notes in Electrical Engineering 280

Young-Sik Jeong
Young-Ho Park
Ching-Hsien (Robert) Hsu
James J. (Jong Hyuk) Park *Editors*

Ubiquitous Information Technologies and Applications

CUTE 2013

Springer

Lecture Notes in Electrical Engineering

Volume 280

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

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.

Young-Sik Jeong · Young-Ho Park
Ching-Hsien (Robert) Hsu
James J. (Jong Hyuk) Park
Editors

Ubiquitous Information Technologies and Applications

CUTE 2013

Springer

Editors
Young-Sik Jeong
Department of Multimedia Engineering
Dongguk University
Seoul
Korea
Republic of (South Korea)

Young-Ho Park
Department of Multimedia Science
Sook Myung Women's University
Seoul
Korea
Republic of (South Korea)

Ching-Hsien (Robert) Hsu
Department of Computer Science and Information Engineering
Chung Hua University
Hsinchu
Taiwan

James J. (Jong Hyuk) Park
Department of Computer Science and Engineering
Seoul University of Science and Technology (SeoulTech)
Seoul
Korea
Republic of (South Korea)

ISSN 1876-1100 ISSN 1876-1119 (electronic)
ISBN 978-3-642-41670-5 ISBN 978-3-642-41671-2 (eBook)
DOI 10.1007/978-3-642-41671-2
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013952938

Printed on acid-free paper

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

Message from the CUTE 2013 General Chairs

On behalf of the organizing committees, it is our pleasure to welcome you to the 8th International Conference on Ubiquitous Information Technologies and Applications (CUTE 2013), will be held in Danang, Vietnam on December 18–20, 2013.

This conference provides an international forum for the presentation and showcase of recent advances on various aspects of ubiquitous computing. It will reflect the state-of-the-art of the computational methods, involving theory, algorithm, numerical simulation, error and uncertainty analysis and/or novel application of new processing techniques in engineering, science, and other disciplines related to ubiquitous computing.

The papers included in the proceedings cover the following topics: Ubiquitous Communication and Networking, Ubiquitous Software Technology, Ubiquitous Systems and Applications, Ubiquitous Security, Privacy and Trust. Accepted papers highlight new trends and challenges in the field of ubiquitous computing technologies. We hope you will find these results useful and inspiring for your future research.

We would like to express our sincere thanks to Steering Chairs: Young-Sik Jeong (Dongguk University, Korea), Mohammad S. Obaidat (Monmouth University, USA), Laurence T. Yang (St. Francis Xavier University, Canada), Hai Jin (Huangzhong University of Science and Technology, China), Chan-Hyun Youn (KAIST, Korea), Jianhua Ma (Hosei University, Japan), Minyi Guo (Shanghai Jiao Tong University, Japan), Bernady O. Apduhan (Kyushu Sangyo University, Japan), Weijia Jia (City University of Hong Kong, Hong Kong). We would also like to express our cordial thanks to the Program Committee members for their valuable efforts in the review process, which helped us to guarantee the highest quality of the selected papers for the conference.

Finally, we would 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.

Jong Hyuk Park, SeoulTech, Korea
Hamid R. Arabnia, The University of Georgia, USA
Cho-Li Wang, University of Hong Kong, Hong Kong

CUTE 2013 General Chairs

Message from the CUTE 2013 Program Chairs

Welcome to the 8th International Conference on Ubiquitous Information Technologies and Applications (CUTE 2013), will be held in Danang, Vietnam on December 18–20, 2013.

The purpose of the CUTE 2013 conference is to promote discussion and interaction among academics, researchers and professionals in the field of ubiquitous computing technologies. This year the value, breadth, and depth of the CUTE 2013 conference continues to strengthen and grow in importance for both the academic and industrial communities. This strength is evidenced this year by having the highest number of submissions made to the conference.

For CUTE 2013, we received a lot of paper submissions from various countries. Out of these, after a rigorous peer review process, we accepted only 107 high-quality papers for CUTE 2013 proceeding, 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 also 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:

- Hamid R. Arabnia, The University of Georgia, USA
- Leonard Barolli, Fukuoka Institute of Technology (FIT), Japan

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!

Young-Ho Park, Sookmyung Women's University, Korea
Robert C.H. Hsu, Chung Hua University, Taiwan
Julien Bourgeois, University of Franche-Comte, France

CUTE 2013 Program Chairs

Organization

Honorary Chair

Seok Cheon Park	Gachon University, Korea

Steering Chairs

Young-Sik Jeong	Dongguk University, Korea
Mohammad S. Obaidat	Monmouth University, USA
Laurence T. Yang	St.Francis Xavier University, Canada
Hai Jin	Huangzhong University of Science and Technology, China
Chan-Hyun Youn	KAIST, Korea
Jianhua Ma	Hosei University, Japan
Minyi Guo	Shanghai Jiao Tong University, Japan
Bernady O. Apduhan	Kyushu Sangyo University, Japan
Weijia Jia	City University of Hong Kong, Hong Kong

General Chairs

Jong Hyuk Park	SeoulTech, Korea
Hamid R. Arabnia	The University of Georgia, USA
Cho-Li Wang	University of Hong Kong, Hong Kong

Program Chairs

Young-Ho Park	Sookmyung Women's University, Korea
Robert C.H. Hsu	Chung Hua University, Taiwan
Julien Bourgeois	University of Franche-Comte, France

Program Vice-Chairs

Track 1. Ubiquitous Communication and Networking:

Soonil Kwon	Sejong University, Korea
Dominique Dhoutaut	UFC/FEMTO-ST Institute, France
Feng Xia	Dalian University of Technology, China

Track 2. Ubiquitous Software Technology:

Ritu Arora	University of Texas at Austin, USA
Xin Chen	University of Hawaii, USA
Kyungeun Cho	Dongguk University, Korea
Lu Liu	University of Derby, UK
John Grundy	Swinburne University of Technology, Australia

Track 3. Ubiquitous Systems and Applications:

Praveen Koduru	Kansas State University, USA
Young-Koo Lee	Kyung Hee University, Korea
Francoise Sailhan	CNAM, France
Zili Shao	The Hong Kong Polytechnic University, Hong Kong

Track 4. Ubiquitous Security, Privacy and Trust:

Seung-Ho Lim	Hankuk University of Foreigh Studies, Korea
Zheng Yan	XiDian Univ., China/Aalto Univ., Finland
Siani Pearson	HP Cloud and Security Research Lab, UK

Workshop Chairs

Min Choi	Chungbuk National University, Korea
Shingchern D. You	National Taipei University of Technology, Taiwan
Ming Li	California State University, USA

International Advisiory Committee

Doo-Soon Park	SoonChunHyang University, Korea
Sung-Kap Cho	Incheon IT Promotion Agency, Korea
Min Hong	SoonChunHyang University, Korea
Bin Xiao	The Hong Kong Polytechnic University, Hong Kong
Han-Chieh Chao	National Ilan University, Taiwan
HeonChang Yu	Korea University, Korea
Hung-Chang Hsiao	National Cheng Kung University, Taiwan
Javier Lopez	University of Malaga, Spain
Jeong-Bae Lee	Sunmmon University, Korea

Jiannong Cao	The Hong Kong Polytechnic University, Hong Kong
Jin Kwak	SoonChunHyang University, Korea
Nammee Moon	Hoseo University, Korea
Ved Kafle	NICT, Japan
Sang-Soo Yeo	Mokwon University, Korea
Kuan-Ching Li	Providence University, Taiwan
Im-Yeong Lee	SoonChunHyang University, Korea
Youn-Hee Han	KoreaTech, Korea
Yi Pan	Georgia State University, USA
Eunmi Choi	Kookmin University, Korea

Publicity Chairs

Joon-Min Gil	Catholic University of Daegu, Korea
Parimala Thulasiraman	University of Manitoba, Canada
Chao-Tung Yang	Tunghai University, Taiwan
Eunyoung Lee	Dongduk Women's University, Korea
Deqing Zou	HUST, China
Sanghoon Kim	Hankyong National University, Korea

Publication Chair

Hwa Young Jeong	Kyung Hee University, Korea

Invited Speaker

Hamid R. Arabnia	The University of Georgia, USA
Leonard Barolli	Fukuoka Institute of Technology (FIT), Japan

Program Committee

Ali Shahrabi	Glasgow Caledonian University, UK
Aziz Nasridinov	Sookmyung Womens University, Korea
Basel Alawieh	Alcatel-Lucent, Canada
Bhekisipho Twala	University of Johannesburg, South Africa
Bo-Chao Cheng	National Chung-Cheng University, Taiwan
Byung-Gyu Kim	SunMoon University, Korea
Byung-Kwon Park	Dong A University, Korea
Changhoon Lee	SeoulTech, Korea
Chen Liu	Florida International University, USA
Chulyun Kim	Gachon University, Korea
Damien Sauveron	University of Limoges, France
Deqing Zou	Huazhong University of Science & Technology, China
Der-Jiunn Deng	National Changhua University of Education, Taiwan

Dong Seong Kim	University of Canterbury, New Zealand
Dugki Min	Konkuk University, Korea
Dumitru Roman	SINTEF, Norway
Eunmi Choi	Kookmin University, Korea
Eun-Young Ahn	Hanbat National University, Korea
Fuu-Cheng (Admor) Jiang	Tunghai University, Taiwan
Giuseppe De Pietro	ICAR-CNR, Italy
Hae-Yeoun Lee	Kumoh National Institute of Technology, Korea
Heonchang Yu	Korea University, Korea
Hongli Luo	Indiana University, USA
Hwa Jin Park	Sookmyung Womens University, Korea
HwaMin Lee	Soonchunhyang University, Korea
Hyo-Sang Lim	Yonsei University, Korea
Imad Saleh	University of Paris 8, France
Jaewoo Kang	Korea University, Korea
Jeonghun Cho	Kyungpook National University, Korea
Jeong-Hyon Hwang	State University of NewYork at Albany, USA
Jeong-Yong Byun	Dongguk University, Korea
Jin Gon Shon	Korea National Open University, Korea
Jin-Hee Cho	U.S. Army Research Laboratory, USA
Jong-Myon Kim	University of Ulsan, Korea
Joon-Ho Woo	Samsung SDS, Korea
Joonho Kwon	Pusan National University, Korea
Joon-Min Gil	Catholic University of Daegu, Korea
JungMin Kim	DaeJin University, Korea
Jun-Ki Min	Korea University of Technology and Education, Korea
Keun Ho Ryu	Chungbuk National University, Korea
Ki-Hoon Lee	Kwangwoon University, Korea
Ki Yong Lee	Sookmyung Womens University, Korea
Kiyoshi Nakabayashi	Chiba Institute of Technology, Japan
Kuan-Chu Lai	National Taichung University, Taiwan
Kuniaki Uehara	Kobe University, Japan
Kuo-Chan Huang	National Taichung University, Taiwan
Kwang Sik Chung	Korea National Open University, Korea
Kwangman Ko	Sangji University, Korea
Kyong-Ho Lee	Yonsei University, Korea
Kyungeun Cho	Dongguk University, Korea
Lam-for Kwok	City University of Hong Kong, Hong Kong
Mei-Ling Shyu	University of Miami, USA
Milan Markovic	Banca Intesa ad Beograd, Serbia
Min Choi	Chungbuk National University, Korea
Min Hong	Soonchunhyang University, Korea
Ming Li	California State University, USA

Minsoo Lee	Ewha Womens University, Korea
Mohamed Ally	Athabasca University, Canada
Neungsoo Park	Konkuk University, Korea
Omaima Bamasak	King Abdulaziz University, Saudi Arabia
Pinaki A Ghosh	Atmiya Institute of Technology and Science, India
Ping-Feng Pai	Nation Chi Nan University, Taiwan
Pyung-Soo Kim	Korea Polytechnic University, Korea
Q. Shi	Liverpool John Moores University, UK
Sanghoon Lee	Yonsei University, Korea
Seng W. Loke	La Trobe University, Australia
Serge Chaumette	University of Bordeaux 1, France
Seung-Ho Lim	Hankook University of Foreign Studies, Korea
Shanmugasundaram Hariharan	Pavendar Bharathidasan College of Engineering and Technology, India
Shu-Ching Chen	Florida International University, USA
Soo-Hyun Park	Kookmin University, Korea
Soo-Kyun Kim	PaiChai University, Korea
Soonil Kwon	Sejong University, Korea
Stefanos Gritzalis	University of the Aegean, Greece
Tae Kyung Kim	Seoul Theological University, Korea
Toshiyuki Kamada	Aichi University of Education, Japan
Tyngyeu Liang	National Kaohsiung University of Applied Science, Taiwan
Uei-Ren Chen	Hsiuping University of Science and Technology, Taiwan
Walter Lee	Tamkang University, Taiwan
Wanquan Liu	Curtin University, Australia
Wan-Sup Cho	Chungbuk National University, Korea
Wei Wei	Xi'an University of Technology, China
Wen-Chi Hou	Southern Illinois University, USA
Won Woo Ro	Yonsei University, Korea
Wookey Lee	Inha University, Korea
Woong-Kee Loh	Sungkyul University, Korea
Yang-Sae Moon	Kangwon National University, Korea
YangSun Lee	Seokyeong University, Korea
Yao-Chung Chang	National Taitung University, Taiwan
Yong Ik Yoon	Sookmyung Womens University, Korea
Yo-Sung Ho	Gwangju Institute of Science and Technology (GIST), Korea
Young-Kuk Kim	Chungnam National University, Korea
Young-Sik Jeong	Wonkwang University, Korea
Zhiwen Yu	Northwestern Polytechnical University, China

Contents

Design and Development of a Driving Condition Collector for Electric Vehicles*

Junghoon Lee[1], Gyung-Leen Park[1,**], Byung-Jun Lee[1], Jikwang Han[2], Joo Kyung Kang[3], Bongsoo Kim[4], and Jinhwan Kim[5]

[1] Dept. of Computer Science and Statistics, Jeju National University
[2] Jinwoo Soft Innovation, Inc.
[3] NextEZ, Inc.
[4] MatisInfo, Inc.,
Jeju-Do, Republic of Korea
[5] Hansung University, Seoul, Republic of Korea
{jhlee,glpark,eothsk}@jejunu.ac.kr, hmurdoc@jinwoosi.co.kr, ez@nextez.co.kr, kbs0416@gmail.com, kimjh@hansung.ac.kr

Abstract. For the sake of obtaining a battery consumption model of reasonable accuracy for electric vehicles, this paper designs and develops a driving condition collector which can achieve the state of charge (SoC) information by interfacing a battery management system. It adds location and time stamps as well as other driving-dependent sensors, creating a spatio-temporal stream. A window-based application, called a road data analysis mapper, manipulates the SoC streams and plots the change in sensor readings, such as slope, speed, temperature, and the like. The battery consumption statistics are collected using this device as a series of battery remaining values in Jeju city. To allow drivers to estimate the battery consumption between two points, actually a subsequence of the whole stream, the power consumption to each evenly spaced reference points is interpolated.

Keywords: electric vehicle, driving condition collector, road mapper, computational intelligence.

1 Introduction

Electric vehicles, or EVs in short, are one of the most important components in the future low-carbon society. They can potentially reduce air pollution and achieve better energy efficiency. Moreover, V2G-enabled EVs can store energy in their batteries and give back to the grid on necessary basis, providing intelligent peak shaving and frequency regulation capabilities [1]. In spite of such advantages, EVs

* This research was financially supported by the Ministry of Trade, Industry and Energy (MOTIE), Korea Institute for Advancement of Technology (KIAT) through the Inter-ER Cooperation Projects.
** Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*, Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_1,

are not yet ready to be widely deployed into our daily lives mainly due to the range anxiety problem. It takes tens of minutes to charge a single EV even with fast chargers, while an EV can drive just 100 *km* [2]. Therefore, computational intelligence must be integrated to EV systems for energy-efficient driving [3]. It can schedule the EV charging operations, estimate when charging request will take place [4], and find energy-efficient routes [5]. Besides, many other computer algorithms can contribute to the convenient EV services.

In the mean time, the most basic concern of drivers lies in whether the EV can reach a destination with current battery remaining, or interchangeably, SoC (State of Charge) [6]. If not, the next question is where to have their EVs charged. Here, the availability of charging facilities and charging load imposed on them must be considered. Those questions can be answered only if we know the battery consumption dynamics on the route that will be taken by the EV. It is different road by road, due to many factors such as curvature, slope, and the like. After all, it is necessary to collect the battery consumption statistics to build a consumption model of reasonable accuracy. In this regard, this paper designs and develops a DCC (Driving Condition Collector) device, which collects SoC along the route, adds time and location stamps, and exports to external devices for further analysis.

2 DCC Design

Basically, an EV connector module, EVC from now on, developed in the enterprise of Jeju Smart Grid Model City, scans the SoC values and provides it to external servers as shown in Figure 1. This product can work with an EV which opens its BMS (Battery Management System) interface, which generates its status to CAN bus according to a vender-specific message format. Two external connection types are supported to report the BMS information. The first one is the cellular network, or CMDA in the Republic of Korea, by which SMS-based status requests and reports are exchanged. The second one is the serial port connection via which SoC information is relayed to a new device. This device can acquire EVC data and combine its own sensor readings. The EVC works on 12V power and mainly remains in sleep mode. It wakes up when the new data is active over the CAN bus. It is the case when the EV power switch is turned on or the EV battery is being charged.

Our DCC has an interface to an EVC to obtain SoC information and embeds various sensors for temperature, humidity, slope, speed, and altitude. It has its own display unit to promptly provide current sensor readings to on-site operators. DCCs can further incorporate data processing logic and are equipped with communication interfaces such as WLAN and cellular networks. To initiate a monitoring session, the DCC device submits a predefined user *id* and a *password* to the EVC. Provided that the *id* is authenticated, the EVC starts reporting SoC values with the period of 1 second. When the DCC device issues an *end* command, this session will be terminated. In addition, for better interoperability, DCC middleware defines the set of allowable interactions between control applications and hardware components including installed sensor devices. It provides APIs and device drivers for each side.

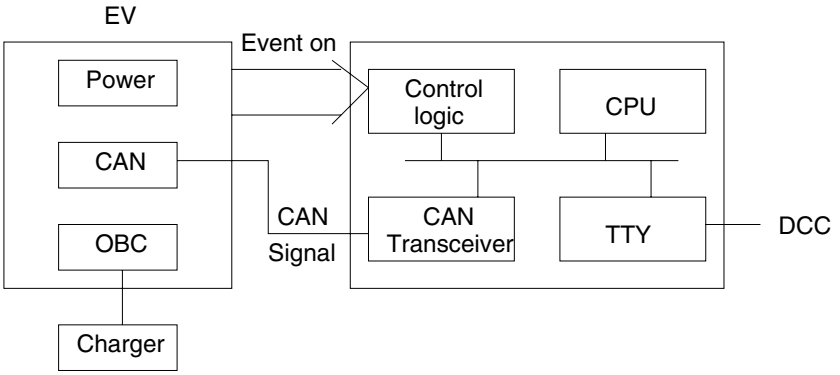


Fig. 1. Block Diagram of EVC modules

3 DCC Record Analyzer

The stream of DCC records are stored in an external memory device such as a SD-card and analyzed in desktop computers. Here, our research team develops a Window-based PC viewer program, and its execution result is shown in Figure 2. Here, if the vehicle network bandwidth is sufficiently available, the analysis can be performed in a telematics server [7]. Named RDAM (Road Data Analysis Mapper), it can load a DCC record file, display each record, and delete invalid records, if necessary. In the figure, each record consists of a time-stamp, a location-stamp, and a series of sensor readings.

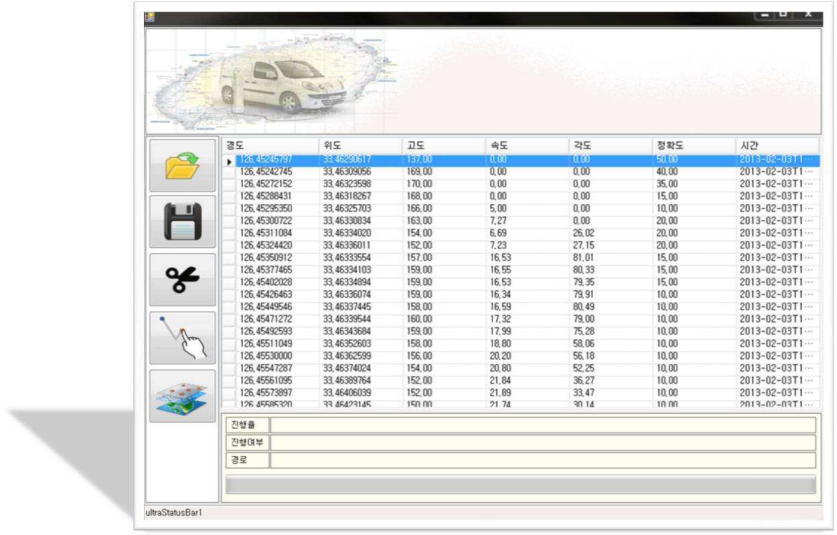


Fig. 2. Road data analysis mapper

RDAM can also display DCC records in a graph form as shown in Figure 3. This figure plots the temporal change in temperature, humidity, and slope during a test drive. According to the visualized pattern for a sample sensor stream, temperature (first graph) remains almost unchanged, while humidity (second graph) increases rather irregularly according to the change in temperature. As the data acquisition is performed along a mountain road, the slope value changes very dynamically as shown in the third and forth graphs. We can integrate diverse statistical analysis schemes for this stream including artificial neural networks and extended Kalman filters [6]. In addition, the route taken by the test drive can be displayed on Google maps.

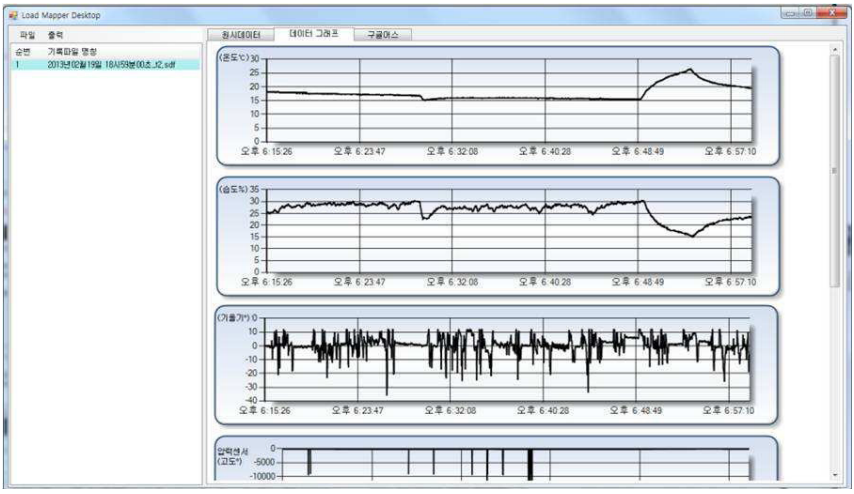


Fig. 3. Analyzer interface

Next, Figure 4 plots the change in SoC when an EV drives the 5 most commonly used routes in Jeju City. Course 1 and Course 2 are in mountain area, Course 3 and Course 4 in plain roads, and Course 5 in urban area. The SoC value is represented by the ratio to the full capacity, and the number of collected records is different course by course due to the difference in their road and thus trip length. In all test drives, the trip length is less than the driving range, so no en-route charging is necessary. SoC begins with difference values in each course, but the change pattern is our concern. For the mountain roads, SoC sometimes increases due to regenerative brake energy. In the down slope, batteries are charged when the driver pushes on the brake, not consuming any battery-stored energy, as explicitly shown in the curve of Course 2.

A driver wants to estimate how much battery will be needed from a start point to an end point. The route can be a subsequence of a course shown in Figure 5. The time interval between two consecutive DCC records ranges from 10 to 30 seconds, as some records can be lost during the delivery from BMS to DCC modules. Hence, we virtually place reference points evenly spaced by 20 *m*. A driver's query can be issued just on those points as shown in Figure 5. Now, taking the route as a straight line, our model approximates power consumption from the start point to each reference point. Next, let a DCC point be a location marked in a DCC record. A reference point falls

between two DCC points, and it is possible to interpolate the consumption from the first DCC point. It must be mentioned that even if the battery consumption between two query points is less than current battery remaining, the whole route must be investigated. As can be inferred from Figure 4, the battery can deplete if the route has steep upslopes.

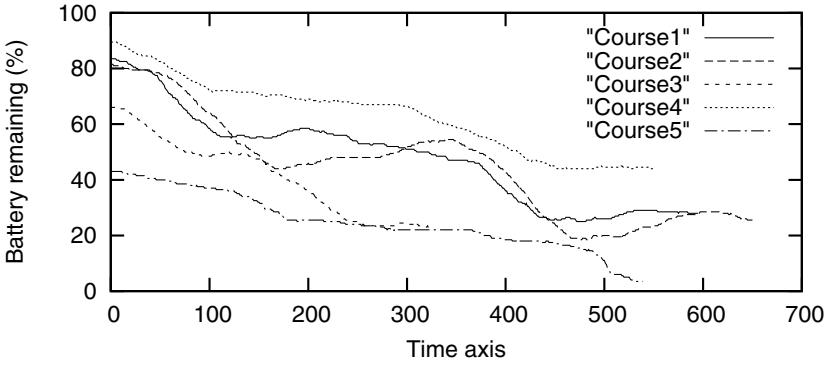


Fig. 4. SoC dynamics

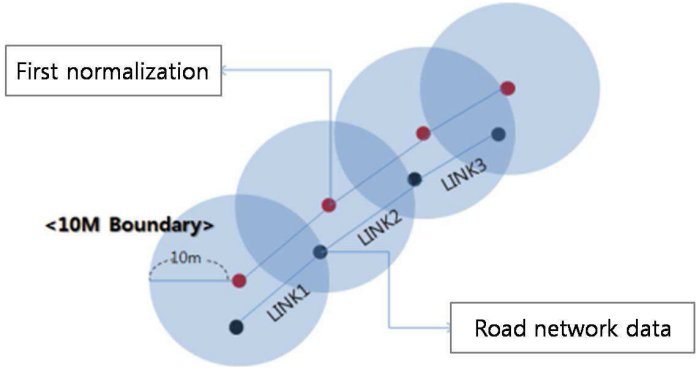


Fig. 5. Distance normalization of battery consumption

4 Conclusions

Just like many other smart grid entities, EVs can significantly benefit from computational intelligence, mainly to overcome their drawbacks in short driving range and long charging time. One of the most fundamental concerns on the drivers' side is whether an EV can reach its destination with current battery remaining, so it is necessary to build a battery consumption model for commonly taken routes. In this regard, to efficiently acquire battery consumption statistics and to obtain the corresponding consumption model, this paper has developed a driving condition collector device, which creates a real-time SoC stream, and stores for further analysis.

In addition, a window-based application visualizes and manipulates the stream, while the interpolated battery dynamics for evenly spaced reference points are built for a query which retrieves power consumption between two points in the route.

As future work, we are planning to collect more data and analyze the battery dynamics to continuously refine our consumption model. Then, an EV-specific route service will be developed, not just considering the distance criteria but also energy consumption and charging schedules. Here, availability of renewable energy on charging stations will be another important issue in making a charging plan [8].

References

1. Sortomme, E., Cheung, K.: Intelligent Dispatch of Electric Vehicles Performing Vehicle-to-Grid Regulation. In: IEEE International Electric Vehicle Conference (2012)
2. Botsford, C., Szczepanek, A.: Fast Charging vs. Slow Charging: Pros and Cons for the New Age of Electric Vehicles. In: International Battery Hybrid Fuel Cell Electric Vehicle Symposium (2009)
3. Frost & Sullivan: Strategic Market and Technology Assessment of Telematics Applications for Electric Vehicles. In: 10th Annual Conference of Detroit Telematics (2010)
4. Sundstrom, O., Corradi, O., Binding, C.: Toward Electric Vehicle Trip Prediction for a Charging Service Provider. In: IEEE International Electric Vehicle Conference (2012)
5. Kobayashi, Y., Kiyama, N., Aoshima, H., Kashiyama, M.: A Route Search Method for Electric Vehicles in Consideration of Range and Locations of Charging Stations. In: IEEE Intelligent Vehicles Symposium, pp. 920–925 (2011)
6. Chen, Z., Qui, S., Masrur, M., Murphey, Y.: Battery State of Charge Estimation Based on a Combined Model of Extended Kalman Filter and Neural Networks. In: International Joint Conference on Neural Networks, pp. 2156–2163 (2011)
7. Hattori, Y., Shimoda, T., Ito, M.: Development and Evaluation of ITS Information Communication System for Electric Vehicles. In: IEEE Vehicular Technology Conference (2012)
8. Freire, R., Delgado, J., Santos, J., Almeida, A.: Integration of Renewable Energy Generation with EV Charging Strategies to Optimize Grid Load Balancing. In: IEEE Annual Conference on Intelligent Transportation Systems, pp. 392–396 (2010)

Electric Vehicle Telematics Services Built Upon Charging Infrastructure Monitoring*

Junghoon Lee[1], Gyung-Leen Park[1,**], Seulbi Lee[1], Jihyun Kang[1], Young-cheol Kim[2], and Seong jun Lee[2]

[1] Dept. of Computer Science and Statistics
[2] Dept. of Management Information Systems,
Jeju National University
[3] EZ Information Technology,
Jeju-Do, Republic of Korea
{jhlee,glpark,gwregx,tankbaby}@jejunu.ac.kr,
nativegod12@naver.com, eziceo@gmail.com

Abstract. Aiming at prompting the wide deployment of electric vehicles by employing intelligent information technologies, this paper designs and develops a telematic service framework built upon charging infrastructure monitoring. This framework builds communication middleware working between each charger and the central server to collect and store real-time status change in availability, operation mode, and current price for each charger. As an application supported in this framework, a charging station tracking system displays the current status of each charging station and charger on the Google map. This application is ported to mobile devices running Android operating systems. Next, a charging station manager program allows station owners remotely change the sales price and manipulate the charger. Finally, for an in-vehicle telematics device, charging station information can create a better tour route combined with the availability of chargers.

Keywords: electric vehicle, charging infrastructure, telematics service framework, real-time monitoring.

1 Introduction

Thanks to many benefits especially in environmental aspects, EVs (Electric Vehicles) are drawing much attention as a key element in the future transport system [1]. For their wide deployment, charging facilities must be installed over the target area, not just in residential places but also in public places such as office buildings and shopping malls [2]. The charging infrastructure is important as the driving range is

* This research was financially supported by the Ministry of Trade, Industry and Energy (MOTIE), Korea Institute for Advancement of Technology (KIAT) through the Inter-ER Cooperation Projects.

** Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_2,

quite short for EVs due to limited capacity of their batteries. Then, location of each charging station must be known to EV drivers any time they want. Such a requirement can be easily met by employing the telematics system already developed in information and communication technology domains [3]. Moreover, intelligent computer algorithms can be integrated into this system, for example, a charging scheduler capable of reducing peak load and other optimization strategies [4].

Using vehicle telematics systems, an information server can monitor static and dynamic objects, create up-to-date information, and provide to vehicles. For EV-telematics systems, the most urgent requirement is to provide accurate and real-time information on EV charging [5]. Not just the location of chargers, this information includes charger types, temporal availability, current price, and the like. Moreover, as not every charger observes the interconnection standard, compatibility with my connector is also of great concern. In this regard, this paper designs and develops a real-time charging station tracking system which continuously collects current state of chargers. This information is made available to EV navigation devices, charging station managers, and mobile applications [6], potentially inviting new EV services.

2 Tracking System Design

In Jeju city, Republic of Korea, about 300 chargers are currently in operation over the area of 1,825 km^2. Most of them are managed by a single administrative agency, and its information server periodically collects the status information of each charger according to the predefined message exchange protocol. Basically, the manager and a managed object communicate via M2M over the 3G or ADSL (Asymmetric Digital Subscriber Line) channels. After Jeju city launched the nation-wide enterprise named *Smart grid model city*, it is pouring its effort to replace all gasoline-powered vehicles with EVs by 2030. The tracking system for the city-wide charging facilities is one of the most fundamental information services for charging infrastructure. As can be seen in Figure 1, it can announce the locations of charging stations to drivers, provide management interface to station owners, and recommend appropriate stations to navigation users.

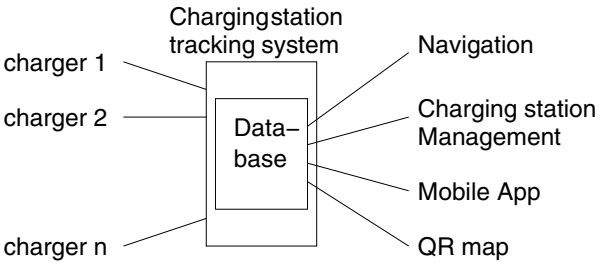


Fig. 1. System overview

Our project team develops communication middleware to interact with chargers under the restricted permission of the management agency. The periodic report includes time stamp, charger type, charging station id, and charger id in its fixed size header. Its information field further specifies charger mode, status, and total amount

of traded electricity. The information server can be aware of whether a charger is operating normally and now charging process is ongoing. Additionally, static information is investigated and stored in our database. The static information includes a station' location specified by longitude and latitude, the number of installed chargers, supported charging connector types, the manufacturer, and the like.

Based on the database tables for charging stations, our project team builds a graphic user interface on top of Google map to visualize the locations and the current state of respective charging stations and their chargers. Figure 2 shows Jeju city map on which the locations of chargers are marked. If a user selects a charger, this GUI will show address, POI, and current state of a charging station in a bubble box. Most of chargers are installed on town areas having high population, office buildings, residential houses, and an airport. Besides, a lot of tour spots facilitate chargers in their parking lots to allow tourists to take a tour while their EVs are charged [7]. This visualizer has a variety of application areas including sharing systems, emergency rescue services, and the like.

Fig. 2. Charging station tracking system

3 Extensions

The Google map-based implementation makes it possible for this visualizer to be easily ported to mobile applications on Android phones. As shown in Figure 3, to say nothing of the Google-supported map functions such as zoom in, zoom out, and pan, the mobile interface provides the drivers with diverse retrieval methods such as finding the nearest fast/slow charging station, availability of a charging facility, and compatibility with their connection cables. After finding the target station, a driver can invoke a path finding request to the Google service specifying the current position obtained from the embedded GPS receiver. Moreover, it is possible to develop a reservation mechanism on a charging station using this mobile interface.

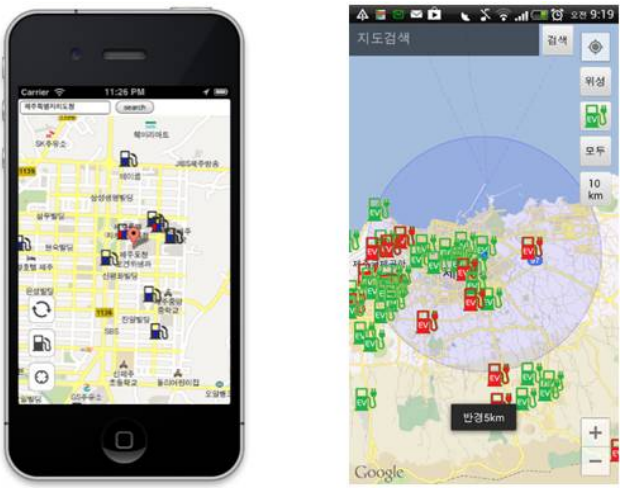


Fig. 3. Mobile application interface

Figure 4 shows a charging station management service, which displays the current states of chargers in a comprehensive user interface. In this example, the station has 4 slows chargers numbered from 1 to 4. Charger 3 is out of order and no chargers are connected to an EV. Using this interface, the manager can set the working environment parameters such as deadlines of charging operations. It also shows the unit sales price of electricity according to the real-time price signal change [8]. In addition, the management statistics provide daily and monthly amount of traded electricity and the corresponding profit. Finally, it is necessary to indicate whether this application is normally communicating with chargers and the EV telematics server to guarantee the accuracy and the timeliness of provided information.

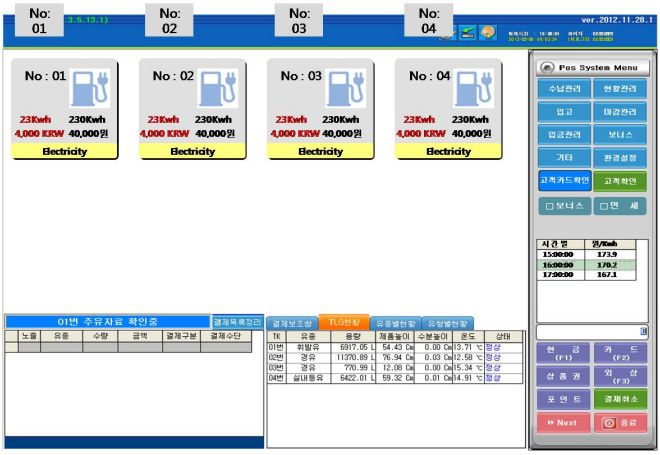


Fig. 4. Charging station operation & management service

Next, Figure 5 depicts the navigation application working with the charging station tracking system. Not just considering the distance criteria, but also the locations of chargers and a charging plan, the navigation service finds an efficient route to the destination. This application works on an in-vehicle telematics device which embraces a set of map layers developed by our project team. Particularly, we can define an additional layer which is made up of EV-specific information such as real-time charging station information, relevant POIs (Point Of Interests), and the like. Moreover, evergrowing vehicle network connectivity makes this information remained almost always up-to-date. As the telematics device has better computing power than mobile phones, more sophisticated mobile applications can run with more detailed map information, including on-spot advertisement.

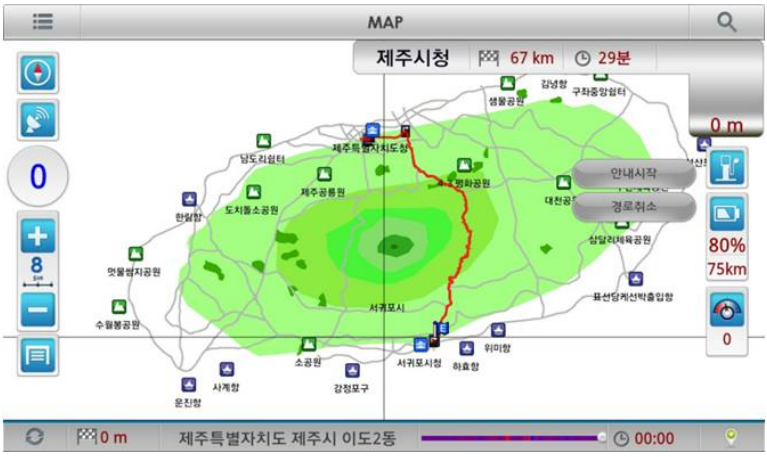


Fig. 5. Charging station information exported to navigation applications

4 Conclusions

Charging infrastructures are being constructed in many countries to accelerate the deployment of EVs. For better efficiency, it is important to announce their status to EV drivers, and an EV telematics system can meet this requirement, taking advantage of intelligent information and communication technologies. This paper has built a charging infrastructure tracking system and a set of relevant applications including a mobile visualizer interface, a charging station management service, and a navigation program combined with charging infrastructure monitoring. On this EV telematics service framework, many new applications can be designed and implemented, making it possible for EVs to penetrate into our daily lives.

As future work, we are planning to design a charging station reservation service to further reduce the waiting time in EV charging by distributing charging requests over multiple stations. In addition, renewable energy integration for charging stations will be intensively researched for better energy efficiency as well as reduction of greenhouse gas emissions.

References

1. Sortomme, E., Cheung, K.: Intelligent Dispatch of Electric Vehicles Performing Vehicle-to-Grid Regulation. In: IEEE International Electric Vehicle Conference (2012)
2. Morrow, K., Karner, D., Francfort, J.: Plug-in Hybrid Electric Vehicle Charging Infrastructure Review. Battelle Energy Alliance (2008)
3. Hattori, Y., Shimoda, T., Ito, M.: Development and Evaluation of ITS Information Communication System for Electric Vehicles. In: IEEE Vehicular Technology Conference (2012)
4. Yamashita, D., Niimura, T., Yoshimi, K., Yokoyama, R., Takamori, H.: Optimal Strategy to Support the Development of Charging Infrastructure for Electric Vehicles towards Low Carbon Emissions. In: IEEE Power and Energy Society General Meeting (2012)
5. Gharbaoui, M., Valcarenghi, L., Martini, B., Castoldi, P., Bruno, R., Conti, M.: Effective Management of a Public Charging Infrastructure through a Smart Management System for Electric Vehicles. In: IEEE Sustainable Transportation Systems Symposium, pp. 1095–1100 (2012)
6. Schapranow, M., Kühne, R., Zeier, A., Plattner, H.: Enabling Real-Time Charging for Smart Grid Infrastructures using in-Memory Databases. In: IEEE Workshop on Smart Grid Networking Infrastructure, pp. 1040–1045 (2010)
7. Lee, J., Kim, S., Park, G.: A Tour Recommendation Service for Electric Vehicles based on a Hybrid Orienteering Model. In: 28th Annual ACM Symposium on Applied Computing, pp. 1652–1654 (2013)
8. Lee, J., Park, G.: Power Load Distribution for Wireless Sensor and Actuator Networks in Smart Grid Buildings. International Journal of Distributed Sensor Networks (2013)

A Chronic Disease Identification Scheme Using Radar Chart Method for Personalized Healthcare System

Sangjin Jeong[1,3], Chan-Hyun Youn[2], and Yong-Woon Kim[3]

[1] Dept. of Information and Communications Engineering, KAIST, Daejeon, Korea
[2] Dept. of Electrical Engineering, KAIST, Daejeon, Korea
chyoun@kaist.ac.kr
[3] Protocol Engineering Center, ETRI, Daejeon, Korea
{sjjeong,qkim}@etri.re.kr

Abstract. Facing to the increasing demands and challenges to personalized disease management, various researches on the personalized healthcare systems which can provide customizable healthcare and patient disease management services have been extensively performed. Among the managed disease, chronic diseases such as metabolic syndrome or diabetes are the main target the long-term diseases care, because the diseases require the real-time monitoring, the multidimensional quantitative analysis, and its classification of patients' diagnosing information. Therefore, to enhance the effectiveness of medical decision process during patient diagnosis, we propose a personalized patient disease identification scheme for effectively diagnosing and show the validity of the proposed scheme.

Keywords: healthcare, chronic disease, radar chart, personalized healthcare.

1 Introduction

Chronic disease becomes an important issue of healthcare systems in many countries. For example, it is forecasted that clinical expenditure for chronic disease in U.S. would be 80% of total medical costs and more that 150 million people might be suffer from chronic disease in 2020. Health status monitoring in out-of-hospital environments particularly patients self-management at home environment has been a major issue of healthcare researchers and developers for long time. Continuous monitoring of health status during daily life activities is essential for effectively managing chronic disease [1]. From a medical service provider's point of view, to provide advanced quality healthcare service for chronic disease, the following issues need to be resolved. Patients require continuous health status monitoring and care over a long term period. Their disease status sometimes may be changed unexpectedly. However, there exist few medical systems that provide any alarm about chronic patient status. The conventional medical examination processes for chronic disease status detection are complicated. The medical systems need to generate reliable outcomes for patients with complex chronic conditions [1]. Therefore, it is important to provide patients with self-management capability and enable patients'

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_3,

own management of disease conditions. The healthcare system are required to assist patients' self-management for chronic condition better through delivering more exact information and suggesting suitable for disease management.

Conventional healthcare systems have focused on providing specific target services only. However, achievements in various ICT technologies have enabled a lot of research on personalized healthcare systems for at home environment. But, the research is mostly focusing on the patients' medication treatment of chronic disease, e.g., to deliver the right treatment, to the right patient, at the right dose and at the right time [2]. Healthcare services, such as health monitoring, medical consultation and so on need to be personalized based on the context of patients' profiles. For efficient provision of personalized healthcare services, it is necessary to accurately identify patients' health status, particularly chronic diseases.

There are several studies for developing personalized disease identification schemes [3][1]. Among them, [1] proposed a novel Patient Status Classification Method (PSCM), which is based on patient tier classification and radar chart priority calculation using surface measure of overall performance (SMOP) theory [4]. The PSCM model for patients with chronic diseases offers automatic medical service procedures in the form of an effective medical information visualization system. It reduces the workload by offering readily available data. The PSCM process contains three parts: the Patient Tier Classifier, the Disease & Complications Identifier, and the Health Risk Quantification [1]. Although radar chart approach (RCA) and SMOP method are very effective ways for identifying goods or best performers while maintain the interdependence of different policy goals in evidence, it is known that those methods have following several weakness [5].

1) Not theory-driven performance indicators
2) Equally weighted performance indicators which are problematic and unjustified
3) No explanation for performance levels, changes and structures of performance
4) No information about efficiency measurement

In order to resolve the weakness of RCA and SMOP above, in this paper, we propose an analytic hierarchy process (AHP) based weighted RCA scheme in order to accurately identify patient disease. We evaluate our proposed scheme by using sample patient physiological data.

In the following sections, the description of proposed patient status classification model for chronic disease care is presented, along with the chronic disease identification procedures.

2 A Chronic Disease Identification Scheme Using Analytic Hierarchy Process and Radar Chart

As we discussed in the previous section, RCA is a very useful method for qualitative data analysis despite of its weakness and it is useful for preliminarily identifying patient's chronic disease and disease status [1]. So, in order to mitigate the weakness

of RCA and SMOP, we have adopted AHP in order to determine the weights of performance indicators and propose a novel patient disease identification scheme based on areal similarity degrees between two radar charts, one displaying the typical characteristics of designated disease by averaging patients medical test results, and the other showing patient's medical test results. We evaluate our proposed scheme by using sample patient physiological data.

AHP is a multi-criteria decision making method developed by Thomas Saaty [6]. AHP allows decision makers to model a complex problem in a hierarchical structure, showing the relationships of the goal, objectives, and alternatives. AHP is made up of several components such as hierarchical structuring of complexity, pairwise comparisons, judgments, an eigenvector method for deriving weights, and consistency considerations. So, we apply AHP to determine the weights of indicators for each medical test results.

Many variables values are put into the same coordinate plane, the area is representation function for the whole quality, and the shape gives the detail characteristics. In the weighted radar chart, every input variable value is expressed by radial r_i of unit circle and w_i is weight coefficient. A circle in which the radius represented input measurement ranges is one could be divided into n parts according to different weight coefficient w_i and the sum of the w_i coefficient is equal 1. On the circle, the n rays represent, the n input variables and the r_i measured value of an input variable will fall in a relevant ray. Connecting the points r_i, which corresponds to measured values of different inputs, a weighted radar chart could be obtained [7]. Since input data for each performance indicator in the weighted radar chart have difference measurement scales, it is important how to transform the original input data to the transformed new input range of weighted radar chart, i.e., transformed input range between zero and one. Since the distribution of medical test results generally follows normal distribution, let the result of i_{th} medical test be x_i and the arithmetic mean and standard deviation of x_i be $E(x_i)$ and $\sigma(x_i)$, respectively. Then, $x_{i_{new}}$, the standardized value of x_i can be written as follows:

$$x_{i_{new}} = \frac{x_i - E(x_i)}{\sigma(x_i)} \tag{1}$$

In order to depict $x_{i_{new}}$ with radar chart whose accepted input range is 0 to 1, it is necessary to transform $x_{i_{new}}$ to new input range. Let $x^*_{i_{new}}$ be the transformation of $x_{i_{new}}$, then $x^*_{i_{new}}$ can be expressed as follows [9].

$$x^*_{i_{new}} = [\tfrac{2}{\pi} tan^{-1}(x_{i_{new}}) + 1] \times \tfrac{1}{2}, \text{ where } 0 \le x^*_{i_{new}} \le 1 \tag{2}$$

Since each medical test result has different influence on disease status, it is necessary to separately weight impact of test results to the disease status. In order to separately weight importance of the types of medical tests, we have utilized the results of risk factor analysis for metabolic syndrome in order to calculate the weights of medical test results. Fig. 1 shows the decomposition of risk factor structure for the metabolic syndrome symptoms based on the correlation analysis results presents in [10].

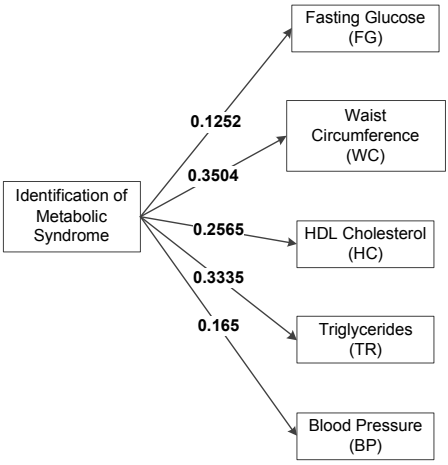


Fig. 1. Decomposition of risk factor structure for the metabolic syndrome symptoms

Pairwise comparison matrix A can be setup based on the hierarchy shown in Fig. 1.

$$A = \begin{array}{c} \begin{matrix} \text{FG} & \text{WC} & \text{HC} & \text{TR} & \text{BP} \end{matrix} \\ \begin{bmatrix} 1 & 0.3573 & 0.4881 & 0.3754 & 0.7588 \\ 2.7988 & 1 & 1.3661 & 1.0507 & 2.1236 \\ 2.0488 & 0.7320 & 1 & 0.7691 & 1.5545 \\ 2.6638 & 0.9517 & 1.3002 & 1 & 2.0212 \\ 1.3179 & 0.4709 & 0.6433 & 0.4948 & 1 \end{bmatrix} \end{array} \begin{matrix} \text{FG} \\ \text{WC} \\ \text{HC} \\ \text{TR} \\ \text{BP} \end{matrix} \tag{3}$$

Where a_{ij} is the relative importance of the i_{th} element in j_{th} indicator criterion level in terms of its contribution to the disease status, n is the rank of this matrix. Fig. 2 provides the numerical ratings recommended for the verbal preferences expressed by the decision maker [6].

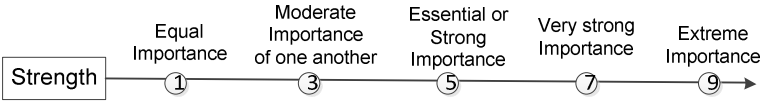


Fig. 2. Pairwise comparison scale for AHP preference

Once the pairwise comparison matrix has been established, the weight of each element being compared can be calculated. In this paper, we have used the logarithmic least square method in order to obtain the weights. The relative weight matrix B can be obtained by solving the following equations [6].

$$\min \sum_{i<j} \sum_{j=1}^{n} [\ln a_{ij} - \ln(\frac{w_i}{w_j})]^2 \tag{4}$$

$$\prod_{i=1}^{n} w_i = 1, \quad w_i > 0 \text{ for } i = 1,\ldots,n \tag{5}$$

Obtained eigenvector W of the hierarchy is as follows.

$$W=[0.1017, 0.2847, 0.2084, 0.2710, 0.1342]=[\text{FG, WR, HC, TR, BP}] \tag{6}$$

In order to check the consistency of the risk factor values in pairwise comparison matrix, a consistency ratio (C.R.) is used to determine the degree of consistency. If $C.R. \le 0.1$, it means that the consistency level is satisfactory. The C.R. and consistency index (C.I.) are defined as follows.

$$C.R. = C.I. / R.I. \tag{7}$$

$$C.I. = \frac{\lambda_{max} - n}{n-1} \tag{8}$$

where λ_{max} is the maximum eigenvalue of the pairwise comparison matrix. The random index (R.I.) is shown in Table 1 [6].

Table 1. Values of random index (R.I.) [6]

Matrix order (n)	1	2	3	4	5	6	7	8
Random index (R.I.)	0	0	0.58	0.9	1.12	1.24	1.32	1.41
Matrix order (n)	9	10	11	12	13	14	15	
Random index (R.I.)	1.45	1.49	1.51	1.48	1.56	1.57	1.59	

By substitute variable above by numerical value, the C.R. of the pairwise comparison matrix can be calculated as follows. Since C.R. is less than or equal 0.1, the consistency level is acceptable.

$$\lambda_{max}=5.00,\ n=5,\ \ C.I. = \frac{\lambda_{max} - n}{n-1} = \frac{5.00-5}{5-1} = 0 \tag{9}$$

$$C.R. = \frac{C.I.}{R.I.} = \frac{0}{1.12} = 0 \tag{10}$$

The computed weights and allocation of angle for medical tests indicators are shown in Table 2. From the table, we can observe that the examination results of waist/hip ratio, triglyceride, and HDL-cholesterol contribute the main factor of metabolic syndrome disease, which comply with the results of [10].

Table 2. Computed weights for each indicator and allocation of angle in radar chart

Examination Test Type	Weight (%)	Allocation of Angle (°)
Fasting Glucose (FG)	0.1017	36.612
Waist Circumference (WC)	0.2847	102.492
HDL Cholesterol (HC)	0.2084	75.024
Triglycerides (TR)	0.2710	97.56
Blood Pressure (BP)	0.1342	48.312

The circle of radar chart is marked off in accordance with the number of disease indicators and the weights calculated above. Some radiate lines are formed by the center of the circle and the marked point. These lines are regard as coordinate axis. Mark the data pretreated before on these coordinate axis and connect the marked points, then polygons for the values of disease indicators can be obtained. This is the radar chart of disease status for patient. Fig. 3 depicts a weighted radar chart using patient's sample medial test results. Table 3 shows the sample medical test results and characteristics of type 2 diabetes mellitus patient [8].

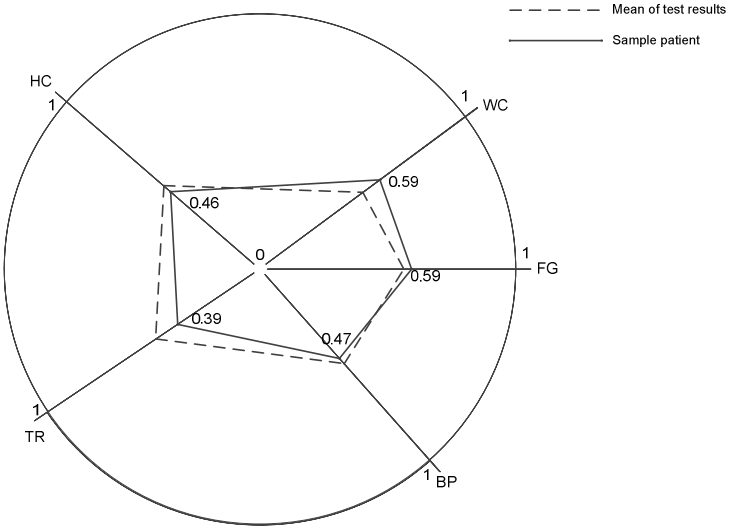


Fig. 3. Weighted radar chart of disease status

Table 3. Characteristics of type 2 diabetes mellitus patient

Examination Test Type	Sample Patient	Characteristics of diabetes mellitus patients (N=108)
Fasting Glucose (FG) (mg/dl)	90	176.4 ± 63.7
Waist Circumference (WC) (cm)	125.0	88.2 ± 5.5
HDL Cholesterol (HC) (mg/dl)	195.8	49.6 ± 13.5
Triglycerides (TR) (mg/dl)	48	163.9 ± 122.9
Blood Pressure (BP) (mmHg)	120	126.4 ± 15.2

3 Evaluation

According to [8], five risk factors of metabolic syndrome for Korean are defined as follows.

Table 4. Thresholds of five risk factors of metabolic syndrome for Korean

Examination Test Type	Thresholds
Fasting Glucose (FG) (mg/dl)	≥ 110mg/dl
Waist Circumference (WC) (cm)	≥ 90 (for man) ≥ 80 (for woman)
HDL Cholesterol (HC) (mg/dl)	< 40 (for man) < 50 (for woman)
Triglycerides (TR) (mg/dl)	≥ 150mg/dl
Blood Pressure (BP) (mmHg)	≥ 130/85

A patient is determined as having metabolic syndrome, if the patient's medical test results exceed the thresholds of risk factors in Table 4 with respect to three more risk factors. Therefore, by comparing the weighted radar chart of the patient with that of thresholds, it is possible to effectively determine whether the patient holds metabolic syndrome. Furthermore, by calculating the area of the two radar charts, we can compute the status of the patient's metabolic syndrome. Fig. 4 shows the two weighted radar charts of the patient and metabolic syndrome thresholds.

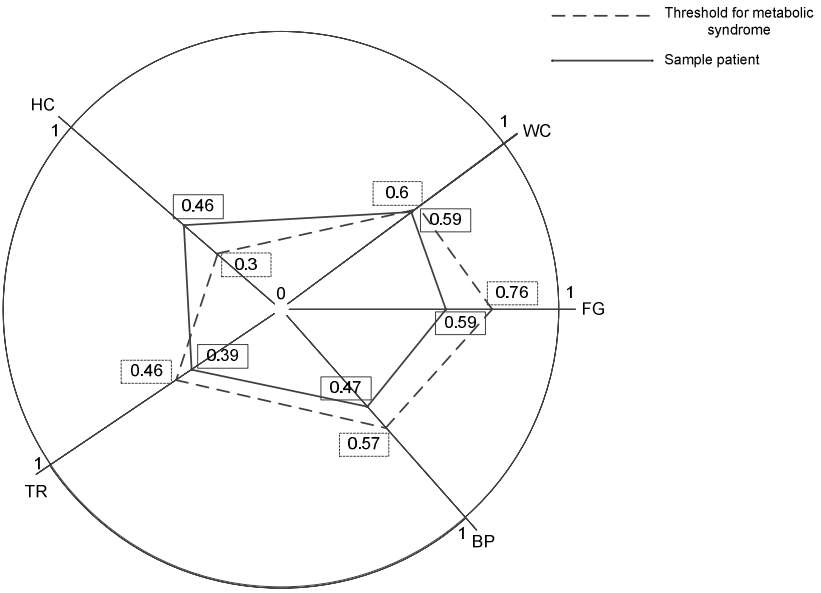


Fig. 4. Two weighted radar charts of the patient and metabolic syndrome thresholds

According to the test results, the patient is not determined as having a metabolic syndrome, because only two test results (fasting glucose and HDL cholesterol) exceeds the threshold. However, we can calculate the status of the patient having metabolic syndrome by comparing the overlapping area between two radar charts. In this example, we can compute the status of metabolic syndrome as 89.6% of metabolic syndrome thresholds.

4 Conclusions

It is known that the radar chart is a very useful method for qualitative data analysis despite of its weakness and useful for preliminarily diagnose patient' status of metabolic syndrome. So, in order to mitigate the known weakness, in this paper, we proposed an analytic hierarchy process based weighted radar chart scheme in order to effectively diagnose patients' chronic disease status, particularly the metabolic syndrome. Then, we presented the validity of the proposed scheme by using sample patient physiological data. Our evaluation results showed that the proposed scheme can be effectively used for medical decision process while physicians diagnose potential patients with metabolic syndrome.

Acknowledgment. This research was supported by the ICT Standardization program of MKE (The Ministry of Knowledge Economy).

References

1. Jeong, S., Youn, C., Shim, E., Kim, M., Cho, Y., Peng, L.: An Integrated Healthcare System for Personalized Chronic Disease Care in Home-Hospital Environments. IEEE Trans. Information Technology in Biomedicine 16(4), 572–585 (2012)
2. Koukias, V.G., et al.: A Personalized Framework for Medication Treatment Management in Chronic Care. IEEE Trans. Information Technology in Biomedicine 14(2), 464–472 (2010)
3. Dahlstorm, O., Timpka, T., Hass, U., Skogh, T., Thyberg, I.: A Simple Method for Heuristic Modeling of Expert Knowledge in Chronic Disease: Identification of Prognostic Subgroups in Rheumatology. In: 21st International Congress of the European Federation for Medical Informatics, pp. 157–162 (2008)
4. Schutz, H., Speckesser, S., Schmid, G.: Benchmarking labour market performance and labour market policies: theoretical foundations and applications. Discussion paper, No. FS I 98-205 (1998), http://hdl.handle.net/10419/43918
5. Schmid, G., Schutz, H., Speckesser, S.: Broadening the Scope of Benchmarking: Radar Charts and Employment Systems. Labor. 13(4), 879–899 (1999)
6. Saaty, T.L.: The Analytic Hierarchy Process. McGraw-Hill, New York (1980)
7. Li, X., Hong, W., Wang, J., Song, J., Kang, J.: Research on the Radar Chart Theory Applied to the Indoor Environmental Comfort Level Evaluation. In: 6th World Congress on Intelligent Control and Automation, pp. 5214–5217 (2006)
8. Kim, H.: Differences in Prevalence and Risk Factors of the Metabolic Syndrome by Gender in Type 2 Diabetic Patients. Korean Journal of Adult Nursing 18(1), 3–9 (2006)
9. Yijing, L., Ming, L.: Evaluation of Drawing Ability Based on Radar Chart. In: International Conference on Information Technology and Computer Science, pp. 574–576 (2009)
10. Shen, B., Todaro, J., Niaura, R., McCaffery, J., Zhang, J., Spiro, A., Ward, K.: Are Metabolic Risk Factors One Unified Syndrome? Modeling the Structure of the Metabolic Syndrome X. American Journal of Epidemiology 157(8), 701–711 (2003)

Wireless Data Collection in Power System

Roman Kuznetsov and Valeri Chipulis

Laboratory of Technical Diagnostics, IACP FEB RAS, Vladivostok, Russia
{kuznetsov,chipulis}@dvo.ru

Abstract. The automatic meter reading (AMR) systems intended for heat energy accounting are discussed. The technology for rapid design and development of AMR systems is suggested. It is allowed to increase the reliability and the effectiveness of data collection process in wireless networks.

Keywords: power system, heat meter, GSM/GPRS modem, wireless communications, information system, M2M.

1 Introduction

In recent years, an intense process to install metering equipment for energy accounting is observed due to the came into force the federal law on energy saving and increasing energy efficiency in the Russian Federation. Heat meters are used to organize the energy accounting in the thermal points of consumers. Heat meter performs the following basic functions: measuring of the heat-transfer medium parameters (flow, temperature and pressure), the calculation of the energy consumption based on the measurements, accumulation and storage measured values in the archives, providing access to the historical data on request via the communication interface. Large-scale implementation of heat meters leads to the creation of data centers for the collection and processing. The continual increase of heat meters and the geographic expansion of heat-power objects with the metering of energy resources require novel approaches to creating data centers by using the modern information technologies and advances in telecommunications. Since 2000 the Institute of Automation and Control Processes FEB RAS in collaboration with the IT-company Infovira has been developing information-analytical systems (IAS) in heat-power engineering [1]. Those systems are aimed to solve the problems of energy accounting, energy audit and energy savings as well as to support to make decision ad hoc tasks associated with the technical diagnosis [2-3] and the effective control of heat-power objects [4]. However, the priority task is to get the measurements with the metering equipment. This problem in the IAS is solved by data collection subsystem (Fig. 1) which is used to provide telemetry data, primary processing and saving of measurements in a database.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_4,

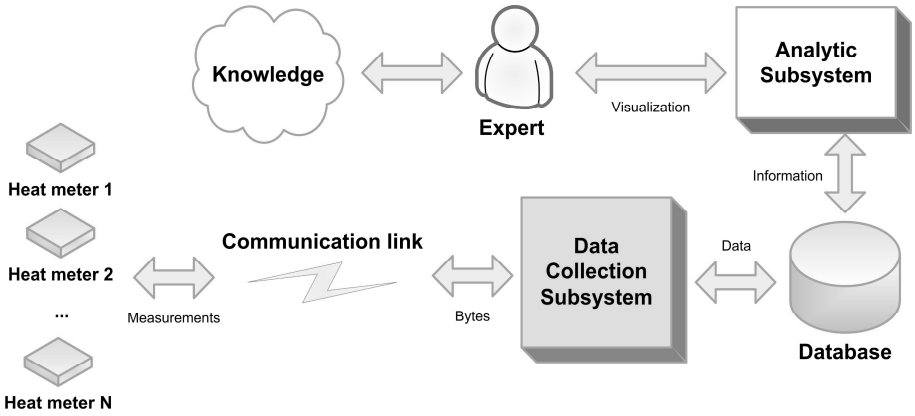


Fig. 1. The architecture of IAS in heat-power engineering

2 Data Collection

Initially there were a few amounts of heat meters. All heat meters had a same type. At that time data collection was a very simple. The notebooks are used for getting data from heat meters by wire connection or optical cable through COM port directly. Progressive increase amount of devices as well as the long distances between heat meters (in different cities or regions) leads to necessity of using of modems for remote data collection. Primarily the software was developed for manually query of heat meter by Public Switched Telephone Network (PSTN). However, the data collection by PSTN modems has several disadvantages. The main problem is the inability of the connection to heat meter during the work day, because at that time the phone is used by people. In the night hours the data collection from a large amount of heat meters using dial up connection is an impossible. Therefore, when the mobile networks have been widely used, the most of PSTN modems are replaced on GSM terminals. The software for automatic data collection by way of mobile networks (GSM) was developed. It was used for mass service of heat meters. The data transmission is provided by CSD technology and based on the schedule is configured by the user.

The data collection system (Fig. 2) was operated during several years. The following drawbacks were identified in the course of operation of data collection system:

- monolithic program at the code level;
- poor portability;
- different data structures for each device driver;
- no support the advantages of the novel mobile communication;
- low reliability;
- lack of performance.

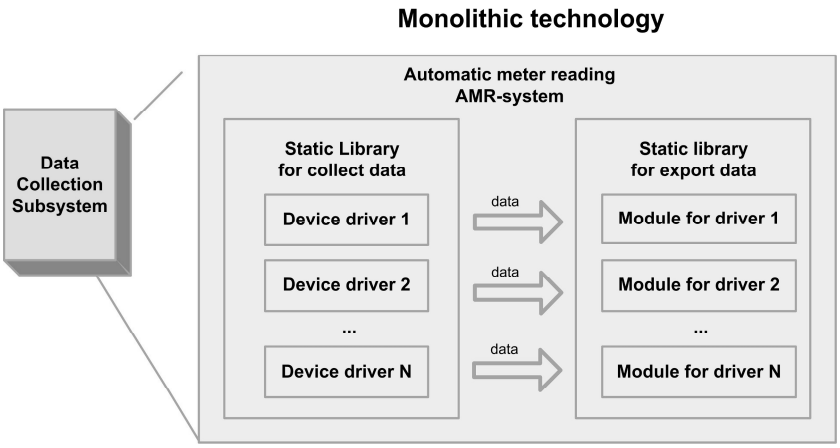


Fig. 2. Previous version of data collection

3 New Technology for Rapid Development

The experience of operation shows that further development of the data collection system is required with taking into account the advanced telecommunications and information technology. The modernization of the data collection system is needed for several reasons. The first, there are new types of metering devices with the more complexity configuration. It makes necessary to abstract from the hardware configuration and settings. The second, the amount of measuring equipments that needs the maintenance is constantly increased. For this reason is required to parallelize data collection for them on a single server or maybe multiple servers. The thirdly, the inclusion of ready-made solutions in the data collection system is one of priority task (such as the support of OPC technology). The last meaningful requirement is the support of different communication channels: communication ports; modems; cellular, wired and wireless networks. The analysis reveals what the improvement of the data collection system is an impossible without a revision of its architecture. Consider the basic innovations used in the new development technology of data collection system (Fig. 3).

3.1 Device Driver

A new device driver model is suggested. According to the new technology the each device driver is represented in the data collection system as a separate software module - a dynamic link library (e.g. DLL). Consequently the supervise application (top-level) loads the library only if the device driver is needed for data collection. The library should be unloaded from memory after this process is completed. The set of device driver functions has been specified and the communication interface with the top-level applications has been unified by using the abstract data structures that are created dynamically in the process of data collection. The unique driver is used for

each device type because there is no common protocol among different manufacturers now. The ability to add new features to the device driver model is provided for future development of the system. The modularity at the level of shared libraries is not supported in the previous system version.

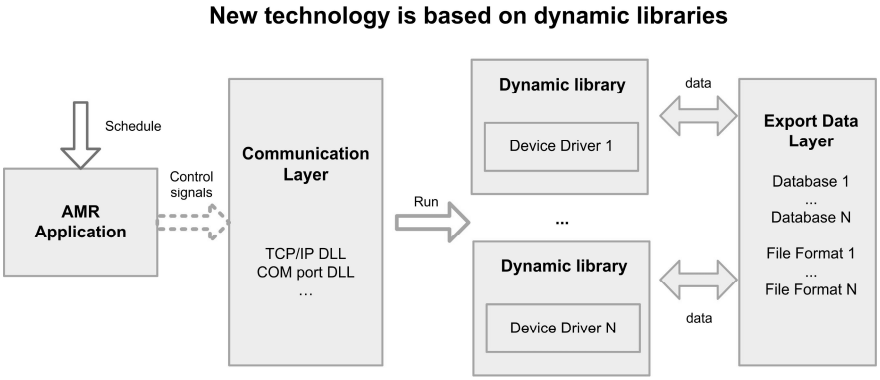


Fig. 3. New data collection system

3.2 Multichannel Data Collection

A multithreaded processing has been realized for simultaneously data collection on all available communication channels and devices. A communication layer (intermediate) is used to establish connection with the meters through the various types of channels. The options of the communication channel are configured in the top-level applications and passed to the device driver. The communication layer is implemented as a set of dynamic link libraries with fixed interfaces. In the previous version of data collection system was used only dial-up by CSD. Also, a significant drawback of the previous version is a single thread of polling. Starting the device drivers in a separate process makes it possible to avoid the critical errors in data collection system as a whole. Because that bugs in the device driver don't lead to the crash of the system. The over execution time of the driver thread is controlled in case of hung-up. In the previous version the main application was restarted with an initial poll schedule due to critical errors in the driver. It adversely affects the efficiency of data collection system.

3.3 Reliability

The new technology provides the possibility of redundancy and duplication of data collection system. The basic idea is to install the automatic data collection on multiple servers like a cluster. It is assumed that the possibility of a separate schedule for each server is organized by using unified information database. This increases both the performance of data collection system and the reliability due to duplication on a hardware level. In previous version the schedule is maintained only for one server of data collection.

3.4 Schedule

The schedule of data collection has been improved in new version as well:

- the setup of the time interval for the data collection during the day has been added (e.g. since 1 am until 3 pm);
- the data collection for each heat meter has been realized by certain days of the week and day of the month (e.g. only on weekdays or during the time of a monthly report);
- the limitation of unsuccessful attempts for each device has been implemented for cost efficiency of data collection;
- temporarily deactivating the data collection from device is a available without removing the information from the common schedule.

Advanced settings in the schedule are to optimize data collection system for more efficiently function in case of a large amount of serviced devices.

3.5 Data Export

The export of collected data is implemented as a separate layer. The data obtained from the meters is transmitted by the device driver into a special library that transforms it in the required format and saves the measurements in the database. The interface between the device drivers and libraries for export data is standardized. It allows binding of the device driver with different data warehouses and vice versa too. This approach is easy to export data into various databases and file formats, because it does not require the development of libraries for each driver.

4 Application and Evaluation

The approbation of new AMR system for data collection was done in real operating conditions for more than 700 heat meters installed at the thermal points around the Primorsky province, Russia. To assess the efficiency of data collection as an increasing number of serviced objects is impossible without the implementation of automated control and diagnostics at all stages of the process. The sequence of events occurring in the process of automatic data collection is logged and stored into database. The visualization of data collection events from database carried out in "Express-analysis" software [2] (Fig. 4). This software is designed to display data collection and verification information for a group of heating objects on the defined time interval (by default one week) in a readily accessible form for the service staff of the thermal points. Detailed information about data collection events during the reporting day for any heat meter, you can see in the dialog box by selecting the appropriate cell in the table. In dialog box on the tab "Events" will be shown a full list of events ordered by time with description of connection results. The connection problem can be found very easily and quickly by using colorization of cells in table.

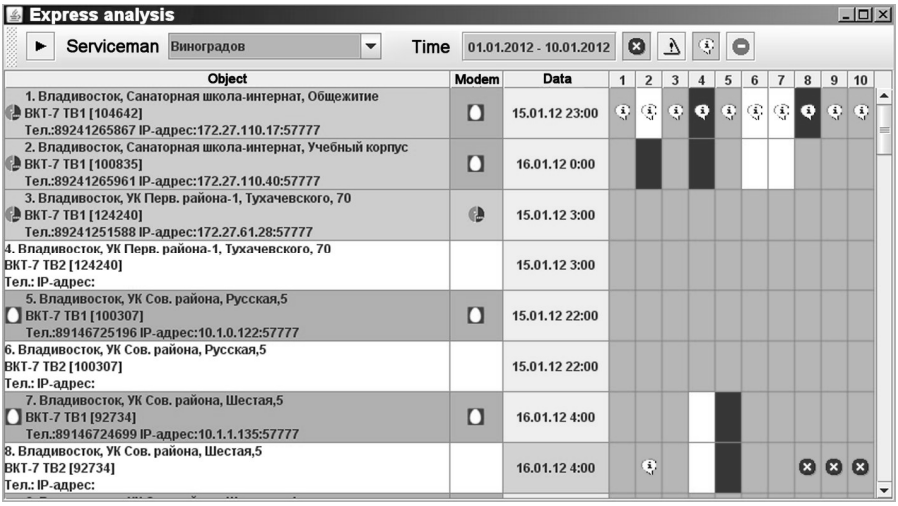


Fig. 4. The summary of data collection

5 Conclusion

The most important results are reached:

- The technology of rapid design and development of data collection systems improving the reliability and the efficiency has been suggested.
- On the basis of the proposed technology has been created new software for data collection.
- The automatic data collection with the energy accounting devices via various communications (including the TCP/IP) has been organized.

In general, the new technology for development of data collection systems improves the quality of information-analytical systems which are implemented at the heat-power system.

References

1. Bogdanov, Y., Chipulis, V.: Information-Analytical Systems of Thermo-Power Engineering. In: Sénac, P., Ott, M., Seneviratne, A. (eds.) ICWCA 2011. LNICST, vol. 72, pp. 116–124. Springer, Heidelberg (2012)
2. Kuznetsov, R.S., Chipulis, V.P.: Express-analysis in the Heat and Power Systems. Lecture Notes of Information Technology (LNIT), vol. 13, pp. 87–92. Information Engineering Research Institute, Delaware (2012)
3. Chipulis, V.P.: Diagnostics of Metrological Defects in Heat Metering Problems. Automation and Remote Control 66(11), 1850–1860 (2005)
4. Chipulis, V.P.: Evaluation of the Reliability of the Results of Measurements in Thermal Power Engineering. Measurement Techniques 48(5), 497–505 (2005)

Automatic LED Lighting System Using Moving Object Detection by Single Camera

Giao Pham Ngoc[1], Suk-Hwan Lee[2], and Ki-Ryong Kwon[1,*]

[1] Dept. of IT Convergence & Application Engineering,
Pukyong National University, Pusan, South Korea
ngocgiaofet@gmail.com, krkwon@pknu.ac.kr
[2] Dept. of Information Security, Tongmyong University, Pusan, South Korea
skylee@tu.ac.kr

Abstract. Smart LED lighting systems have been interested and developed for spaces such as buildings, rooms, garages and marine. The light sensor and motion sensor to control the automatic brightness of LEDs are very effective for automatic LED control system. In this paper, we address an idea which replaces light sensors, motion sensors by a camera. Our system integrates image processing to control an illumination system based on LEDs for buildings, rooms and garages, assume that an array of LEDs is installed on the wall. Here, our system detects moving object using adaptive Gaussian mixture model (AGMM), estimates the mapped positions of objects using inverse perspective transformation and a dimmed LED, and calculate the brightness of image from only a single camera. We implemented our system using LED array board with AVR micro controller and verified that our system is effective for automatic LED control system in specific room.

Keywords: Automatic LED control system, Moving object detection, AGMM, Inverse perspective mapping.

1 Introduction

In recent years, usage of LED light has been increasing. They are applied in mobile phone, LED television, lighting and so on. In addition, LED technology has been researched by many companies and research centers with advanced countries. At this point, the LED light is used the most smart lighting applications in spaces such as home, parking lot [1], buildings [2, 3] and now it can be substituted for the existing lights because of low cost, durability and saving energy. For smart lighting applications in spaces such as buildings, rooms, street so on, the LED light has a variety of advantages compared with the existing lights. First of all, it is easy to interwork with other electronic modules such as sensors [1-4] and communicated modules to provide new services, saves energy [1, 9] and can be controlled more elaborately because LED is a kind of electronic components. However, LED can be

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_5, © Springer-Verlag Berlin Heidelberg 2014

dimmed by PWM (Pulse Width Modulation) [5] or current control to change its brightness easily. Furthermore, it has a low power characteristic, whose power consumption is much lower than the existing lights.

One of visions – based LED control system is to integrate LED system with other devices or interfaces such as light sensors, motion sensors or TCP/IP, wireless to automatically control, manage for applications as smart home, and green house or energy saving. The purpose of integration LED system with other devices or services is optimal, automatic control and management. Smart LED lighting systems use light sensors to determine the brightness of environment and motion sensors to detect motions for control turn on/off and adjust the brightness of LEDs follow position of motion and brightness of environment [1-4]. These systems often use many motion sensors, light sensor and other devices while each motion sensor just controls a LED in LEDs system [1-3].

Previously, we proposed an idea to remove the role of light sensors and motion sensors by a camera [7] to detect the brightness of environment and motion of objects, and then control LEDs system. In previous paper [7], we integrated a camera with LEDs system to automatically control turn on/off and adjust the brightness of each individual LED follow position of moving object and illumination of environment based on video processing. Our algorithm in previous proposed system used a camera to replace role of light sensor and motion sensors in previous systems but it just detected and processed an object. In this paper, we continue to improve our previous system by detecting many objects and automatically to control many LEDs which are corresponded to objects. Related problems and experimental results will be presented in next sections.

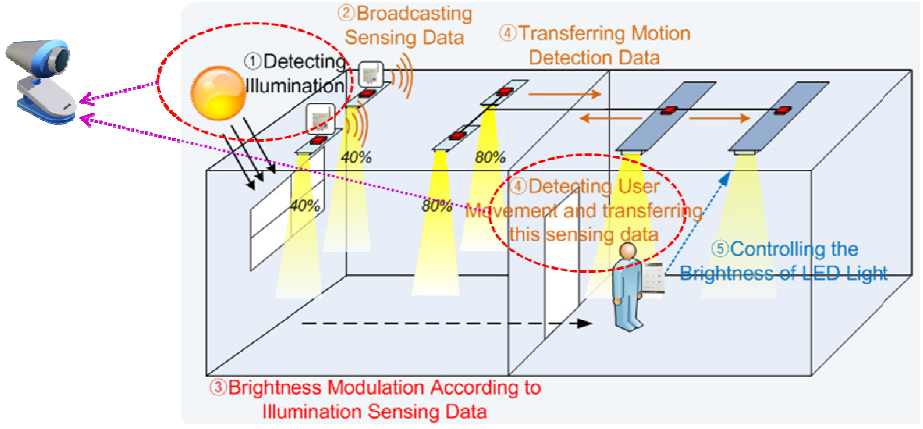


Fig. 1. Proposed scenario for automatic LED control system

2 Proposed Automatic LED Control System

2.1 The Scenario of Our System

The scenario of our system is described as an example of figure 1. 1) A single camera is calculating illumination of daylight in a room. 2) Main processor is broadcasting

the sensing data and 3) determining the brightness modulation according to illumination sensing data. Steps 1), 2), 3) keep going without moving object. 4) When a camera detects one or more moving persons, a camera captures them. 5) Main processor detects moving objects from captured images or video and light of LEDs that are positioned on head of person. The brightness of LEDs is adjusted by the brightness of environment which is computed via the brightness of received image from camera.

The structure of our system includes following units. Single camera module that gets video sequence, the embedded board or person computer (PC) for video processing, interface module that transfers data to LED board of LED driver and LED array, LED driver that receives data from embedded board/PC, LED array that displays on the position of moving object, and the power supplies.

2.2 Data Processing of Our System

Our system operates on embedded board/PC and LED board. The process on embedded board/PC includes video processing, interface and transfer data between computer and LED board, as shown Figure 2(a). The process on LED board gets data from computer, and controls LEDs, as shown figure 2(b).

2.2.1 Detect Moving Object

There are many existing algorithms for the detection of moving multiple objects. Most of them are based on motion estimation and background subtraction. Similar as You's method [6], we firstly use adaptive Gaussian mixture model (AGMM) to model background and get regions of moving objects by background subtraction. Thus, defined AGMM with maximum components Max as the probability density $p(\vec{x}|X_T, Bg + Fg)$, we decide the pixel belongs to background if

$$p(\vec{x}|X_T, Bg + Fg) = \sum_m^M \pi_m N(\vec{x}: \vec{\mu}_m, \sigma_m^2) < c_f. \tag{1}$$

X_T is a data of training set and Bg and Fg indicate backgrounds and foregrounds. M is the current number of component and $\vec{\mu}_m, \sigma_m^2$ are the estimate mean and variance value. π_m is the weighting value and c_f is threshold value.

Second, we perform the motion estimation to discriminate moving object from dynamic component of background. Set kth search region of ith frame as b_k^i and motion distance of b_k^i as n_k^i.

$$n_k^i = \begin{cases} n_k^{i-1} + |\lambda_k^i|, & O\left(g\left(b_k^i\right), B_k^{i-1}\right) < \varepsilon \\ 0, & O\left(g\left(b_k^i\right), B_k^{i-1}\right) > \varepsilon \end{cases} \tag{2}$$

where $|\lambda_k^i|$ is motion vector. $O\left(g\left(b_k^i\right), B_k^i\right)$ is calculated similar regions with the motion state $g\left(b_k^i\right)$ and the set of close regions in previous frame B_k^i. We judge the moving target region by the following equation.

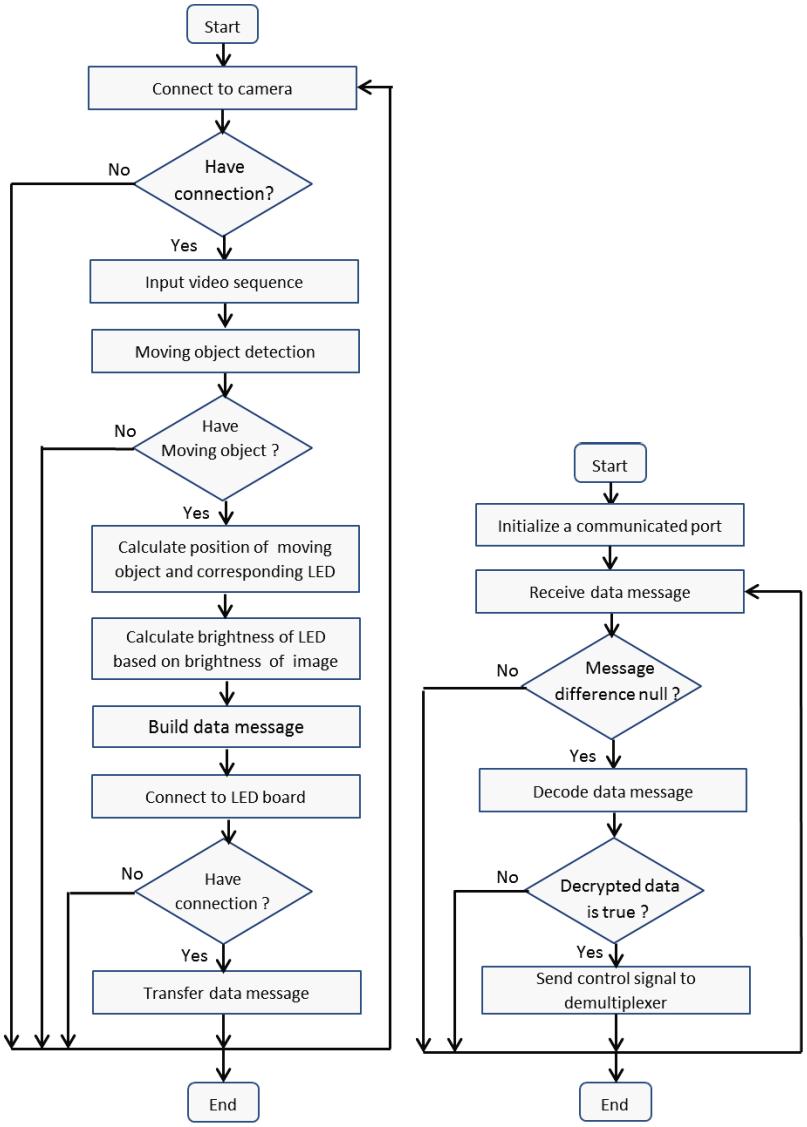


Fig. 2. Algorithm data processing for system; (a) algorithm on embedded board/PC and (b) algorithm on LED board

$$f(n_k^i) = \begin{cases} 1, & n_k^i \geq \varphi\sqrt{w^2 + h^2} \\ 0, & x < \varphi\sqrt{w^2 + h^2} \end{cases}. \tag{3}$$

where w and h are the width and height of frame and φ is a specific weighting value. When $f(n_k^i) = 1$, the region b_k^i is judged as the moving target region.

2.2.2 Mapping Moving Object to LED Module

We estimate the position of moving object on the floor to find corresponding LED on the wall, and then control turn on/off using previous our method [7]. Suppose that the camera is setup on the side of wall while LED array is installed on the top. In the real space *OXYZ* with *OX* is width, *OY* is length and *OZ* is height; we have Hr is the height of camera, α is the camera angular aperture, β is the angle between optical center of camera (*Ct*) and *OY* horizontal, and the position of moving object on the floor which is attached *OXY* space, has (*x*, *y*, *0*) coordinates in *OXYZ* space. We intend to find LED on top which corresponds to position of moving object in real space based on the position of moving object on image. To solve that problem, we use inverse perspective transform [8]. Assume that, we receive image has size *(h, w)* with *h* is the height of image; *w* is the width of image; and on image space, moving object has coordinates *(u, v)*. The floor has size (*S, L*) with *S* is the width of floor, and *L* is the length of floor. We calculate coordinates of moving object in real space by inverse perspective transform.

$$x(u,v) = Hr \times cot\left[\left(\beta - \frac{\alpha}{2}\right) + u \times \frac{\propto}{h-1}\right] \times sin\left[v \times \frac{\propto}{w-1} - \frac{\propto}{2}\right] + \frac{S}{2} \tag{4}$$

$$y(u,v) = Hr \times cot\left[\left(\beta - \frac{\alpha}{2}\right) + u \times \frac{\alpha}{h-1}\right] \times cos\left[v \times \frac{\propto}{w-1} - \frac{\propto}{2}\right] \tag{5}$$

2.2.3 LED Brightness Control

The brightness of dimmed LED depends on the brightness of environment when it is day time or night time. Here, we get brightness of received image instead of brightness of environment, and calculate brightness of LED to control dimming based on brightness of received image [9] because the brightness of received image is depended on the brightness of environment. Suppose the max dimming value of LED in night time is Max_{dim}. First we calculate the brightness *I* of received RGB color image

$$I = \frac{1}{2}\left(max\left(\frac{R+G+B}{3}\right) + min\left(\frac{R+G+B}{3}\right)\right) \tag{6}$$

and calculate the dimming level *D* of LED.

$$D = Max_{dim} - I \tag{7}$$

$\max\left(\frac{R+G+B}{3}\right)$, $\min\left(\frac{R+G+B}{3}\right)$ are maximum and minimum value among gray values. Generally, Max_{dim} closes to 255, which corresponds to 8 bits color image. Therefore, the value range of dimming level is from (255 – *I*) to 255, and if *I* is zero, we can dim 255 levels.

LED brightness depends on the duty cycle of pulse width modulation (PWM). Given the period of pulse sequence τ and the lifetime of PWM Tp, a LED brightness *B* can be determined by two parameters.

$$B = \frac{Tp}{\tau} \tag{8}$$

Given an array of $N \times N$ LED modules with brightness B_{ij} and $X \times Y$ transformed position, positions of $(x(u,v), y(u,v))$ should be mapped accurately to specific LED modules. We divide $X \times Y$ positions to $N \times N$ squares that are denoted as $\mathbf{S} = \{S_{ij} | i,j \in \{1,N\}\}$. If a position $(x(u,v), y(u,v))$ of moving object is within a square S_{ij}, a B_{ij} LED for S_{ij} and neighbor LED of 3×3 will turn on to each of weight values.

$$B_{i+k,j+l} = w_{k,l} \times \frac{Tp}{\tau} \times s \text{ for } k,l \in [-1,1], w = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 8 & 2 \\ 1 & 2 & 1 \end{bmatrix} /22 \quad (9)$$

w is 3×3 mask for weight values. s is a scale value for normalization. Figure 3 shows an example of neighbor LEDs for a transformed position $(x(u,v), y(u,v))$. Our experiment used $w = [1]$ for simple implementation.

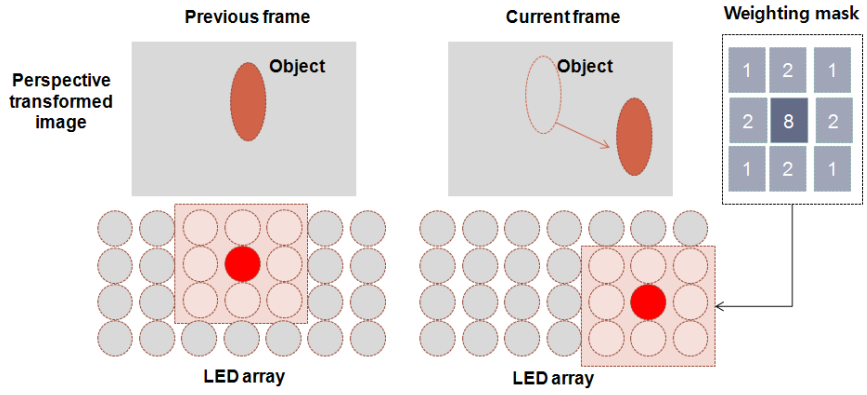
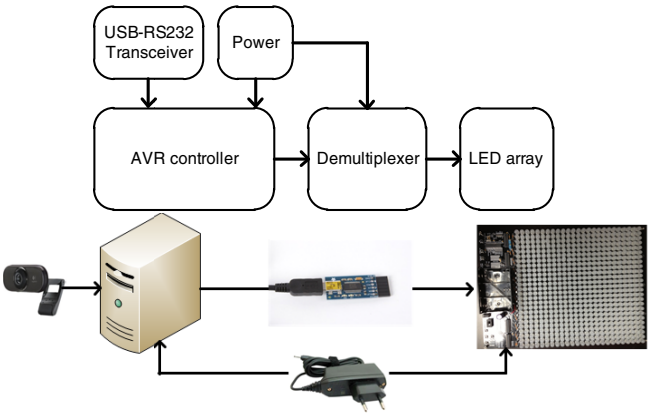


Fig. 3. LED array lighting using 3×3 weight mask

3 Experiment and Results

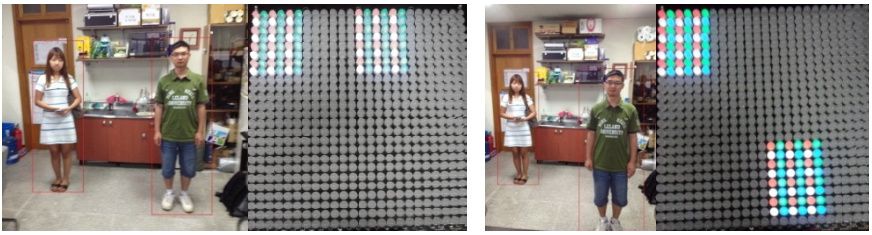
As description in section 2, our system is implemented and experimented on computer and LED board. We designed and implemented LED board as figure 4(a). The hardware structure of LED board includes 5 parts; LED array (RGBW LED), De-muxs (CD74HC154, STP16DP05), AVR micro controller (Atmega328P), Power supply (Battery 5V or Adapter DC 5V), and USB-RS232 transceiver (USB cable, Max232). The connected diagram of system is presented as figure 4(b). Computer is connected to LED board by USB-RS232 transceiver. The experimental environment is the lobby of a building. The camera is installed on the side of wall, and connected to computer.

The system always detect positions of persons in the room to control block LEDs. First time, two persons have positions, which are corresponded to block LED1 (block 1x1) and block LED3 (block 1x3) on figure 5(a). Then a person moves to position to correspond to block LED11 (block 3x3) on figure 5(b).



(a) (b)

Fig. 4. (a) Block diagram of LED board and (b) connected diagram of system



(a) (b)

Fig. 5. Experimental results; (a) two persons start entering central of room and (b) a person moves towards

4 Conclusion

In this paper, we designed, experimented an automatic LED illumination control system using a camera, based on moving object detection and the brightness of received image on LED board. Future, we will implement and experiment with a large capacity LED system, and evaluate the level of energy consumption, and compare with current others.

Acknowledgement. This work was supported by IT/SW Creative research program supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2013-H0502-13-1023), the framework of international cooperation program managed by National Research Foundation of Korea (2012K2A1A2032979) and Brain Busan (BB21).

References

1. Hong, I., Byun, J., Park, S.: Intelligent LED Lighting System with Route Prediction Algorithm for Parking Garage. In: INTELLI 2012: The First International Conference on Intelligent Systems and Applications, pp. 54–59 (2012)
2. Hwang, Z., Uhm, Y., Kim, Y., Kim, G., Park, S.: Development of LED Smart Switch with Light-weight Middleware for Location-aware Services in Smart Home. IEEE Transactions on Consumer Electronics 56(3), 1395–1402 (2010)
3. Hong, I., Byun, J., Park, S.: Design and Implementation of Power-aware LED Light Enabler with Location-aware Adaptive Middleware. IEEE Transactions on Consumer Electronics 56(1), 231–239 (2010)
4. Faheem, I., Siddiqui, A., Im, B.K., Lee, C.: Remote management and control system for LED based Plant Factory using ZigBee and Internet. In: ICACT 2012, February 19-22, pp. 942–946 (2012)
5. Singh, R., Bhardwaj, A., Chauhan, Y., Jain, R.: PWM Strategies in 32-Bit Micro controller for Interior White LED Down Panel. International Journal of Computer Applications 58(22), 25–32 (2012)
6. You, Y., Gong, S., Liu, C.: Adaptive Moving Object Detection Algorithm Based on Background Subtraction and Motion Estimation. IJACT: International Journal of Advancements in Computing Technology, AICIT 5(6), 357–363 (2013)
7. Giao, P.N., Lee, S.-H., Kwon, K.-R.: An Intelligent LED Illumination Control System Using Camera. JCIT: Journal of Convergence Information Technology 8(12) (August 2013)
8. Li, M., Zhao, C., Hou, Y., Ren, M.: A New Lane Line Segmentation and Detection Method based on Inverse Perspective Mapping. JDCTA: International Journal of Digital Content Technology and its Applications, AICIT 5(4), 230–236 (2011)
9. Cho, H., Kwon, O.: A Backlight Dimming Algorithm for Low Power and High Image Quality LCD Applications. IEEE Transactions on Consumer Electronics 55(2), 839–844 (2009)

Audio Recorder Identification Using Reduced Noise Features

Chang-Bae Moon, HyunSoo Kim, and Byeong Man Kim

Department of Computer Software Engineering,
Kumoh National Institute of Technology Daehak-ro 61, Gumi, Korea
moonyeses@naver.com, deltakor@gmail.com, bmkim@kumoh.ac.kr

Abstract. In this paper, we propose an audio recorder identification method as one of digital forensic technologies, where Wiener filter is used to extract noise sounds of recorders and their features are extracted by MIRtoolbox. A recorder identification model is generated by training SVM with the extracted noise features. To improve the identification performance, a new feature reduction method which uses inter-classes standard deviations of features is adopted. The experimental results for 11 audio recorders show 1% improvement over the method with no feature reduction. The improvement is not too noticeable as expected, but the number of features can be reduced up to one third of the method with no feature reduction. The results also show that the proposed feature reduction method is competitive over the other well-known methods such as PCA, LDA and R-squared.

Keywords: Digital Forensic, Wiener Filter, SVM, Feature Reduction, Standard Deviation.

1 Introduction

Thanks to the development of computer and Internet, diverse audio editing programs are easily available to users and audio file can be easily revised by using one of those programs. As the result, an audio file can be forged, thereby causing secondary damage [1].

For audio contents, audio watermarking technologies [2, 3, 4] have been studied to protect copyright or integrity/originality, but in the case that original sound is already sold, it is difficult to insert additional information again. From this point of view, it is more efficient to claim copyright by identifying recorder rather than identifying copyright by using watermark.

Some researchers [1, 5, 6, 7, 8, 9] have studied to identify the equipment producing image or audio, but have been done mainly for image; Moon et al.[1] propose an audio identification method using audio noise features; Kraetzer et al. [5] propose an identification method for peripheral recording equipments and recording environments; Garcia-Romero et al. [6] propose an identification method for Microphones and wired telephones; Choi et al.[7] propose an identification method for printers; and research of [8, 9] propose identification methods for cameras.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_6,

This paper proposes an audio recorder identification method using noise features of recorder as like Moon et al. [1], but 11 recorders are used for experiments and a new feature reduction method is used while the method of Moon et al. uses 5 recorders and no feature reduction method.

2 Related Studies

Choi et al. [7], Matsushita et al. [8] and Lukas et al. [9] suggested the method to identify color printers or camera equipments. and] suggested the method to identify camera equipment. Choi et al. made use of the fact that each printer has slightly different printing method (printer noise) when it print materials. Matsushita et al. and Lukas et al. used characteristics of hardware, that is, CCD elements.

Among the studies to claim copyright of audio data, the representative method is to directly insert additional information of copyright holder into sound source. The methods [2, 3, 4] are ones of those methods. They inserted additional information of producer in the original sound and detect that information when the problem of copyright arises. However, the original sound should be processed to insert additional information. In the case that the original sound is already sold, it is difficult to insert additional information again.

Kraetzer et al. [5] proposed a method to identify microphones and the recording environments of digital audio samples by using audio steganalysis features, where an inter device analysis with different device characteristics was performed while intra device evaluations (identical microphone models of the same manufacturer) were not considered.

Garcia and et al. [6] performed a study on the automatic identification of acquisition devices when only access to the output speech recordings was possible. To alleviate the effects of the speech content variability, they used a statistical characterization of the frequency response of the device contextualized by the speech.

In [1], Moon et al. used the slight different noise features between audio files which is caused by the design and the integrated circuit of each recorder but cannot be identified by the audience. Wiener filter was used to extract noise features of recorders and the recorder was identified by the classification model trained with SVM using the extracted noise features. But they used 5 recorders and no feature reduction method was considered while 11 recorders, in this paper, are used for experiments and a new feature reduction method for improving classification performance.

3 Our Recorder Identification Method

The proposed identification method consists of 2 phases; the phase 1 is for training and phase 2 for identification. In the training phase, the SVM (Support Vector Machine) model is trained with the reduced noise features of recorders and their class information together and, in identification phase, recorders are identified by the classification model generated through the phase 1.

In training or testing phase, feature extraction process is very important and its quality affects the overall performance of the proposed method. The extraction process consists of 3 steps; (1) the noise sound is extracted by Winner filter from a recorded sound; (2) its noise features are extracted by MIRtoolbox [10] and are normalized; (3) noise features are reduced by the method in Section 3.3 called inter-classes standard deviation method.

3.1 Extracting Noise Sound with Wiener Filter

To extract the noise sound, the sound that noises are removed from the original sound should be acquired first and then this sound should be removed from the original sound. In this paper, Wiener filter provided by [13] is used to remove noises.

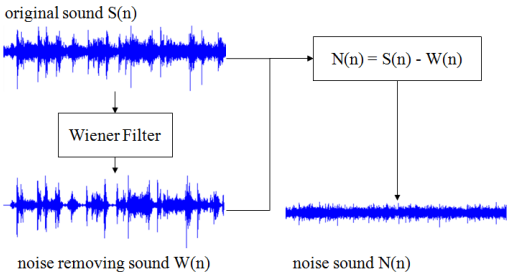
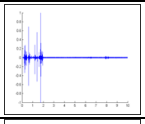
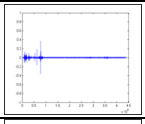
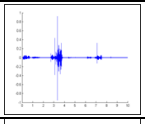
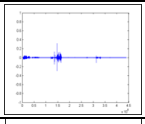
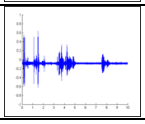
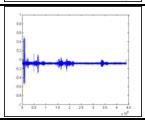
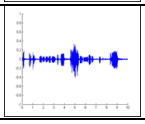
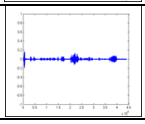


Fig. 1. Process to extract noise sound by using Wiener filter

The examples of noise sounds acquired from some equipments by using Wiener filter are shown in Table 1: The first column of table shows original signal sound; the second column shows noise sound.

Table 1. Examples of Original sound and Noise sound

Company (model)	Original Sound $S(n)$	Noise Sound $N(n)$	Company (model)	Original Sound $S(n)$	Noise Sound $N(n)$
COWON (cws-cown-d2)			IRIVER (IREI-B30)		
MOBIBLUE (DHA-1700)			SAMSUNG (YP-Q2AB)		

3.2 Extracting Features of Noise

In this paper, the 391 features extracted from MIR toolbox of Lartillot[10] are used. When features are extracted by using MIR toolbox, NaNs are occurred. As NaN is the value that cannot be expressed in number, the feature with at least one NaN is

removed and a feature is also removed when the feature value is same to all sound. The final number of features used in our experiments is 347. Table 2 is an example of features extracted by using MIR toolbox.

Table 2. Examples of features acquired from MIR toolbox

Company (model)	Dynamics/ Rms			…	Tonal/ HCDF		
	Mean	Std	Slope		Period Freq	Period Amp	Period Entropy
COWON (cws-cown-d2)	0.981	0.183	0.624	…	11.41	0.893	0.9722
MOBIBLUE (DHA-1700)	0.968	0.249	0.478		26.63	0.903	0.9714
…	…		…			…	
SAMSUNG (YP-U5AWH)	0.962	0.271	0.468	…	26.63	0.904	0.9722
SAMSUNG (YP-U5QR)	0.924	0.381	0.297		13.31	0.884	0.9727

The maximum value and minimum value are different from feature to feature. So, in this paper, a feature value is normalized between -1 and +1.

3.3 Feature Reduction

If 347 features acquired by using MIR toolbox are wholly used, negative effect can be caused by noise features. Thus, we use the feature reduction method that uses the inter-class standard deviation as a selection criterion.

The feature reduction method is composed of 3 steps as mentioned previously. In the first step, the mean point of each class for each feature which is calculated by averaging values of the relevant feature of data in each class is acquired. Namely, the mean point is calculated by Equation (1). The mean points of classes for each feature are illustrated in Fig. 2(a) where the dots indicated on the bars are mean points.

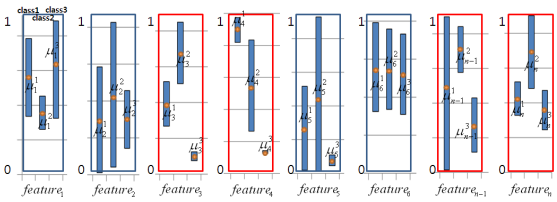
$$\mu_n^c = \frac{1}{k}\sum_{i=1}^{k} nf_{c,n}^i \tag{1}$$

where, k is the number of data in c'th class and μ_n^c is the mean value of c'th class of n'th feature.

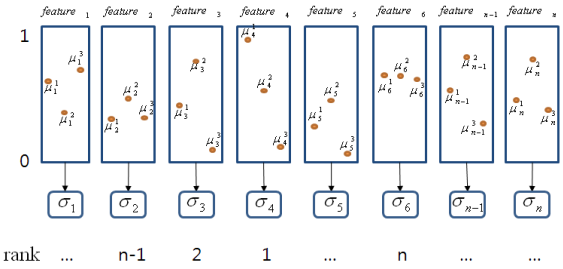
In the second step, standard deviation of mean points for each feature is calculated and used to select features having good discrimination power. Namely, let the number of class be m and μ_j^1, μ_j^2, … and μ_j^m be the mean point of the j'th feature for each class respectively. Then, their standard deviation is acquired by Equation (2) (see Fig. 2(b)).

$$\sigma_j = \sqrt{\frac{1}{m}\sum_{c=1}^{m}(\mu_j^c - \overline{\mu_j})} \tag{2}$$

where, $\overline{\mu_j}$ is the average of μ_j^1, μ_j^2, …, and μ_j^m.



(a) The mean point of each class for each feature



(b) Standard deviation of mean points for each feature

Fig. 2. Illustration of the proposed reduction process

In the last step, after getting the standard deviations for all features, features are ranked by their standard deviation and features whose rank is over certain level are finally selected. Higher standard deviation, higher rank. The proposed feature reduction method is based on the assumption that the inter-class standard deviation of a good feature is higher than a bad feature.

4 Experiments

To measure the performance, in this paper, we use LIBSVM[11, 12] among support vector machine for identifying multi classes. To collect evaluation data, 11 equipments of 5 companies are used in 3 regions: a house, street and a fast food restaurant. Each recording time is over 30 minutes and the recorded data are segmented with the interval of 10s. Finally, we have acquired total of 5,934 samples (see Table 3).

To evaluate our approach, 5-fold cross validation is used, in which the data is split into 5 equal-sized groups and in 20 runs, each group is once selected as a test set, while other 4 groups are used for training. The final result is averaged for 20 runs. Table 4 shows the best identification performances of the proposed method, R-squared method, LDA, and PCA, where “All features” means no feature reduction method is applied. Experiments are performed for 4 kernel functions of SVM: linear, non-linear, radial basis, and sigmoid.

Table 3. Number of Sample in Experiments ((a) : Home, (b) : fast food restaurant, (c) Road)

Company	Model	(a)	(b)	(c)	Total
COWON	CWS-COWN-D2	180	180	180	540
MOBI BLUE	DAH-1700	180	180	180	540
PHILIPS	GA30006-0026USA	179	179	179	537
IRIVER	IREI-B30	180	180	180	540
	IREI-E150	180	180	180	540
SAMSUNG	YP-Q2AB	179	179	179	537
	YP-R1AB	180	180	180	540
	YP-T7F	180	180	180	540
	YP-U5AW	180	180	180	540
	YP-U5AWH	180	180	180	540
	YP-U5QR	180	180	180	540
Total number of sample					5,934

As shown in Table 4, the performance becomes worse when LDA or PCA is used, whereas the performance becomes slightly better when R-squared method or the proposed method is used.

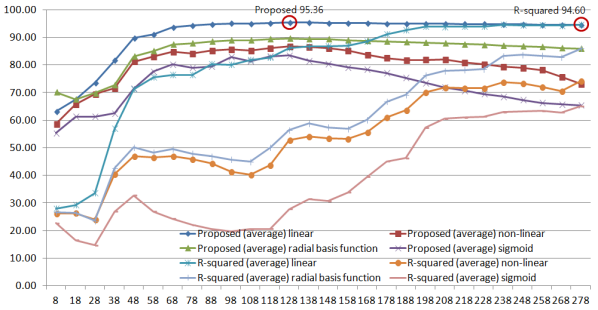
Table 4. Identification performance (n-fold (n=5), average of 20 runs)

Feature Reduction	linear	non-linear	radial basis function	sigmoid
Proposed (128)	95.36 %	86.73 %	89.59 %	83.44 %
R-squared (278)	94.60 %	74.20 %	85.92 %	65.02 %
LDA	67.78 %	14.62 %	58.79 %	52.72 %
PCA	67.94 %	8.34 %	60.20 %	54.96 %
All features	94.58 %	72.91 %	85.61 %	64.42 %

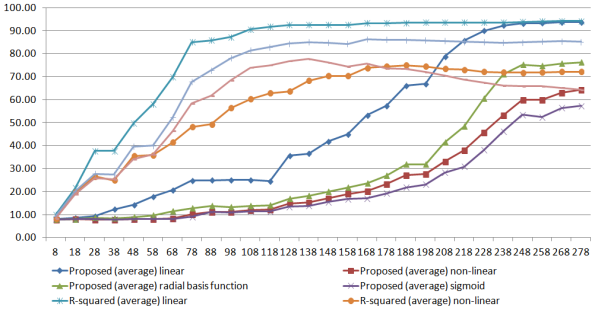
To show the usefulness of the proposed feature reduction method, two types of experiments are further taken; 10 features with good discriminating power are added incrementally from 8 in the first type of experiments and 10 features with bad discriminating power in the second type.

Fig. 3(a) shows the results of the first type of experiments. The performance improvement of the proposed method over the R-squared method is insignificant from the viewpoint of accuracy, but significant when considering the number of features; the best performance (95.36%) of the proposed method is achieved when using linear kernel and 128 features while the best performance (94.60%) of the R-squared method when using linear kernel and 278 features.

A shown in Fig. 3(b) which shows the results of the second type of experiments, the performance of the R-squared feature reduction method is sharply increased in the range of the number of features 38~78 while the performance of the proposed method in the range 168~228. This means good features for identification of the proposed method might be located in the higher rank than the R-squared method and so the proposed method is better than R-squared method in selecting good features for audio recorder identification.



(a) Identification performance of Type I



(b) Identification performance of Type II

Fig. 3. Identification performance (proposed vs. R-squared)

5 Conclusion

In this paper, we proposed audio recorder identification method as a digital forensic technique on audio data. For verification of integrity and originality of digital audio data, the audio watermarking technology can be a good choice. However, the original sound should be processed to insert additional information and so it is difficult to insert additional information again for the sounds already sold. In this respective, the proposed method is more comprehensive than the watermarking technology.

Also, in this paper, to further improve the identification performance, a new feature reduction method which uses inter-classes standard deviations of features is proposed. The experimental results for 11 audio recorders show 1% improvement over the method with no feature reduction. The improvement is not too noticeable as expected, but the number of features can be reduced up to one third of the method with no feature reduction. The results also show that the proposed feature reduction method is competitive over the other well-known methods such as PCA, LDA and R-squared.

Currently, only different audio recorder models are considered. For the proposed method to be more practical, identical audio recorder models should be considered. The proposed method also does not guarantee the reliability of digital audio data. For

reliability verification, forged audio data should be detected. In the next study, we will focus on this topic.

Acknowledgements. This paper was supported by Research Fund, Kumoh National Institute of Technology.

References

1. Moon, C.B., Kim, H.S., Yi, M.H., Kim, B.M.: Experiments on Audio Recorder Identification. In: Proceeding of the 3rd International Workshop on Ubiquitous Computing & Applications, Hong Kong, p. 21 (2012)
2. Cveji, N.: Algorithms for Audio Watermarking and Steganography, Oulun yliopisto (2004)
3. Dutta, P., Bhattacharyya, D., Kim, T.H.: Data Hiding in Audio Signal: A Review. International Journal of Database Theory and Application 2(2), 1–8 (2009)
4. Ru, X.M., Zhang, H.J., Huang, X.: Steganalysis of audio: Attacking the steghide. In: Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, vol. 7, pp. 3937–3942 (2005)
5. Kraetzer, C., Oermann, A., Dittmann, J., Lang, A.: Digital audio forensics: a first practical evaluation on microphone and environment classification. In: Proceedings of the 9th Workshop on Multimedia & Security, pp. 63–74. ACM (2007)
6. Garcia-Romero, D., Espy-Wilson, C.Y.: Automatic acquisition device identification from speech recordings. In: IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), Dallas, pp. 1806–1809 (2010)
7. Choi, J.H., Im, D.H., Lee, H.Y., Oh, J.T., Ryu, J.H., Lee, H.K.: Color laser printer identification by analyzing statistical features on discrete wavelet transform. In: 16th IEEE International Conference on Image Processing (ICIP), pp. 1505–1508. IEEE (2009)
8. Matsushita, K., Kitazawa, H.: An improved camera identification method based on the texture complexity and the image restoration. In: Proceedings of the 2009 International Conference on Hybrid Information Technology, pp. 171–175. ACM (2009)
9. Lukas, J., Fridrich, J., Goljan, M.: Digital camera identification from sensor pattern noise. IEEE Transactions on Information Forensics and Security, 205–214 (2006)
10. Lartillot, O., Toiviainen, P.: A Matlab toolbox for musical feature extraction from audio. In: International Conference on Digital Audio Effects, pp. 237–244 (2007)
11. Ryu, S.-J., Lee, H.-Y., Cho, I.-W., Lee, H.-K.: Document forgery detection with svm classifier and image quality measures. In: Huang, Y.-M.R., Xu, C., Cheng, K.-S., Yang, J.-F.K., Swamy, M.N.S., Li, S., Ding, J.-W. (eds.) PCM 2008. LNCS, vol. 5353, pp. 486–495. Springer, Heidelberg (2008)
12. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2(3), 27 (2011)
13. Wiener filter, `http://www.mathworks.com/matlabcentral/fileexchange/7673-wiener-filter`

Performance Evaluation of a Generalized Music Mood Classification Model

Min Kyun Song, HyunSoo Kim, Chang-Bae Moon, and Byeong Man Kim

Department of Computer Software Engineering,
Kumoh National Institute of Technology Daehak-ro 61, Gumi, Korea
{smkasd,moonyeses}@naver.com, deltakor@gmail.com,
bmkim@kumoh.ac.kr

Abstract. This paper focuses on building a generalized mood classification model with many mood classes instead of a personalized one with few mood classes. Two methods are adopted to improve the performance of mood classification. The one of them is a new feature reduction based on standard deviation of feature values, which is designed to solve the problem of lowered performance when all 391 features provided by MIR toolbox used to extract features of music. Our experiments show that the feature reduction method suggested in this paper has better performance than that of the conventional dimension reduction methods, R-Squared and PCA. As performance improvement by feature reduction only is limited, the modular neural network approach is used additionally. The experiments also show that the method improves the performance effectively.

Keywords: Mood Classification, Feature Dimension Reduction, Modular Neural Network.

1 Introduction

A lot of music has been released these days and Internet enables users to access such music more easily and faster. Due to the increase in music contents, the classification system of music is required to easily find and manage music. Music is traditionally classified by use of metadata information inputted by users or system administrators. However, such an approach cannot properly classify if relevant music information is insufficiently provided, thereby causing wrong result when users search. The manual approach is also quite burdensome to apply it to vast quantity of music. For these reasons, a technology which can directly extract or induce genre or mood from music is required.

Some studies [1-4] have tried to detect music mood using acoustic features like tempo, timbre, pitch and so forth. However, they have focused on personalization for each user rather than generalization or have used few mood classes. We, in this paper, focus on building a generalized mood classification model with many mood classes instead of a personalized one with few mood classes.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_7,

First, to construct a generalized mood classification model, the mood information of music segments are collected from about 200 subjects. Then, the representative mood of a music segment is defined based on the definition in [16]. Finally, a general mood classification model is built up via modular neural network by using the representative mood of music segments and their acoustic features.

We use a new feature reduction method based on the standard deviation of feature values to improve the performance of mood classification, which is designed to solve theproblem of lowered performance when all 391 acoustic features provided by MIR toolbox[5] are used. Our experiments show that the feature reduction method has better performance than that of the conventional dimension reduction methods, R-Square and PCA[6]. As performance improvement by feature reduction only is limited, modular neural network approach is used additionally. The experiments also show that the modular approach improves the performance effectively.

2 Related Studies

The existing music mood models include Russel Model[7] , Hevner Model[8] and Thayer Model [9]. Since both the Russell and Hevner models use adjectives to describe emotions, ambiguity arises if adjectives have multiple meanings. Thus, this paper uses the 2 dimensional mood model of Thayer where music mood is expressed by vector value composed of arousal and valence; arousal indicates the strength of stimulation felt by listener in music; valence indicates stability of sound. Fig. 1 shows the relationships among 12 adjectives of mood and emotion in Thayer's two-dimensional model.

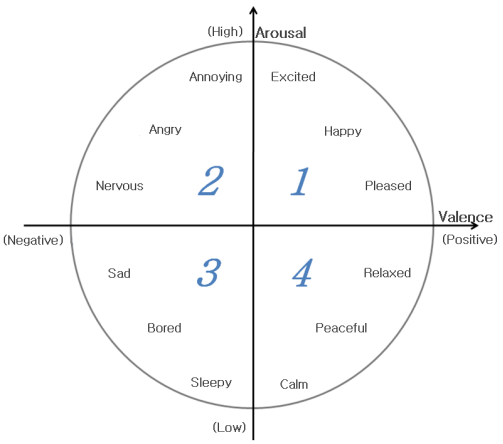


Fig. 1. Two-Dimensional Thayer Model

Feng et. al. [10] propose the method to classify mood into 4 classes, e.g., happiness, sadness, anger, fear by using tempo and articulation feature. Li and Ogihara[11] propose a method to detect mood by using features like timbre texture,

rhythm, pitch, etc. In this method, 13 adjective groups made by Farnsworth[12] by reorganizing the Hevner's inspection list[13] are used as mood classes. Yang et al[4] use a fuzzy-based method to solve the ambiguity of expression happening when only a single mood is allowed, where they express the mood of music with the strength of several moods in number. However, they point out that the fuzzy-based method may fail to deal with the subjective preference of individuals in personalization service[14, 15]. So, they use AV (Arousal and Valence) coefficient composed of two real number values between -1～1 of each axis of the two-dimensional mood model of Thayer instead of a single mood class. They use two regressors for modeling AV coefficients collected from users. To get AV coefficients, the data are collected as the subjects directly input AV values for each music piece. They also suggest the personalized detection method considering user's similar groups (professional/non-professional depending on the degree of understanding of music).

3 Construction of a Generalized Music Mood Classification Model

Music mood classification structure in this paper is shown in Fig. 2, where music segments are used for experiments instead of full music. For music segments in a training set, their acoustic features are extracted and reduced by the proposed method and the reduced features are finally used as the input of modular neural network. From AV coefficients of a segment collected from the subjects, the representative AV coefficient is defined and used as the target of modular neural network. Using the classification model learned via modular neural network, the mood of a new music segment is predicted with its reduced features.

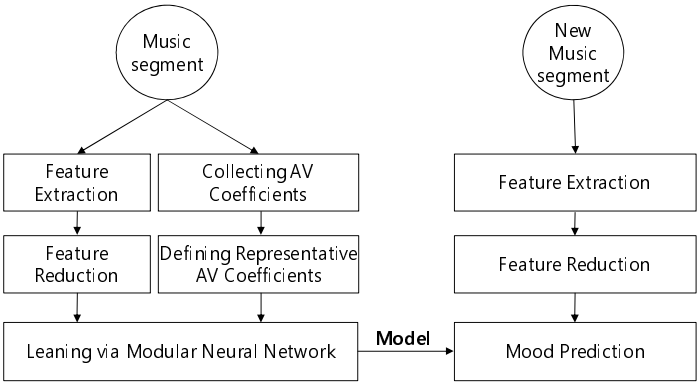


Fig. 2. Structure of mood classification

3.1 Mood Data of Music Segments

We use music segments used in [16]. The data were collected using 202 subjects who were asked to identify the moods evoked by music segments. A total of 281 music

segments were extracted from 101 songs; 47 of these segments were randomly selected and provided to the subjects; the subjects gave mood information for 44 music segments on average. Subjects were asked to rate a segment for several moods with totally maximum 5 points.

Table 1. Mood Data of Music Segments

No.	Annoying	Angry	Nervous	Sad	Bored	Sleepy	Calm	Peaceful	Relaxed	Pleased	Happy	Excited
1								2	3			
2							2	2			1	
…												
8895	1										1	3
8896								1			3	1

Table 1 shows the mood data of the music segments. The values of each line in Table 1 are the scores entered by one subject. Subjects might assign different moods to the same music segment, but it is necessary to define one representative mood per segment; during training phase, only the representative mood of each music segment is used. The representative mood of each segment is calculated based on the definition in [16]. Table 2 shows the number of music segments for each representative mood.

Table 2. Number of music segments for each representative mood

mood	number of Data	mood	number of Data
Angry	10	Nervous	7
Annoying	48	Peaceful	38
Bored	12	Pleased	5
Calm	51	Relaxed	18
Excited	30	Sad	10
Happy	32	Sleepy	20

3.2 Feature Extraction and Reduction

In this paper, 391 features of a music segment are extracted by MIRtoolbox[6]. However, NaNs — values that cannot be expressed numerically— may occur; 347 features are obtained by eliminating features including at least one NaN. According to the prior test, if the 347 features acquired by using MIR toolbox are wholly used, the negative effect in performance is caused by noise features. Thus, we use the feature reduction method proposed in this paper to remove some noise features.

The feature reduction method uses the basic standard deviation, which is composed of 3 processes: normalization, calculation of the mean point and selection of features. Through the process of normalization, feature values are converted into values in the range of 0～1 because each feature has different maximum value. After normalization, it is required to get the mean point of each class for each feature, which is acquired by averaging values of the relevant feature of data belonging to each class. Namely, it is calculated by Equation (1).

$$\mu_n^c = \frac{1}{k}\sum_{i=1}^{k} \tilde{d}_{i,n}^c \tag{1}$$

where, k means the number of data belonging to class c and μ_n^c means the average of their normalized n'th feature values.

After getting the mean points for each feature, their standard deviation is calculated and used to select features having good discrimination power. Let the number of class be m and μ_j^1, μ_j^2, … and μ_j^m be the mean point of the j'th feature for each class respectively. Then, their standard deviation is acquired by Equation (2).

$$\sigma_j = \sqrt{\frac{1}{m}\sum_{c=1}^{m} (\mu^c{}_j - \overline{\mu}_j)^2} \tag{2}$$

where, $\overline{\mu}_j$ is the average of μ_j^1, μ_j^2, …, and μ_j^m.

Finally, the features are ranked by their standard deviation and then the features whose rank is over certain level are selected.

3.3 Mood Learning Using Modular Neural Network

The structure of neural network built in this paper is shown in Fig. 3. Input layer uses features obtained by applying the feature reduction method in 3.2. The number of features used in our experiments is preset to 50 from the prior experiments. The module layer is composed of the hidden layer of each module so that the number of hidden nodes can be controlled. There are 2 nodes in output layer for each module.

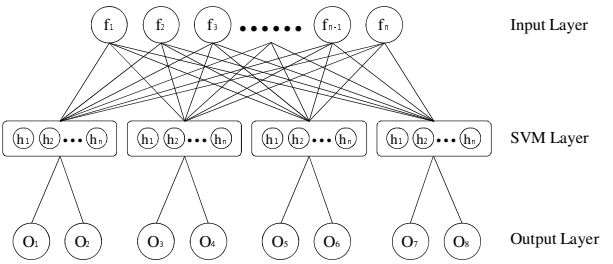


Fig. 3. Modular neural network structure

In designing the structure of MNN (Modular Neural Network), we intend to make each module cover two opposite moods. So the network should be composed of 6 modules to include 12 moods, but there are 4 modules in Fig. 3. Namely, 4 moods are excluded. This is why the number of data used in this paper is not enough for the 4 moods - Angry, Nervous, Pleased and Sad (See Table 2). Therefore, in our experiments, a module is constructed to cover two moods being most dissimilar

among the remaining 8 moods: the first module for Annoying & Peaceful, the second module for Bored & Relaxed, the third module for Calm & Excited, and the last module for Happy & Sleepy. During the test phase, given the feature vector of a new music segment to the input layer, the class that the new segment belongs is determined by $argmax(o_1, o_2, o_3, o_4, o_5, o_6, o_7, o_8)$. For instance, if o_1 is the largest value among $o_1, o_2, o_3, o_4, o_5, o_6, o_7$ and o_8, the mood is Annoying and if o_2 is the largest value, the mood is Peaceful and so forth.

4 Experiments

In this paper, the number of hidden layer is changed to 5, 10, 20, 30, 40, 50 and 100 for test, the Leave-one-out Cross-validation method is used for validation and the model is learned for 10,000 iterations. To show the effectiveness of the feature reduction methods proposed in this paper, the performance of our methods is compared with the performance of the feature reduction methods, R-Square and PCA. The performance of non-modular neural network is also compared with the performance of MNN for showing the effectiveness of MNN.

The performance of the feature reduction methods proposed in this paper is measured for non-modular neural network. As shown in Fig. 4, the best performance is 21% when features are not reduced and only features with any "NaN" value are removed. When the feature reduction method based on the basic standard deviation is applied, however, the best performance of 39% is achieved when 5 hidden nodes are used. The performance is improved by 18% over the overall features.

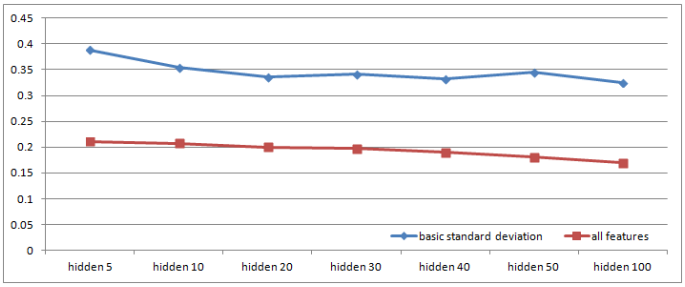


Fig. 4. Performance of suggested feature reduction methods

The performance of MNN over non-modular neural network is shown in Fig. 5, where the feature reduction method based on the basic standard deviation is used. The difference in performance is 4% when 5 hidden nodes are used, 11% when 10 hidden nodes are used, 20% when 20 hidden nodes are used, 19% when 30 hidden nodes are used, 21% when 40 hidden nodes are used, 17% when 50 hidden nodes are used and 19% when 100 hidden nodes are used. In average, the performance of MNN is improved by 16% over the non-modular neural network.

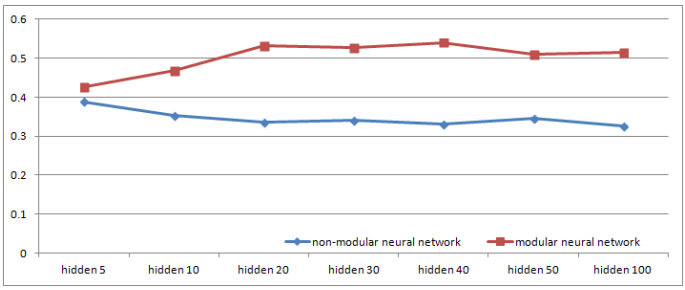


Fig. 5. Comparison of performance between general neural network and MNN

The performance when our feature reduction method and MNN are combined is shown in Fig. 6. For comparison, the performance when the existing two representative feature reduction methods, R-Square and PCA, replace our reduction methods is measured. For convenience, the number of features used in learning is fixed to 50. The result shows the performance is 54% with 40 hidden nodes for the basic standard deviation. We also achieve the performance of 23% when all features are used, 53% when PCA is used and 43% when R-Square is used.

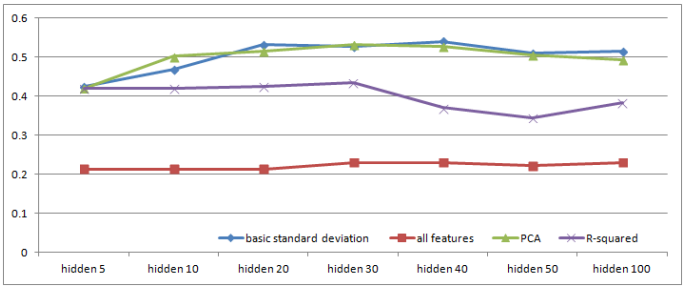


Fig. 6. Performance when feature reduction method and MNN are combined

5 Conclusion

To improve mood classification performance, a feature reduction method based on the basic standard deviation is suggested. The experiments show the suggested feature reduction method has better performance than that of the conventional feature reduction methods, R-Square and PCA. Also, the experiments show that the modular neural network approach is effetive in improving the performance.

The more detailed study on the representative mood selection, feature reduction and learning method is required to improve classification performance. Further, more general learning and classification of music mood will be possible if mood data on music is collected from more people for more music.

Acknowledgements. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2010-0021097).

References

1. Lee, J.I., Yeo, D.G., Kim, B.M., Lee, H.Y.: Automatic Music Mood Detection through Musical Structure Analysis. In: International Conference on Computer Science and its Application, pp. 510–515 (2009)
2. Lu, L., Liu, D., Zhang, H.J.: Automatic mood detection and tracking of music audio signals. IEEE Transactions on Audio, Speech, and Language Processing 14(1), 5–18 (2006)
3. Yang, Y.H., Liu, C.C., Chen, H.H.: Music Emotion Classification: A Fuzzy Approach. In: Proceedings of the 14th Annual ACM International Conference on Multimedia, pp. 81–84 (2006)
4. Singh, P., Kappor, A., Kaushik, V., Maringanti, H.B.: Architecture for Automated Tagging and Clustering of Song Files. International Journal of Computer Science Issues 7(4), 11–17 (2010)
5. Lartillot, O., Toiviainen, P.: A Matlab toolbox for musical feature extraction from audio. In: International Conference on Digital Audio Effects, pp. 237–244 (2007)
6. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer (2006)
7. Russell, J.A.: A Circumplex model of affect. Journal of Personality and Social Psychology 39(6), 1161–1178 (1980)
8. Hevner, K.: Experimental studies of the elements of expression in music. The American Journal of Psychology 48(2), 246–268 (1936)
9. Thayer, R.E.: The Biopsychology of Mood and Arousal. Oxford University Press (1989)
10. Feng, Y., Zhuang, Y., Pan, Y.: Popular music retrieval by detecting mood. In: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 375–376 (2003)
11. Li, T., Ogihara, M.: Detecting emotion in Music. In: Proceedings of the International Symposium on Music Information Retrieval, pp. 239–240 (2003)
12. Farnsworth, P.R.: The social psychology of music. The Dryden Press (1958)
13. Hevner, K.: Expression in music: a discussion of experimental studies and theories. Psychological Review 42(2), 186–204 (1935)
14. Yang, Y.H., Su, Y.F., Lin, Y.C., Chen, H.H.: Music emotion recognition: the role of individuality. In: Proceedings of the International Workshop on Human-Centered Multimedia, pp. 13–21 (2007)
15. Yang, Y.H., Lin, Y.C., Su, Y.F., Chen, H.H.: A regression approach to music emotion recognition. IEEE Transactions on Audio, Speech, and Language Processing 16(2), 448–457 (2008)
16. Moon, C.B., Kim, H., Lee, H.A., Kim, B.M.: Analysis of relationships between mood and color for different musical preferences. Color Research & Application (online published 2013)

Reliable Transmission for Remote Device Management (RDM) Protocol in Lighting Control Networks

Sang-Il Choi[1], Sanghun Lee[1], Seok-Joo Koh[1], Sang-Kyu Lim[2], Insu Kim[2], and Tae-Gyu Kang[2]

[1] Kyungpook National University, South Korea
[2] Electronics Telecommunications Research Institute, South Korea
{overcycos,lah1290}@gmail.com, sjkoh@knu.ac.kr,
{sklim,iskim,tgkang}@etri.re.kr

Abstract. The Remote Device Management (RDM) protocol is used to manage lighting devices in the lighting control network. RDM provides bi-directional communication between lighting devices based on the modified DMX512 data link so as to perform device discovery, configuration, monitoring, and management of the devices connected through the DMX512 network. However, the RDM protocol does not provide a reliable packet transmission. To overcome this limitation, we propose the two retransmission schemes: sequence-based retransmission, and timer-based retransmission. In the existing RDM protocol, each response packet uses the same Transaction Number (TN) with that of the request packet from the controller. In the proposed schemes, we use TN to identify a lost or error message. For example, if a response packet with a specific TN number does not arrive within a pre-specified time, the controller will retransmit the concerned request message. From the numerical analysis, it is shown that the proposed retransmission schemes can give better reliability than the existing RDM protocol, and the sequence-based and timer-based retransmission schemes provide the different performance, depending on network environments.

Keywords: RDM, Lighting Devices, Retransmission, Reliability.

1 Introduction

One of the primary issues on the lighting control system is how to effectively control a lot of lighting devices in the network [1]. Some protocols for control of lighting devices have so far been made in the PLASA Technical Standards Program (TSP) [2], which include the Digital Multiplex 512-A (DMX512-A) and the Remote Device Management (RDM) [3, 4].

The RDM protocol provides the device discovery, configuration, monitoring, and management of console or other controlling devices connected through a DMX512 network. The RDM protocol is based on the pulling system. That is, only the controller can first initiate communication, and the devices will respond to the request of controller. However, in the RDM protocol, there is no retransmission mechanism. So, any error or loss of packet will not be recovered. From this reason, if the status of

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_8,

the DMX512 network becomes worse with a large packet error rate, the controller cannot manage the lighting devices effectively due to the error or loss of the request or response packets.

To address this limitation, in this paper, we propose the two retransmission schemes for RDM protocol. The proposed retransmission schemes can be used to effectively provide reliable packet transmission in the RDM protocol, compared to the existing RDM protocol, as done in the Transmission Control Protocol [5].

This paper is organized as follows. Section 2 describes the data transmission mechanisms of the existing RDM protocol. In Section 3, we describe the proposed two retransmission schemes for RDM. Section 4 analyzes and compares the existing and proposed schemes in terms of the reliability and performance. Section 5 concludes this paper.

2 Packet Transmission Mechanisms in RDM

In RDM, each communication is only performed between controller and devices, and it is assumed that there is only one controller in the network and all devices are managed by the controller. Fig. 1 shows the network model of the RDM protocol in the lighting control networks.

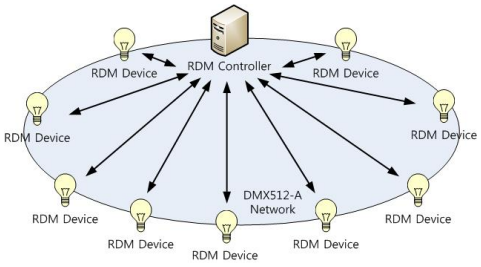


Fig. 1. Network model of RDM in lighting control networks

In the existing RDM protocol, there is no packet retransmission process. Fig. 2 shows the packet transmission mechanism of the existing RDM Protocol.

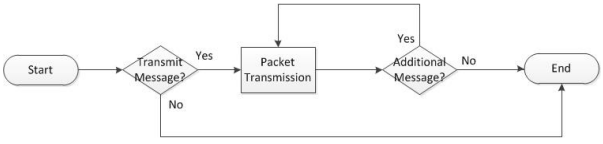


Fig. 2. Packet transmission of the existing RDM protocol

First, if the controller has a message to send, it makes a request packet including the associated control information and sends it to the device over the DMX512 data link. Then, the controller waits for the response packet from device for a specific time

(e.g., 2.1mS). If there is no response message in a specific time, the controller will decide that the response packet is lost and may send the request packet again. However, if the response packet arrives from devices, the controller just checks the Transaction Number (TN) field of the response packet to map between a request packet and a response packet and the checksum field to determine whether the packet has any error. After the reception of the response packet is completed, the controller resumes sending the request packets. When the controller has no additional request packet to send, the data transmission process is completed. As shown in Fig. 2, the existing RDM protocol has no consideration for recovery of the lost or error packets. Thus, if there are some errors in the response packet, the controller will just check an error using the checksum field. The controller does not attempt to recover the errors. Fig. 3 summarizes the data transmission procedures of the existing RDM scheme.

```
1. Sends the Request packet to Device or Devices
2. Waits response packet for Device or Devices
2.1 Response packet is received, Controller process them
2.2 Response packet is not arrived, Controller does not work
3. Prepares next Request packet
3.1 Has next Request packet, Sends that packet
3.2 No next Request packet, Controller does not work
```

Fig. 3. Procedure for data transmission of existing RDM scheme

In addition, in the RDM protocol, there is a negative acknowledgement (NACK) packet to indicate that the responder is unable to reply with the requested command, and this NACK packet includes a reason for the error. However, after receiving the NACK packet, there is no retransmission process to recover the error packet. The NACK-based retransmission scheme may be considered for the RDM protocol. However, we do not discuss this scheme in this paper, because this scheme is only performed at a packet format error condition and cannot handle the packet loss condition.

3 Proposed Retransmission Schemes

To provide the reliable packet transmission for RDM, we propose the two retransmission schemes: sequence-based and timer-based schemes. For each scheme, we use the TN field of RDM packet. In addition, we define a buffer in the controller for temporarily storing packets for retransmission in the event of packet loss or error.

3.1 Sequence-Based Retransmission

In the sequence-based retransmission scheme, only the TN of a response packet is used. In this case, the packet retransmission is only performed when the order of TN contained in the response packet indicates a loss or error. With this concept, after only the further subsequent response packet arrives at the controller, the retransmission of previous packet can begin. Fig. 4 shows the packet retransmission mechanisms of the sequence-based retransmission scheme.

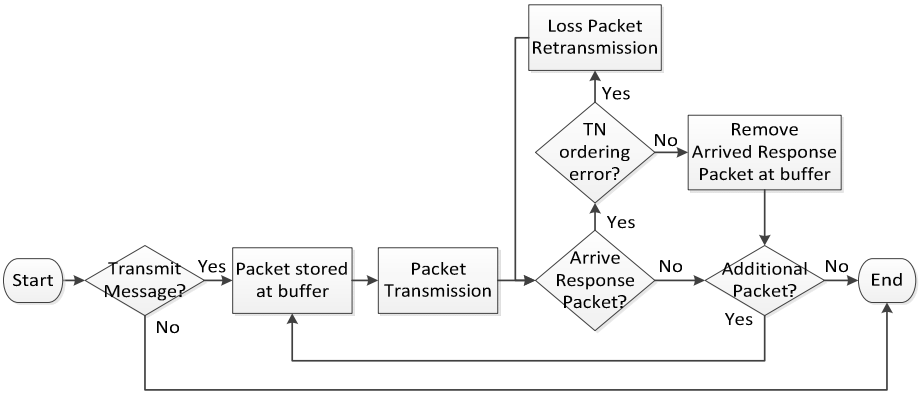


Fig. 4. Sequence-based retransmission scheme

First, if the controller has a message to send, it makes a request packet including the control information, temporarily stores the packet in the buffer, and then it sends the request packet over the DMX512 data link. Then, the controller waits for the response packet from the device. It is expected that the response packet has the same TN with that of the request packet that the controller sent. In this scheme, we focus on the loss of request or response packet, so we do not consider the reception of corrupted packet. If the response packet does not arrive, the controller waits for a specific time (2.1mS) and checks additional messages for device. On the other hand, if a response packet arrives from the device, the controller checks the order of TN for the response packet. After checking the order of TN, if the associated TN is different from the expected value of sequence counter at controller, it performs the retransmission of the lost request packet. On the other hand, in the successful reception case, the controller will remove the packet stored in the buffer. When the controller has no additional request packet to send, the data transmission process is completed.

Fig. 5 summarizes the data transmission procedures for the sequence-based retransmission scheme.

```
1. Stores the Request packet at the Buffer
2. Sends the Request packet to Device or Devices
3. Waits Response packet for Device or Devices
3.1. Response packet is received
3.1.1. Processing the Response packet
3.1.1.1. Transaction Number of Response packet is out of sequence
3.1.1.1.1 Resends the Request packet that has missing Transaction Number
3.1.1.2. Transaction Number is ordering
3.1.1.2.1. Removes stored packet at the Buffer
3.2. Response packet is not arrived
3.2.1. Controller does not work
4. Prepares next Request packet
4.1. Has next Request packet
4.1.1. Sends that packet
4.2. No next Request packet
4.2.1. Controller does not work
```

Fig. 5. Procedure of sequence-based retransmission scheme

3.2 Timer-Based Retransmission

In the timer-based retransmission scheme, we use the TN of the response packet with temporary buffering. In addition, we define a retransmission timer for retransmission of each lost packet at the controller. This retransmission scheme also focuses on the packet loss rather than the packet error, as done in the sequence-based retransmission scheme. So, the packet retransmission is only performed when the retransmission timer is expired by a packet loss. Fig. 6 shows the packet transmission flow in the timer-based retransmission scheme.

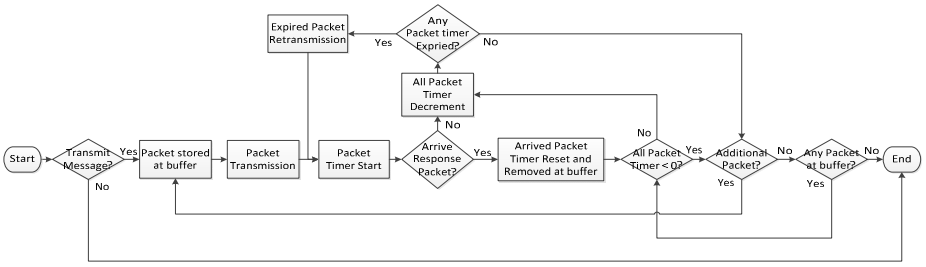


Fig. 6. Packet transmission flow in the timer-based retransmission scheme

If the controller has a message to send, it stores the packet in the buffer, and sends it over the DMX512 data link. Then, the controller activates the retransmission timer for the transmitted request packet and waits for the response packet. If the associated response packet arrives before the retransmission timer is expired, the controller will process the response message successfully. Otherwise, if the response packet does not arrive with the retransmission time interval, the controller confirms that the request packet is lost for any reason, and the request packet will be retransmitted by the controller. Fig. 7 summarizes the data transmission procedure in the timer-based retransmission scheme.

```
1. Stores the Request packet at the Buffer
1.1. Timer value of Request packet is set to specific value
2. Sends the Request packet to Device or Devices
3. Waits Response packet for Device or Devices
3.1. Response packet is received
3.1.1. Processing the Response packet
3.1.1.1. Initiates the Timer Value of Request packet
3.1.1.2. Removes stored packet at the Buffer
3.2. Checks the Timer Value of each Request packet at Buffer
3.2.1. Timer value is expired
3.2.1.1. Resends the Request packet that has expired Timer value
3.2.2. Timer value is not expired
3.2.2.1. Update the Timer value using spent time
4. Prepares next Request packet
4.1. Has next Request packet
4.1.1. Sends that packet
4.2. No next Request packet
4.2.1. Controller does not work
```

Fig. 7. Data transmission procedure of timer-based retransmission scheme

4 Performance Analysis by Simulation

To evaluate the performance of the proposed retransmission schemes, we analyze the reliability and the packet transmission completion for the existing RDM protocol, the sequence-based retransmission scheme, and the timer-based retransmission scheme in various network conditions.

We define the parameters used for analysis in Table 1.

Table 1. Parameters used for numerical analysis

Parameter	Description
N_a	Total number of 'a'(e.g., N_{packet} : number of total transmitted packet)
R_b	Ratio of 'b' (e.g., R_{loss} : loss ratio of DMX512 link)
T_c	Time value of 'c'(e.g., T_{timer} : timer period of controller, T_{packet} : packet generation interval by controller)

For numerical analysis, we made the simulation code for each scheme by using the MATLAB [6]. In the analysis, we set a single request-response transation time over the DMX512-A link as 2.1mS + 3.484mS = 5.584mS by referring to [6].

Fig. 8 shows the impact of the request packet generation time interval at the controller on goodput. The goodput is defined as the number of successfully transmitted data packets divided by the total transmission completion time. In this figure, the timer-based retransmission schemes give lower goodput performance because the timer-based schemes tend to spend so long time to recover the lost packets by using a long retransmission timer. However, as the packet generation time interval gets larger, all retransmission schemes show better goodtput than the existing RDM protocol.

Fig. 9 shows the impact of the packet loss rate on goodput. In this figure, as the packet loss rate increases, the existing RDM protocol provides the lowest goodput performance. This is because the goodput is decreased with more packet losses. Among the candidate retransmission schemes, the sequence-based retransmission and the timer-based retransmission (with the retransmission timer of 5.584ms) schemes show better performance than the other schemes.

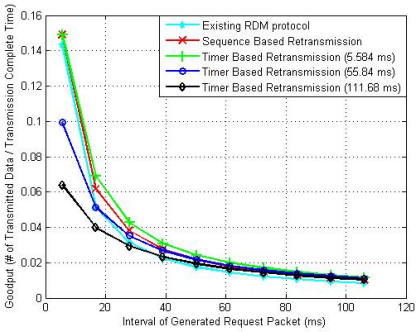


Fig. 8. Impact of packet generation interval on goodput

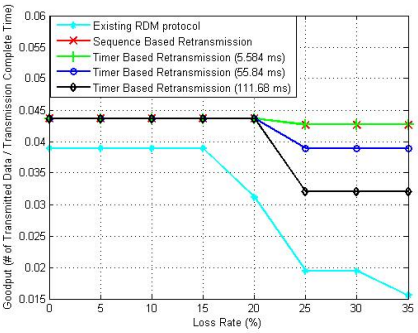


Fig. 9. Impact of packet loss rate on goodput

Fig. 10 shows the impact of the packet generation time interval on the transmission completion time. In this figure, we can see that the timer-based retransmission scheme (with the timer of 111.68ms) shows the longest transmission completion time due to the long recovery time. We can also see that the sequence-based retransmission scheme gives the performance between the two timer-based schemes with the timer of 5.584ms and the timer of 55.84ms.

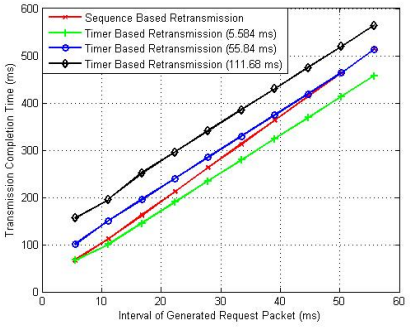


Fig. 10. Impact of packet generation interval on transmission completion time

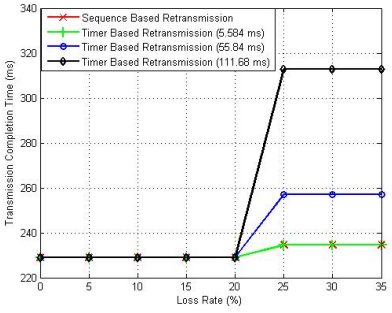


Fig. 11. Impact of packet loss rate on transmission completion time

Fig. 11 shows the impact of the packet loss rate on the transmission completion time. In this figure, it is shown that the sequence-based scheme provides the same performance with the timer-based scheme with the timer of 5.584ms, and those two schemes give better performance than the other two schemes, when the packet loss rate is greater than 20%.

From the analysis, we see that the proposed sequence-based and timer-based retransmission schemes can recover the lost packets effectively and provide the reliable transmission in the lossy network. Among the candidate retransmission schemes, the sequence-based retransmission scheme gives robust and better performance than the timer-based schemes. However, if we set the retransmission timer more appropriately, the timer-based scheme can be used more effectively.

5 Conclusions

In this paper, we proposed the two retransmission schemes for the RDM protocol: sequence-based retransmission and timer-based retransmission. The sequence-based scheme is based on the Transaction Number of the request packet, whereas the timer-based scheme uses the retransmission timer.

By numerical analysis, the proposed retransmission schemes are compared with the existing RDM protocol in terms of reliability and transmission completion time. From the results, it is shown that the retransmission schemes can recover the lost packet effectively in the lossy network. Among the candidate retransmission schemes, the sequence-based retransmission scheme gives robust and better performance than the timer-based schemes. However, if we set the retransmission timer more appropriately, the timer-based retransmission scheme can provide the best performance.

Acknowledgment. This research was supported by the MSIP support program of NIPA (NIPA-2013-H0301-13-2004), and ICT Standardization program of MSIP (Ministry of Science, ICT & Future Planning).

References

1. Roo, Y.-S., Jang, W.-C.: Implementation of Ubiquitous Lighting Network System Using Embedded Linux. In: Proceeding of the Institute of Electronics Engineers of Korea, pp. 613–614 (2007)
2. Professional Lighting And Sound Association (PLASA), http://tsp.plasa.org/
3. American National Standard Institute(ANSI), Asynchronous Serial Digital Data Transmission Standard for Controlling Lighting Equipment and Accessories, American National Standard E1.11 (2008)
4. American National Standard Institute(ANSI), Remote Device Management Over DMX512 Networks, American National Standard E1.20 (2006)
5. Cerf, V.G., Kahn, R.E.: A Protocol for Packet Network Intercommunication. IEEE Transaction on Communications Com-22(5) (1974)
6. Homepage of MATHWORK, http://www.mathworks.co.kr/

Configuration of Tracking Area Code (TAC) for Paging Optimization in Mobile Communication Systems

Hyung-Woo Kang[1,*], Woo-Ju Kim[1], Seok-Joo Koh[1],
Hyon-Goo Kang[2], and Jung-Bae Moon[2]

[1] Kyungpook National University, Daegu, Korea
{hwkang0621,kachukun}@gmail.com, sjkoh@knu.ac.kr
[2] SK Telecom Acess Network Lab, Seoul, Korea
{hyongoo.kang,jbmoon}@sk.com

Abstract. Recently, the mobile telecommunication traffics have been rapidly increasing due to the growth of smart phone services. In this paper, we propose a new scheme for configuration of TAC (Tracking Area Code) to maximize the paging success rates in the LTE-based mobile communication networks. The proposed scheme includes the initial configuration of TAC, the local optimization algorithm, and the re-clustering algorithm for further improvement of the TAC configuration. From the performance analysis with real traffic data of service provider, we see that the proposed TAC configuration scheme can improve the paging success rates in the LTE networks, compared to the existing TAC configuration scheme.

Keywords: LTE, Tracking Area Code, Paging, Optimization.

1 Introduction

With the popularity of smart phones, mobile communication has been rapidly changes from 3GPP to Long Term Evolution (LTE). Recently, LTE is emerging as the new mobile communication technology [1-3].

A mobile user typically moves around in a zone that is composed of many cells in a mobile communication system. When a call request to a specific user arrives, the cellular system should page the user in the cells to locate the user in the network. It is noted that the cellular systems require efficient methods to find a specific mobile user in the paging process [4-6]. In particular, the paging success rate is a very important factor in the design of the paging areas. A paging area is defined by a Tracking Area Code (TAC).

In this paper, we propose a new scheme for configuration of TACs to improve the paging success rates in the paging process. A TAC consists of a group of cells to which a paging signal is broadcast in the paging process. The proposed TAC configuration scheme can be used to increase the paging success rate and to reduce the overhead of paging traffic in the LTE networks.

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_9,

This paper is organized as follows. In Section 2, we discuss the TAC optimization issue and how to configure the TAC in the mobile communication system. In Section 3, we describe the proposed TAC configuration algorithm. Section 4 describes the experimental analysis and comparison with the existing scheme. Finally, Section 5 concludes this paper.

2 TAC Optimization in Mobile Communication System

2.1 TAC Configuration

At present, most of the mobile operators configure the TAC in an arbitrary way, in which only the topological location information of all cells are considered and the network manager manually configures a group of cells as a TAC. TAC is defined as a group of cells. The TAC is coded into a 2-bytes hexadecimal digit, in which the first byte represents the area code, called the Tracking Area List. When a call request is required for a specific user, if the user was registered with a TAC, then a paging signal will be broadcast to all of the cells contained in the TAC.

Usually, when a mobile user is connected to a cell in network attachment phase, the user is assigned to the TAC that the cell is associated with. If the user moves to a new cell in the dormant mode, the TAC of the user may be changed, if the TAC of the new cell is different from that of the old cell. Accordingly, the TAC should be configured by considering the mobility and traffic of users in the mobile network.

The Self-Organizing Network (SON) system is used in the LTE-based mobile communication in order to configure and manage a mobile network. For TAC configuration, the TAC optimization S/W provides the information of optimized TAC configuration to the SON system. Figure 1 describes the overview of SON system [7].

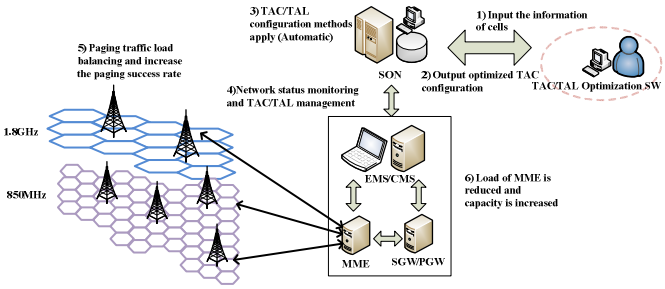


Fig. 1. SON system and TAC Configuration

With such TAC configuration, when a paging request to a specific user arrives, the paging operations are performed as follows. In the first paging, the paging request message (or signal) will be broadcast to the cells contained in the TAC. If the paging request fails (i. e., no response to the paging request from the mobile user), then the second paging is performed, in which the paging request will be broadcast to all of the cells in the area. Therefore, it is important to optimize the TAC configuration so as to

maximize the paging success rate of the first paging and thus to reduce the paging traffics generated in the network.

3 Proposed TAC Configuration Algorithms

3.1 TAC Optimization Model

For TAC optimization, we first make a mathematical optimization model for TAC configuration. Given a network area with many cells, the goal is to find an optimal TAC configuration (mapping from a group of cells to a TAC) by considering traffic and user mobility between cells. The Paging Success Rate (PSR) is used as the objective function of the optimization model. The PSR represents the success probability of the first paging.

Figure 2 shows the mathematical model of TAC optimization.

Objective function: Maximize PSR

$$\mathbf{PSR = \Sigma_{k\in M} \Sigma_{i\in N} \Sigma_{j\in N} \lambda_i \times P_{ij} \times X_{ik} \times X_{jk}}$$

Constraints: feasibility conditions

$$\mathbf{\Sigma_{k\in M} X_{ik} = 1, \text{ for all } i \in N \text{ (TAC assignment condition)}}$$

$$\mathbf{\Sigma_{i\in N} X_{ik} \leq S_{TAC}, \text{ for all } k \in M \text{ (TAC size condition)}}$$

$$\mathbf{\Sigma_{i\in N} \lambda_i \times X_{ik} \leq C_{TAC}, \text{ for all } k \in M \text{ (PTL condition)}}$$

$$\mathbf{d_{ij} \times X_{ik} \times X_{jk} \leq D_{TAC} \text{ for all } i, j \in N, k \in M \text{ (distance condition)}}$$

$$\mathbf{X_{ik} = 1 \text{ or } 0, \text{ for all } i \in N, k \in M}$$

Fig. 2. TAC Optimization Model

In the optimization model, we use the following variables and parameters:

- P_{ij}: mobility ratio that user moves from cell i to cell j;
- λ_i: sum of paging traffics in cell i;
- d_{ij}: distance between cell i and j;
- N: a group of cells in TAL, with the size of n;
- M: a group of TAC in TAL, with the size of m;
- S_{TAC}: the maximum number of cells that a TAC can contain;
- C_{PTL}: the maximum paging traffic load allowable for a TAC; and
- D_{TAC}: the maximum distance between two cells allowable for a TAC.

On the other hand, we consider the following feasibility conditions:

- TAC assignment condition: each cell should be assigned to at least one TAC;
- TAC size condition: the number of cells assigned to a TAC cannot exceed S_{TAC};
- Paging Traffic Load (PTL) condition: the paging traffic load for a TAC cannot exceeds C_{PTL};
- Distance condition: the distance between two cells in TAC cannot exceed D_{TAC}.

3.2 Proposed TAC Configuration Schemes

Based on the TAC optimization model described in Section 2, we propose the TAC configuration algorithms in this section. The proposed algorithms consist of initial configuration of TAC, local optimization algorithm, and re-clustering algorithm for further improvement of the TAC configuration.

Figure 3 shows the overall sequence of the proposed algorithms.

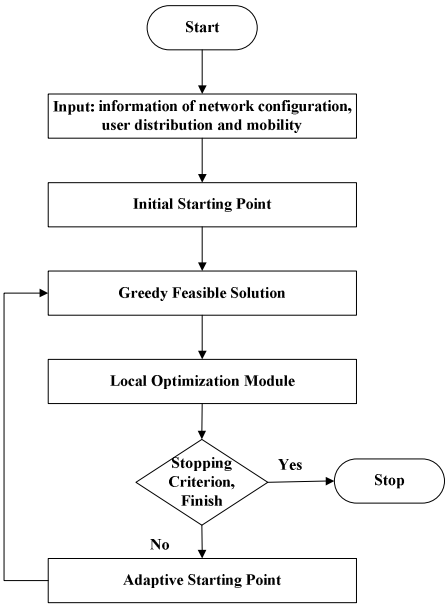


Fig. 3. Overall Algorithm for TAC Configuration

As input to TAC configuration, we consider the mobility ratio and paging traffic per cell and the distance between cells. The Initial Starting Point (ISP) algorithm is used to select an initial cell for each TAC as the starting point to overall algorithm. Based on these initial points, the Greedy Feasible Solution (GFS) algorithm is used to obtain a feasible TAC configuration, in which all of the cells are assigned to one of the TACs in the network. In the GFS algorithm, each cell will be assigned to a TAC in a stepwise way so as to maximize the PSR objective function while satisfying the feasibility conditions.

After the GFS algorithm is performed, we apply the Local Optimization Module (LOM) algorithm to find a more optimized solution. In LOM, we try to optimize the TAC configuration by changing the TAC of a cell to the other TAC. Finally, the Adaptive Starting Point (ASP) algorithm is used to explore the other solution space for further optimization, in which an initial stating cell per TAC will be re-calculated and then the optimization algorithms (GFS and LOM) are performed again. Such procedures are repeated until the stopping criterion is satisfied. The details of each algorithm are described below.

3.2.1 Initial Starting Point (ISP)

The ISP algorithm is the first step of overall optimization algorithm, in which an initial cell is randomly selected for each TAC.

3.2.2 Greedy Feasible Solution (GFS)

In GFS, a feasible TAC configuration is obtained, based on the initial cells of ISP. For each TAC, we select the cell with the largest mobility rate moving from the initial cell among the other cells that have not assigned to any TAC. The selected cell will be assigned to the TAC, if the feasibility conditions are still satisfied by this assignment. GFS will be repeated until all cells are assigned to one of the TACs.

3.2.3 Local Optimization Module (LOM)

The LOM algorithm finds more optimized configuration of TAC. Based on the configuration obtained by GFS, we try to change the TAC of a cell to another TAC. If the PSR value is improved by the TAC change, we perform the change of TAC for the cell. Otherwise, if the PSR value is not improved or the feasibility condition is not satisfied, we do not change the TAC of the cell. These procedures are performed until no improvement of PSR is made for all cells.

3.2.4 Adaptive Starting Point (ASP)

After LOM is completed, the ASP algorithm is used to explore much more solution space in the optimization process. For each TAC given by LOM, we find the 'center' cell, which is obtained by calculating the distances of the other cells from a candidate center cell within the TAC. We define the center cell as the cell with the minimum distance to the other cells.

In ASP, a center cell is calculated for each TAC and it is used as an initial point to GFS and LOM. This ASP algorithm is performed until a pre-specified stopping rule (e. g., totally 10 times)

4 Experimental Results and Analysis

4.1 Test Network Environment

In the experiments, we use the real-world data of network topology, user traffic, and the mobility rates to neighboring cells, which are based on the information of SK Telecom in Korea. The proposed TAC configuration scheme was applied to a target area, and we calculate the PSR value.

For all of the experiments, the default values for parameters are set as follows: N is different for TAL, M = 11, S_{TAC} = (N / M) + 10, C_{TAC} = 2100, D_{TAC} = 1/2 * the maximum distance between cells in TAL.

4.2 Results and Discussion

In experiments, the distance between two cells is calculated by using the Euclidean distance. The probability of moving from a cell to another is assumed to be 20%,

which is represented as α. That is, a mobile user remains within the current cell with a probability of 0.8. Then, the Paging Success Probability (PSP) can be calculated as follows, as shown in Figure 4.

Paging Success Probability (PSP)

$PSP = (1-\alpha) + \alpha \times PSR / \lambda,$

where $\lambda = \Sigma_{i \in N} \lambda_i$, α = the probability of moving from a cell to another

Fig. 4. Calculation of PSP

Table 1 shows the comparison of the existing and proposed schemes for TAC configuration, which is applied to the target area that is coded 19.

Table 1. Comparing the Existing and Optimization TAC configuration of target area

	Existing TAC Configuration			**Optimized TAC Configuration**		
TAC Index	**# of cells**	**maximum distance**	**PTL**	**# of cells**	**maximum distance**	**PTL**
1900	51	12624	684	64	9565	968
1901	56	4041	829	63	5990	688
1902	48	4629	671	64	9624	264
1903	44	2714	680	64	4489	710
1904	72	6325	1083	62	6436	524
1905	68	6789	943	64	5673	787
1906	53	6138	494	65	6917	904
1907	55	3797	569	61	6236	584
1908	5	4002	60	57	6795	685
1909	5	1152	73	71	8115	401
190A	193	7995	571	5	4002	60

In the table, we can see that the original (existing) TAC configuration is composed of a total of 642 cells and is made up with a number of 11 TACs. It is noted that the TAC 190A contains 193 cells, which violates the feasibility condition of S_{TAC}, whereas the TAC 1908 and 1909 have only 5 cells. This implies that the existing TAC configuration is not balanced in the viewpoint of the TAC size. This leads to a lower Paging Success Rate. The standard deviation of PTL is 91268. Overall, the existing TAC configuration gives the Paging Success Probability (PSP) of 89.39.

On the other hand, the optimized TAC configuration assigns the total 642 cells to 11 TACs. Most of the TACs have the TAC size of a minimum 61 cells and up to 71 cells, except for TAC 190A. This is because the TAC 190A is a subway TAC, and thus this TAC cannot be further optimized. Note that all TACs give the nearly equal size, except the subway TAC, in the proposed optimization scheme. The standard deviation of PTL is 67268, which is lower than the existing configuration. The PSP of the proposed scheme is calculated as 91.58, which is greater than the existing scheme.

Figure 5 compares the PSP values of target area for the existing and proposed schemes as phases of optimization. From the figure, we can see that the proposed scheme provides higher PSP than the existing scheme. That is, the proposed optimization scheme improves the PSP compared with the original configuration.

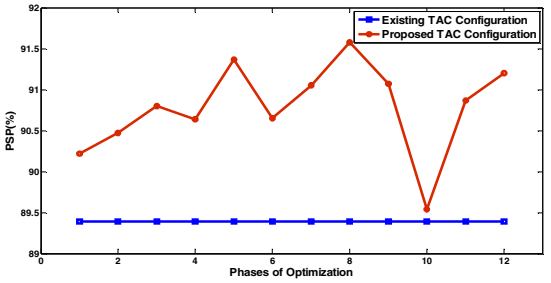


Fig. 5. Paging Success Probability of target area

Figure 6 and 7 shows the distribution of TACs for the target area, in the viewpoint of physical location, for the original and optimized TAC configuration schemes.

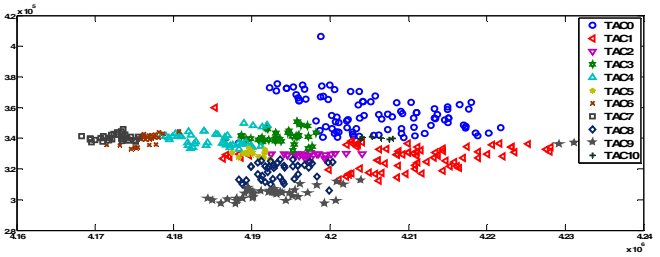


Fig. 6. Existing TAC configuration

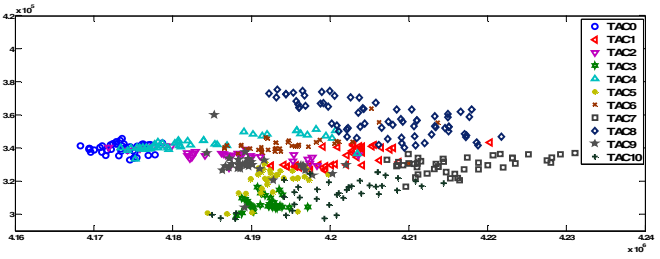


Fig. 7. Proposed TAC configuration

From the figure, we can see that the proposed scheme gives more compact and balanced TAC configurations that the existing scheme. This is because the proposed scheme applies the feasibility conditions for TAC configuration, and the PSP is maximized by including the neighboring cells into the same TAC.

5 Conclusions

In this paper, we presented a new scheme for configuration of TAC to maximize the paging success rates in LTE-based mobile communication networks. The proposed scheme is composed of a set of sub-algorithms: Initial Starting Point (ISP), Greedy Feasible Solution (GFS), Local Optimization Module (LOM), and Adaptive Starting Point (ASP). In the TAC optimization, we also considered the feasibility conditions to get a balanced configuration of TACs.

By the experimentations for real-world data of network topology and user traffic, the proposed optimization scheme is compared with the original existing scheme in the perspective of paging success probability. From the results, we can see that the proposed scheme provides more optimized TAC configurations than the existing scheme by maximizing the paging success probability. It is also noted that the proposed scheme gives a more compact and balanced TAC configuration than the existing scheme by using the feasibility conditions. It is expected that the proposed TAC optimization scheme can be used in the real-world mobile communication networks to maximize the paging success rates and to reduce the paging traffic load.

Acknowledgment. This research was supported by the Basic Science Research Program of NRF(2010-0020926), and by the MSIP support program of NIPA(NIPA-2013-H0301-13-2004).

References

1. Astély, D., et al.: LTE: the evolution of mobile broadband. IEEE Communications Magazine 47(4), 44–51 (2009); 44-51.3GPP TS 36.300, Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN): Overall Description Stage 2 (2008)
2. Ulvan, A., Bestak, R., Ulvan, M.: The study of handover procedure in LTE-based femtocell network. In: Wireless and Mobile Networking Conference, WMNC (2010)
3. Modarres Razavi, S.: Tracking Area Planning in Cellular Networks: Optimization and Performance Evaluation. Dissertation of Linköping (2011)
4. Bar-Noy, A., et al.: Paging mobile users in cellular networks: Optimality versus complexity and simplicity. Theoretical Computer Science (2012)
5. Lyberopoulos, G.L., et al.: Intelligent paging strategies for third generation mobile telecommunication systems. IEEE Transactions on Vehicular Technology 44(3), 543–554 (1995)
6. 3GPP TS 36.902, Evolved Universal Terrestrial Radio Access Network (E-UTRAN): Self-configuring and self-optimizing network (SON) use cases and solutions (2009)

OpenFlow-Based Implementations of Distributed ID-LOC Mapping System in Mobile Internet

Nak-Jung Choi, Ji-In Kim, Jin-Ho Park, and Seok-Joo Koh

School of Computer Science and Engineering, Kyungpook National University, Korea
{peaceful7007,jiin16,jin.ho.paul.park}@gmail.com,
sjkoh@knu.ac.kr

Abstract. The future Internet will be evolved to mobile-oriented environments, and thus the mobility support is a key issue in the design of future Internet. This paper proposes a distributed identifier-locator mapping system (DMS) for the future mobile-oriented Internet environment. The proposed DMS scheme is implemented over Linux platform by using the OpenFlow and Click Router software. From the experimentations over the real testbed network of Korea Research Education Network (KOREN), we can see that the proposed DMS scheme can perform the mobility management operations effectively for mobile Internet hosts.

Keywords: Mobile Internet, ID-LOC Separation, Distributed ID-LOC Mapping System, OpenFlow Implementation, Testbed Experimentation.

1 Introduction

It is noted that the current Internet was historically designed for fixed network environment, rather than for the mobile network environment. This has enforced Internet to add a lot extensional features to satisfy the mobility requirements, as shown in the example of Mobile IP (MIP) [1, 2]. However, this patch-on approach seems to be just a temporal heuristic to the problems in the mobile environment, rather than an optimization. Thus, a variety of research activities have been made to design the future Internet for mobile environment, which include eMobility [3], 4WARD [4], FIND [5], MobilityFirst [6], GENI [7], and AKARI [8]. It is also noted that many challenging works are in progress with the identifier-locator separation principle, as shown in Host Identity Protocol (HIP) [9], Locator-Identifier Separation Protocol (LISP) [10], and Identifier-Locator Network Protocol (ILNP) [11].

With these observations, we design the architecture of Mobile Oriented Future Internet (MOFI) [12] to support the mobile environment of future Internet. The MOFI is designed with the following functional blocks: Host Identifier and Local Locator (HILL) and Distributed Mapping System (DMS). The details of the MOFI architecture are described in [12].

This paper specifies a distributed mapping system (DMS) to support the identifier-locator (ID-LOC) mapping management based on the ID-LOC separation. The DMS control operations include HID-LOC binding for HID-LOC registration and LOC

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_10,

query for data delivery. The DMS functional entities include Local Mapping Controller (LMC) at AR and Global Mapping Controller (GMC) at gateway (GW). For intra-domain communication, LMC is used to maintain the Local Map Register (LMR) to keep the list of HID-LOC mapping for the hosts. For inter-domain communication, GMC is used to maintain a Global Map Register (GMR) for HID-LOC mapping management for the hosts in the other domain.

This paper is organized as follows. Section 2 briefly summarizes the MOFI architecture. In Section 3 we describe the HID-LOC mapping control operations for DMS. Section 4 describes the Linux-based implementation using the OpenFlow and Click Router platform and the experimental results performed over the Korea Research and Education Network (KOREN) network. Section 5 concludes this paper.

2 MOFI Architecture Overview

2.1 Host ID and Local Locator (HILL)

HID is used to identify a mobile host. LOC is used to represent the location of a host in the network and also used for delivery of data packets. In addition, a LOC is a locally routable address that has only to be locally unique in the concerned network. As a LOC, we use the IP address of an access router that a host is attached to. These IP addresses may be local in the network.

Figure 1 shows the data delivery operations with global HID-based communication and local LOC-based routing in MOFI [12].

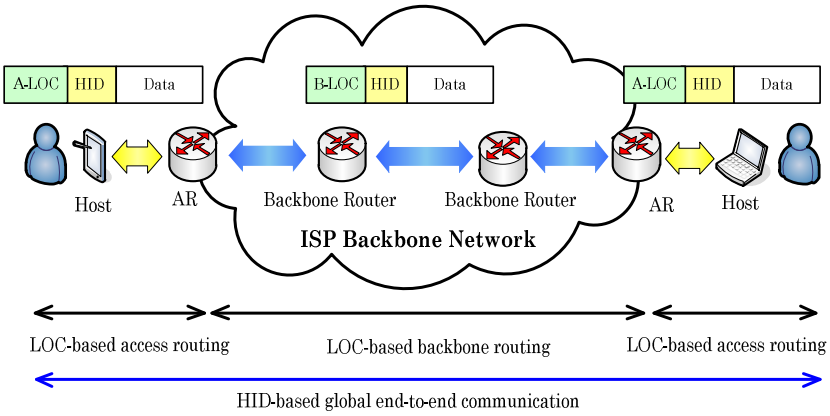


Fig. 1. HID-based Communication and LOC-based Routing

In the figure, the communication between two hosts is performed with HIDs, whereas LOCs are used for packet delivery. For packet delivery, an access LOC (A-LOC) is used as LOC within access network, whereas the IP address of AR will be used as backbone LOC (B-LOC) in the backbone network. The data packet routing is performed locally in the access and backbone networks.

2.2 Distributed HID-LOC Mapping System (DMS)

The mapping between HID and LOC of MOFI is managed by DMS. For description of DMS, we divide the network reference model into intra-domain and inter-domain cases. And for DMS MOFI assume that many of Cache and Register. [Table 1] is show the Cache and Register information.

Table 1. Caches and Registers

Category	Entity	Cache/ Register	Usage
Intra domain	AR	LBC	Data forwarding (host – AR)
		DFC	Data forwarding (AR –AR , AR – GW)
	LMC	LMR	DHT-based mapping control
Inter domain	GW	DFC	Data forwarding (GW – GW)
	GMC	GMR	Domain-based mapping control

The mapping between HID and LOC is managed by DMS. For description, we divide the network reference model into intra-domain and inter-domain cases. Figure 2 shows the overall network model for HID-LOC mapping control of DMS, in which a domain represents the network domain of an ISP.

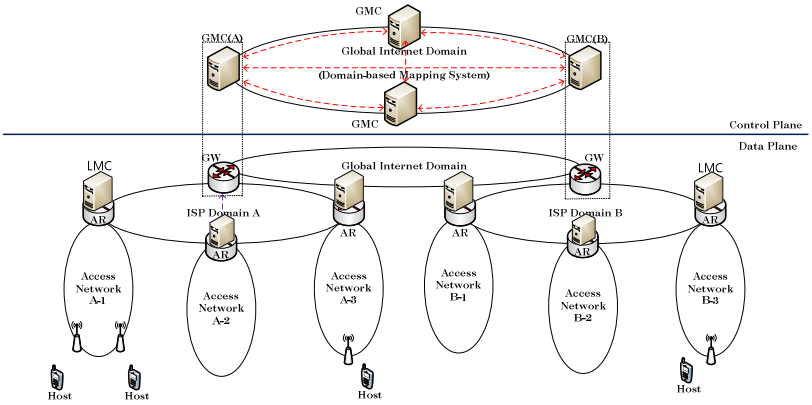


Fig. 2. Mapping Control Model of DMS

In the figure, the control plane is separated from the data plane. In the data plane, an Access Router (AR) maintains a Local Mapping Controller (LMC) that maintains the list of HID-LOC of the locally attached hosts. Each domain gateway (GW) maintains its Global Mapping Controller (GMC), which is used to the HID-LOC mapping information for all of its domain hosts.

3 HID-LOC Mapping Control in DMS

The HID-LOC mapping control by DDMS is further divided into 'intra-domain' and 'inter-domain' cases. This paper focuses on the 'inter-domain' mapping control. The discussion of intra-domain case is omitted in this paper. Because intra-domain is substitute Openflow Click system. First I show this table. [Table 2] shows the list of the messages for mapping control.

Table 2. Control Messages for Mapping Control

Message	Full Name	From	To
HBR	HID Binding Request	Host/GMC	GMC/GMC
HBA	HID Binding ACK	GMC/GMC	GMC/Host
LQR	LOC Query Request	GMC	GMC
LQA	LOC Query ACK	GMC	GMC

The HBR and HBA control messages are exchanged between host and GMC, or between GMCs. On the other hands, the LQR and LQA messages are exchanged between GMCs. Each control message is encapsulated into UDP.

3.1 HID-LOC Binding Operation

With network attachment of host, the HID-LOC operation is performed, in which HID and LOC of the host will be registered with LMC. LMC will maintain and update its Local Mapping Register (LMR) the information of bindings between HIDs and LOCs for all of the hosts that are attached to the LMC/AR.

In HID-LOC binding, the HID Binding Request (HBR) and HID Binding ACK (HBA) messages are exchanged as shown in Figure 3. After the HID-LOC binding between host and LMC/AR, the LMC/AR of the host sends a HBR message to the GMC/GW. The HBR message contains the HID of the host and the LOC (IP address of LMC). Based on the received HBR, the GMC updates its Global Mapping Register (GMR) and responds with a HBA message to LMC of the host.

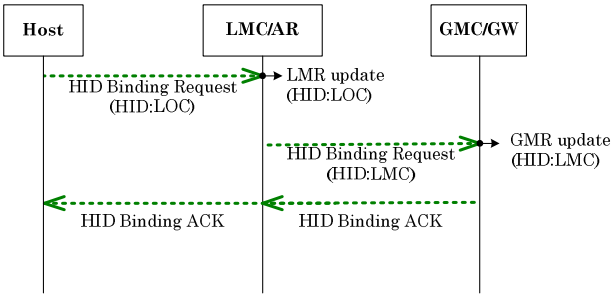


Fig. 3. HID-LOC binding between host and LMC/GMC

3.2 LOC Query and Data Delivery Operations

In the LOC query operation, LMR and GMR will be used to find the AR and GW that a mobile host is attached to. Figure 4 shows the LOC query and data delivery operations for inter-domain case. In this case, we assume that a sending host (SH) is attached to LMC/AR1 and a receiving host (RH) is connected to LMC/AR2 via GW1and GW2.

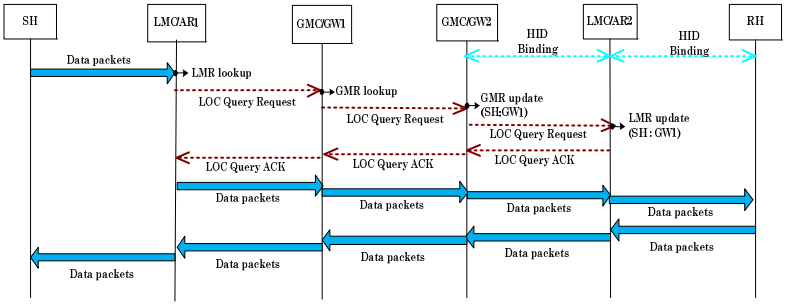


Fig. 4. LOC query and data delivery for inter-domain case

When a data packet arrives from SH, the LMC/AR1 sends a LOC Query Request (LQR) message to GMC/GW1. On reception of a LQR message, GMC1 sends LQR to GMC2, and then GMC2 will send the LQR message to LMC/AR2. After looking up the LMR, LMC/AR2 responds with a LOC Query ACK (LQA) message to GMC/GW2. Then, GMC/GW2 sends LQA message to GMC/GW1 and further to LMC/AR1. AR1 of SH can now send a data packet to RH via GW1, GW2 and AR2.

4 OpenFlow-Based Implementation and Experimentation

To validate the proposed DMS scheme, the mapping control operations were implemented using UDP socket programming [13], OpenFlow [14] and Click Modular Router [15] over the Linux platform.

4.1 Intra-Domain Communications

For intra-domain implementation, we use the OpenFlow software and an arbitrary IPv4 address as the HID. Figure 5 shows the network model for intra-domain implementation. For implementation of AR/LMC, we employ OpenFlow switch and NOX controller. The OpenFlow switch will function as AR, and it forwards the data packets. All of the LMRs in the domain are implemented over the OpenFlow NOX Controller. In HID-LOC binding operation, a flow table is maintained by OpenFlow Switch so as to manage all of the hosts that are attached to the AR. The LMR managed by NOX Controller will maintain the list of all hosts that are attached in the domain. The Packet-in and Packet-out messages of OpenFlow are used as the HBR and HBA

messages of DMS. In the LOC query operations for data delivery, the flow table of Openflow switch will be updated. It is noted that the Packet-in/Flow-Mod messages of OpenFlow correspond to the LQR/LQA messages of DMS.

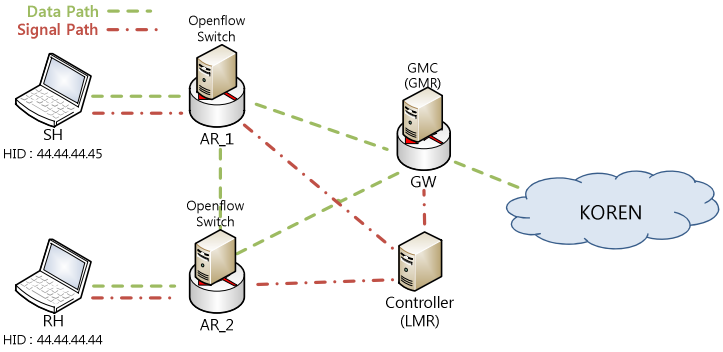


Fig. 5. Intra-domain implementation by OpenFlow and Click Router

Figure 6 shows the packet capturing results for intra-domain communication by using the WireShark tool [16]. The test environment is the same with Figure 5. From the figure, we can see that the data packets are delivered between the two hosts with the help of the DMS operations.

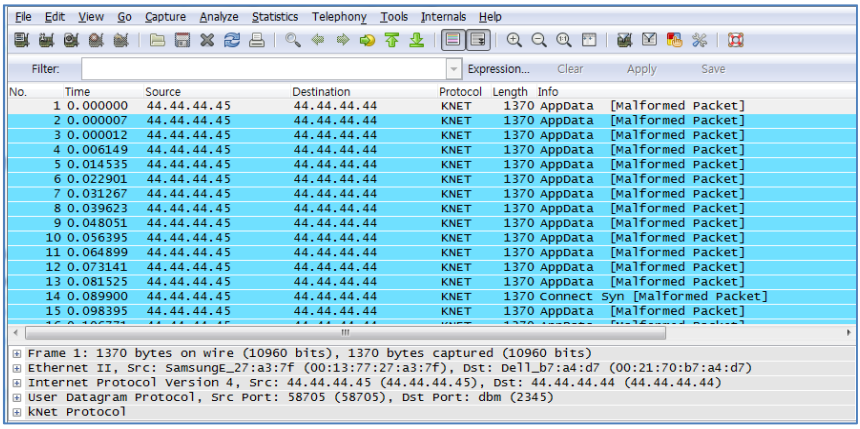


Fig. 6. Intra-domain Packet Capture

Figure 7 shows the information of LMR that is updated by the HID-LOC binding operation. From the figure, 2c2c2c2c is the hexadecimal representation of the HID (44.44.44.44), and the number of 210.114.94.174 represents the location of AR that the host is attached to. From the figure, we can see that the LMR of LMC is updated as per the HID-LOC binding operation.

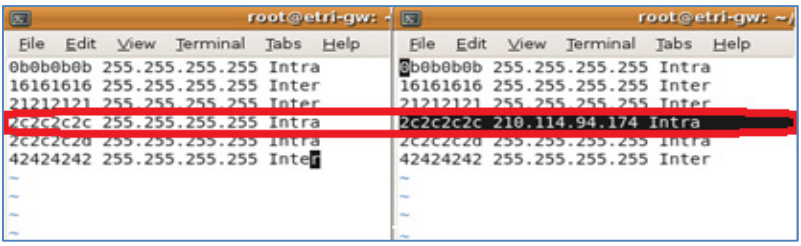


Fig. 7. LMR update

4.2 Inter-Domain Communications

For implementation of inter-domain mapping control, we use the UDP socket programming. Fig. 8 shows the testbed configuration for inter-domain implementation of the proposed DMS scheme. There are the two domains, one is located at the KNU site and another one is located at the ETRI site, and the two domains are connected by Korea Research and Education Network (KOREN) [17].

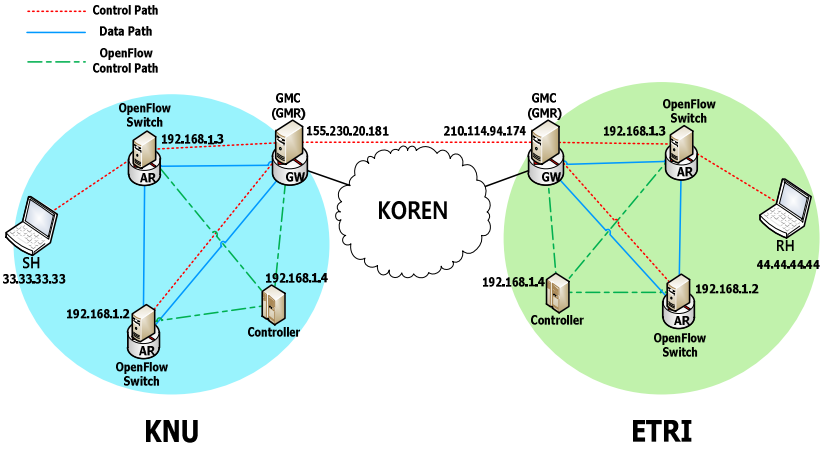


Fig. 8. Testbed Network for Inter-domain Implementation

Figure 9 shows the packet capturing results that are measured by GW for inter-domain communication between KNU and ETRI domains. From the figure, we can see that the data packets are delivered between the source and destination hosts via the domain GWs. In addition, the data packets are encapsulated by using the IP-in-IP tunneling scheme at each GW.

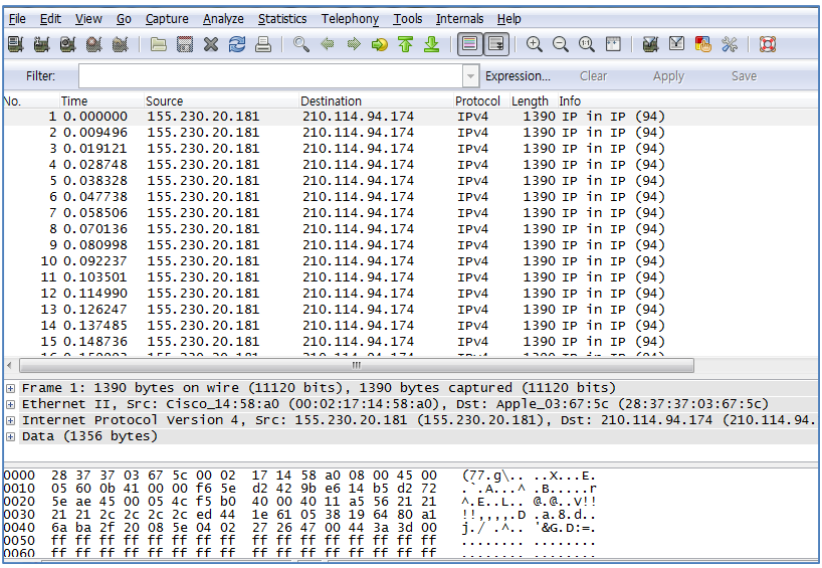


Fig. 9. Gateway packet capture

5 Conclusions

Until now, we have described the distributed identifier-locator mapping system (DMS) for the future mobile-oriented Internet environment. For validation, we implemented the proposed DMS scheme over Linux platform by using the OpenFlow and Click Router software. From experimentations over the real testbed network of Korea Research Education Network (KOREN), we can see that the proposed DMS scheme can perform the mobility management operations effectively for mobile Internet hosts.

For further works, the implementations and experimentations of DMS/MOFI need to be further extended by considering the real world network environments.

References

1. IETF RFC 5944, IP Mobility Support for IPv4 (revised November 2010)
2. IETF RFC 3775, Mobility Support in IPv6 (June 2004)
3. eMobility Project, `http://www.emobility.eu.org/`
4. 4WARD Project, `http://www.4ward-project.eu/`
5. Future Internet Design (FIND), `http://www.nets-find.net/`
6. Mobility First: Future Internet Architecture, `http://mobilityfirst.winlab.rutgers.edu/`
7. Global Environment for Network Innovations (GENI), `http://www.geni.net/`
8. The AKARI Project, `http://akari-project.nict.go.jp/eng/`
9. Host Identity Protocol (HIP), `http://www.ietf.org/html.charters/hip-charter.html`

10. Locator Identifier Separation Protocol (LISP), http://www.ietf.org/html.charters/lisp-charter.html
11. Identifier Locator Network Protocol (ILNP), http://ilnp.cs.st-andrews.ac.uk/
12. Mobile Oriented Future Internet (MOFI), http://www.mofi.re.kr
13. Forouzan, B.A.: TCP/IP Protocol Suit. McGraw-Hill Press (2012)
14. OpenFlow, http://www.openflow.org/
15. Kohler, E.: The Click Modular Router. Ph. D. Thesis. MIT (2000)
16. Wireshark, http://www.wireshark.org
17. KOREN, http://www.koren.kr/koren/kor/

A Study on Performance Comparison of Cloud Architectures Using Nested Virtualization*

HeeSeok Choi, TaeMuk Lyoo, JongBeom Lim, Daeyong Jung,
Jihun Kang, Taeweon Suh, and Heonchang Yu**

Department of Computer Science Education, Korea University, Seoul, Korea
{hsrangken,lyoo12,jblim,karat,k2j23h,suhtw,yuhc}@korea.ac.kr

Abstract. In recent years, cloud computing has become a significant technology trend because of various advantages including cost savings, flexibility, high availability, and scalability using virtualization technology. However, one of the concerns for using cloud computing is security. In fact, there are multiple attack surfaces in virtualized environments. In this paper, we build a fault tolerant cloud architecture using nested virtualization. With our constructed cloud architecture, we argue that a malicious virtual machine cannot subvert the whole virtual machines on the physical host machine. To support this, we compare the performance of two types of virtual machines (i.e., nested virtual machines and regular virtual machines). Results provide encouraging support for the validity of our cloud architecture with negligible performance degradation while having fault tolerance.

Keywords: Cloud architecture, Hypervisor, Nested virtualization, Virtual machine.

1 Introduction

Cloud computing has gained a momentum and is transforming the Internet-based computing infrastructure. Cloud computing offers huge opportunities to the IT industry and dynamically scalable virtual machines provisioned as a service over the Internet on a pay-per-use basis as needed. The key technology of cloud computing is virtualization, by which a virtual machine that acts as a real computer machine with some degree of performance degradation can be created and managed.

However, multiple attack surfaces exist in virtualized environments as shown in Figure 1. Once a virtual machine is exploited by an attack or rootkit, the virtual machine may affect other virtual machines within the physical host machine. Specifically, for example, the attacked virtual machine also can exploit other virtual machines by masquerading the source address of the sender and commanding malicious source codes. Indeed, if the network interface controller of the attacked

* This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2013056346).
** Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_11, © Springer-Verlag Berlin Heidelberg 2014

virtual machine is set to the promiscuous mode, the attacked virtual machine can sniff all the packets that are going through all the virtual machines on the physical host machine (this attack scenario is indicated by circled one in Figure 1).

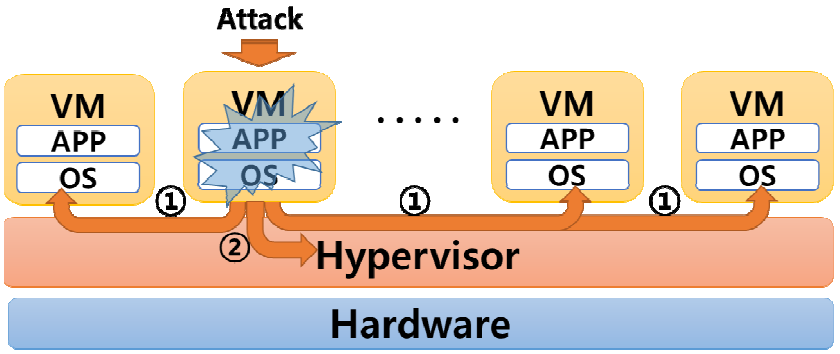


Fig. 1. Attack surfaces in virtualized environments

Another potential attack surface is that the attacked virtual machine may directly exploit the hypervisor (this attack scenario is indicated by circled two in Figure 1). In this case, once the hypervisor is exploited by one of the virtual machines existing on the physical host machine, all the virtual machines may be shut down regardless of their intention.

In this paper, we build a fault tolerant cloud architecture using nested virtualization [1] to mitigate these security threats. With our constructed cloud architecture, we argue that a malicious virtual machine cannot subvert the whole virtual machines within the physical host machine. We note that with nested virtualization, a regular virtual machine can provision other virtual machines from inside itself. To support the argument, we compare the performance of two types of virtual machines (i.e., nested virtual machines and regular virtual machines) by executing benchmark programs.

The rest of this paper is organized as follows. We discuss about security threats in virtualized environments and related works in Section 2. Section 3 presents our proposed cloud architecture using nested virtualization; performance results for benchmark programs are also shown in this section. Finally, Section 4 gives our conclusion.

2 Related Works

There have been a number of security threats in virtualized cloud environments. In [2], the authors have shown that third-party cloud computing services such as Amazon's EC2 (Elastic Compute Cloud) and Microsoft's Azure can introduce security vulnerabilities. More specifically, they show that it is possible to map the internal cloud infrastructure, identify where a particular virtual machine is to reside by using side channel attacks between virtual machines to extract information from a target virtual machine.

In [3], the authors have shown that there was a vulnerability in Amazon's EC2 services to signature wrapping attacks. In this scenario, a cloud user can replace the SOAP (Simple Object Access Protocol) request message with other codes (i.e., launch a number of virtual machine instances), by exploiting the SOAP message security validation vulnerability. In [4], the authors have shown that existing virtual machine solutions were vulnerable to malicious codes such as crashme and iofuzz, and therefore, virtual machines can easily be compromised.

To remove attack surfaces in virtualized cloud infrastructures, NoHype [5] has been proposed. In NoHype architecture, the authors claim that a virtual machine can be provisioned without the virtualization layer, in which a malicious party has the opportunity to attack. In fact, once a virtual machine is initialized, the virtual machine runs without the need of hypervisor on the system. However, when initializing a virtual machine in NoHype architecture, it actually needs to get hypervisor's help, and there are numerous tricks to operate virtual machines.

In our approach, however, we let the virtualization layer exist, contrast to NoHype architecture. Furthermore, we added the virtualization layer to the cloud architecture using nested virtualization. In our cloud architecture, attack surfaces can be limited to a virtualization layer because, technically speaking, diffusing the attack from a nested virtual machine to the whole virtualization layers is unrealistic.

Apart from the security threats, there are a number of virtualization research efforts to improve the performance of the virtualization. For example, the authors in [6] have suggested an approach to I/O virtualization, called self-virtualized devices to improve I/O performance by offloading select virtualization functionality to the device.

As we shall see, the performance of virtual machines gradually decreases, as the number of virtualization layers increases. However, if we adopt the I/O virtualization technology as well as processor virtualization, the amount of performance degradation will be reduced. Note that in our cloud architecture using nested virtualization, the I/O virtualization technology is not applied due to the system's limitation.

3 System Architecture and Performance Evaluation

In this section, we present the cloud architecture using nested virtualization, which eliminates the security threats. Then, we show the performance results of virtual machines including regular virtual machines and nested virtual machines by using SPEC CPU 2006 benchmark programs [7].

3.1 Experimental Architecture

Figure 2 shows the differences between the traditional virtualization architecture (bare-metal) and the one that is used in this paper to evaluate the performance of virtual machines of each layer (nested virtualization). In the traditional bare-metal architecture as shown in Figure 2(A), virtual machines are running on top of physical host machine and the virtual machines do not have the ability to provision another virtual machine from inside itself. On the contrary, in our experimental cloud architecture, the one-layered virtual machines can also provision other virtual machines as shown in Figure 2(B).

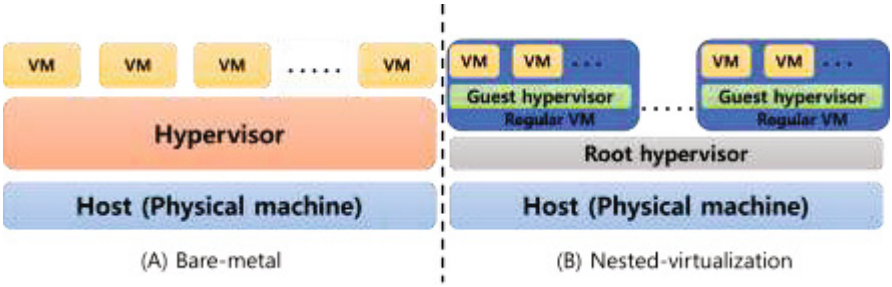


Fig. 2. Virtualization layers: bare-metal (A) and nested-virtualization (B)

Henceforth, we use the following terms to refer to different kinds of machines including host and virtual machines.

- **Physical host machine**: a physical machine that provisions virtual machines and can directly access physical resources. It resides in between physical resources and the host operating system.
- **Regular virtual machine**: a virtual machine running on top of the physical host machine (layer-1 virtual machine). It also has the hypervisor that runs other virtual machines on top of itself.
- **Nested virtual machine**: a virtual machine running on top of the regular virtual machine (layer-2 virtual machine). It does not provision another virtual machine from itself in our architecture.

The purpose of the performance comparison is to figure out how much the performance difference is between regular virtual machines and nested virtual machine. To this end, we constructed a cloud architecture using nested virtualization as shown in Figure 2(B). Table 1 shows the specification of physical host machine in our experimental cloud architecture.

Table 1. Specification of physical host machine

Category	Specification
CPU	Intel Core i5-2500 CPU @ 3.30GHz, quad core
RAM	8GB
HDD	SAMSUNG HD502HM 500GB
OS	Linux (x86_64, kernel version: 3.9.6-200)
Hypervisor	KVM

Table 2 shows the configuration of virtual machines (regular virtual machine and nested virtual machine). Note that the regular virtual machine has the hypervisor (KVM) and the nested virtual machine does not. To evaluate the performance of regular virtual machines and nested virtual machines, we use SPEC CPU 2006 benchmark programs, such as bzip2, mcf, gobmk, hmmer, sjeng, libquantus, h264ref, and omnetpp.

Table 2. Configuration of virtual machines

Category	Regular Virtual Machine	Nested Virtual Machine
Number of VCPU	4	2
RAM	8GB	2GB
Size of Disk Image	100GB	20GB
OS	Linux (x86_64, kernel version: 2.6.32-358)	Linux (x86_64, kernel version: 2.6.32-358)
Hypervisor	KVM	-

3.2 Performance Results

To compare the performance of each layer of virtual machines and physical host machine, we performed the aforementioned benchmark programs on the machines. Figure 3 shows performance comparisons of benchmark programs on physical host machine, regular virtual machine, and nested virtual machine. Note that when performing benchmark programs on a regular virtual machine, we let the physical host machine remain idle. Likewise, when performing benchmark programs on a nested virtual machine, we let the physical host machine and the regular virtual machine remain idle.

When comparing the results, the average execution time of the nested virtual machine is about 400.2 seconds, which takes about 7% longer than that of the regular virtual machine (about 374.2 seconds) and takes about 12% longer than that of the physical host machine (about 357.1 seconds). Even though there exists some degree of performance of overheads on each virtualization layer, the costs are negligible in some scenarios. Indeed, as the virtualization layer increases, some attack surfaces on the virtualization layer can be eliminated.

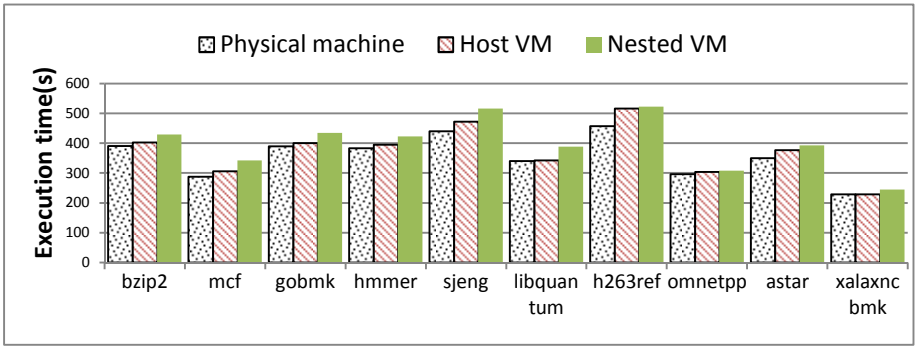


Fig. 3. Performance comparisons of benchmark programs on physical host machine, regular virtual machine, and nested virtual machine

Figure 4 shows performance comparisons of benchmark programs as the number of nested virtual machine increases from one to three. The nested virtual machines (up to three) are configured to perform benchmark programs simultaneously sharing the physical resources at the same time. Note that in this experiment, the number of regular virtual machine is one. In other words, nested virtual machines (up to three) run on the regular virtual machine.

The performance results show that when the number of nested virtual machine is two, the average execution time is about 537.9 seconds, which takes about 34% longer than that when the number of nested virtual machine is one. When the number of nested virtual machine is three, the average execution time is about 639.7 seconds, which takes about 60% longer than that when the number of nested virtual machine is one and takes about 19% longer than that when the number of nested virtual machine is two. Obviously, as the number of nested virtual machine increases, the performance degradation occurs. However, the portion of performance degradation increases linearly, as the number of nested virtual machine increases.

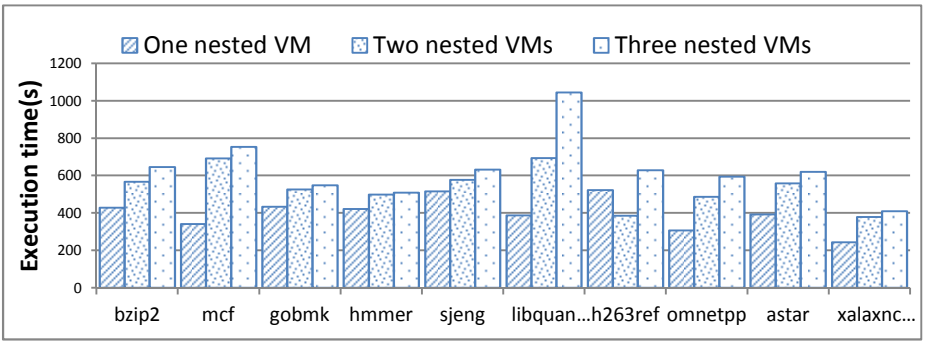


Fig. 4. Performance comparisons of benchmark programs as the number of nested virtual machine increases from one to three

Figure 5 shows performance comparisons of benchmark programs as the number of regular virtual machine increases from one to two. Note that in this experiment, the number of nested virtual machine is one and two, when the number of regular virtual machine is one and two, respectively.

When comparing the results with the previous experiments, the phenomenon is similar to the previous ones. As the number of regular virtual machine increases, the portion of performance degradation increases linearly, not doubling the execution time. When the number of regular virtual machine is two, the average execution time is about 470.2 seconds, which takes about 17% longer than that when the number of regular virtual machine is one.

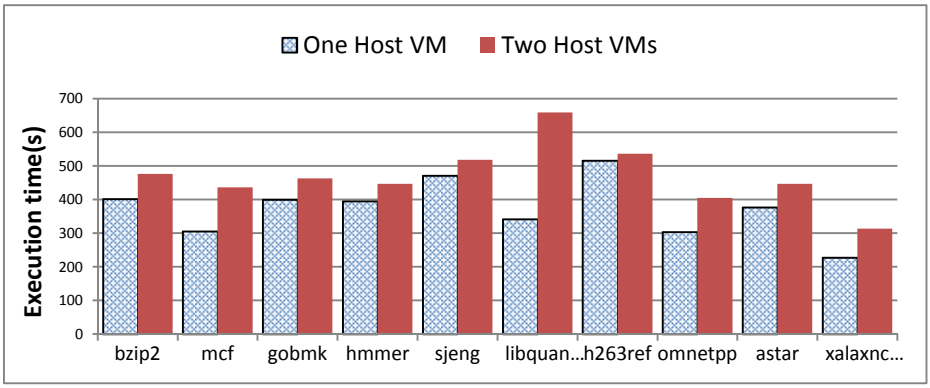


Fig. 5. Performance comparisons of benchmark programs as the number of regular virtual machine increases from one to two

4 Conclusion

In this work, we constructed a fault tolerant cloud architecture against attack surfaces in virtualized environments using nested virtualization. To figure out the performance overheads of nested virtualization compared to single-layer virtualization and physical host machine, we performed various benchmark programs of SPEC CPU 2006. Although the presented cloud architecture involves some performance costs due to an additional virtualization layer, the benefits of this abstraction layer do not always outweigh the costs. Future work includes exploring the performance overheads in cluster environments using the same cloud architecture described in this paper, and performing parallel processing benchmark programs such as PARSEC [8] in cluster environments. In addition, we will focus on performance benefits on cloud architectures in many-core computing environments using nested virtualization. Furthermore, we will construct a set of realistic hostile environments, in which one virtual machine could subvert other virtual machines, to demonstrate the security benefits of our cloud architecture.

References

1. Ben-Yehuda, M., Day, M.D., Dubitzky, Z., Factor, M., Har'El, N., Gordon, A., Liguori, A., Wasserman, O., Yassour, B.-A.: The turtles project: design and implementation of nested virtualization. In: Proceedings of the 9th USENIX Conference on Operating Systems Design and Implementation, pp. 1–6. USENIX Association, Vancouver, BC (2010)
2. Ristenpart, T., Tromer, E., Shacham, H., Savage, S.: Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds. In: Proceedings of the 16th ACM Conference on Computer and Communications Security, pp. 199–212. ACM, Chicago (2009)

3. Gruschka, N., Iacono, L.L.: Vulnerable Cloud: SOAP Message Security Validation Revisited. In: Proceedings of the 2009 IEEE International Conference on Web Services, pp. 625–631. IEEE Computer Society (2009)
4. Ormandy, T.: An empirical study into the security exposure to hosts of hostile virtualized environments. In: Proceedings of the CanSecWest Applied Security Conference, pp. 1–10 (2007)
5. Keller, E., Szefer, J., Rexford, J., Lee, R.B.: NoHype: virtualized cloud infrastructure without the virtualization. SIGARCH Comput. Archit. News 38, 350–361 (2010)
6. Raj, H., Schwan, K.: High performance and scalable I/O virtualization via self-virtualized devices. In: Proceedings of the 16th International Symposium on High Performance Distributed Computing, pp. 179–188. ACM, Monterey (2007)
7. Henning, J.L.: SPEC CPU2006 benchmark descriptions. SIGARCH Comput. Archit. News 34, 1–17 (2006)
8. Bienia, C., Kumar, S., Singh, J.P., Li, K.: The PARSEC benchmark suite: characterization and architectural implications. In: Proceedings of the 17th International Conference on Parallel Architectures and Compilation Techniques, pp. 72–81. ACM, Toronto (2008)

Recognizing Text in Low Resolution Born-Digital Images

Minh Hieu Nguyen, Soo-Hyung Kim, and Gueesang Lee*

Department of Electronics and Computer Science,
Chonnam National University, Gwangju, South Korea
hieunp131@gmail.com, {shkim,gslee}@jnu.ac.kr

Abstract. Since born-digital images usually have low resolution, they are distinctly different from natural scene images. Extracting text information from born-digital images has been an increasing interest in document analysis and recognition field. We propose an automatic method to recognize word from low-resolution color image. First, the image is smoothed by using the bilateral filter, which preserves edge information. Then, it is binarized using global thresholding method and cleaned from noise. Finally, the open source Optical Character Recognition engine, with the incorporation of a post-processor trained on knowledge of English language, is applied to obtain meaningful words from the binary image. We experiment the proposed system on ICDAR 2011 and music sheet dataset, and the result shows better performance than several previous works.

Keywords: text recognition, born-digital image, low resolution.

1 Introduction

With the development of Internet and digital devices, images are frequently used to embed textual information. The use of images as text carriers stems from a numerous online situations. In some cases, images are used as the decorations (titles, headings), or to attract attention (advertisements). In other cases, images can be utilized to hide information (images in spam emails used to avoid text-based filtering), even to distinguish a human apart from a computer (CAPTCHA tests).

For that reason, automatically extracting text from born-digital images is an interesting study as it would provide the preliminary for a number of applications such as improved indexing and retrieval of Web content, enhanced content accessibility, content filtering for advertisements or spam emails, etc.

Some examples of born-digital images can be seen in Fig. 1. On the exterior, born-digital text images are pretty similar to real scene text images. The fact is they are distinctly different. While scene text images captured by camera are high-resolution, born-digital images are inherently low-resolution with text digitally overlaid on. Born-digital images are usually made in small sizes for the purposes of transferring online and displaying on a monitor. Furthermore, born-digital images might suffer

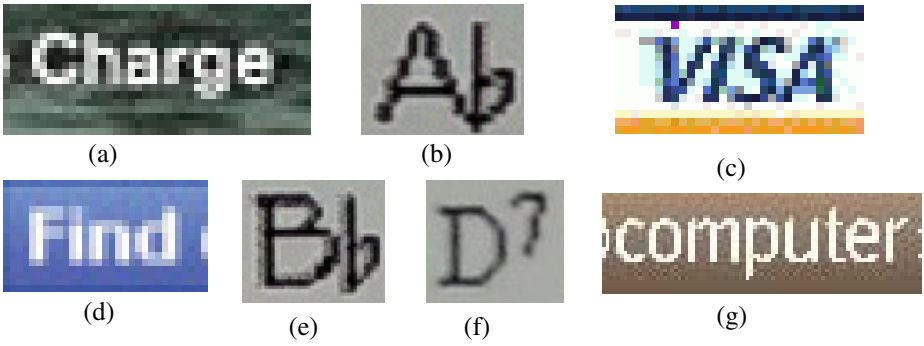
* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_12,

from compression artifacts and severe anti-aliasing, but they don't have the illumination and other problems of scene images. Therefore it is not necessarily true that methods developed for one domain would work in the others [1].

In comparison to publications on text extraction from complex images, which are obtained from video frames, real-scenes, book and magazine covers, there has been little work published specifically focused on born-digital images. One particular interest in the born-digital domain is the application of image-spam filtering, and there have been a few approaches that attempt to classify email images as legit or spam [2]. Those approaches are generally based on features such as the scope of text in the image, so they stop at the bounding box images of text localization step, rather than attempting to extract and analyze the textual content [3]. Nevertheless, the benefits of using extracted textual content from spam images for filtering are substantial [4], even if the successful recognition rate is limited.

In the rest of the paper we present all stages of our system in order to recognize text from born-digital images. They include pre-processing, binarization and post-processing. The related works is summarized in Section 2. Section 3 describes the proposed system. In Section 4, before results are shown, we briefly examine the dataset and performance evaluation methodology. We conclude this paper by giving some conclusions and ideas for future works in Section 5.



(a) (b) (c) (d) (e) (f) (g)

Fig. 1. Examples of low resolution born-digital images

2 Related Works

Recently, many methods have been proposed or experimented on born-digital word images. This section will summarize some methods that have the same input and output with our proposed system.

In "TH-TextLoc / TH-OCR" by C. Yang, C. Liu and X. Ding, the algorithm focused on text binarization rather than on the OCR. Adaptive local binarization method was used to create coarse text components followed by morphological opening to separate consecutive connected characters. For recognition, TH-OCR 2007 engine [5] was used. In [6], the approach first uses the L0 norm smoothing to increase the edge contrast of the input web images. The images were then binarized on each

color channel by the method in [7]. A connected component analysis was followed to identify the possible character components. Finally, the candidate characters were recognized by Tesseract [8] and ABBYY OCR engines [9].

In [10], the paper showed the improvement in image binarization, and the consequent increase in the recognition performance of OCR engine on the word image. The algorithm was exhaustively experimented by varying the gamma and stroke width threshold value. Commercial Omnipage OCR was applied on the images after being processed by the binarization algorithm. In [11], single characters were recognized using a classification approach based on K-Nearest Neighbors (KNN) and gradient direction features. In [12], a skeleton-based binarization method was developed in order to separate text from background. ABBYY and Tesseract OCR were used.

3 Proposed Method

3.1 Overview

The flowchart of our method is shown in Fig. 2. There are 4 main stages: Preprocessing, Binarization, Post-binarization, and applying the OCR.

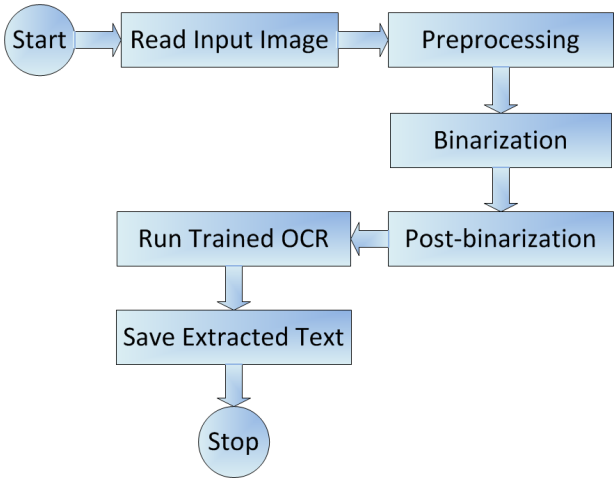


Fig. 2. System flowchart

3.2 Preprocessing

Technologies from standard OCR are often customized for high resolution documents, while text in born-digital images comes in different resolutions and with heterogeneous background. Therefore, in the first stage, the input image is converted to grayscale. To resolve the problem of low-resolution images, the image with height

smaller than 200 pixels will be scaled up 4 times via bi-cubic interpolation. Then, two dimensional bilateral filtering is used for edge-preserving smoothing [13].

3.3 Binarization

In the second stage, due to its simplicity and considerable performance, Otsu global thresholding is selected to binarize the grayscale image (Fig. 3). The algorithm assumes that the image to be thresholded contains two classes of pixels or bi-modal histogram (e.g. foreground and background) then calculates the optimum threshold separating those two classes so that their combined spread (intra-class variance) is minimal [14]. The optimal threshold value k* is found at by maximizing the following objective function

$$\sigma^2(k*) = max_k \frac{[\mu_T \omega(k) - \mu(k)]^2}{\omega(k)[1-\omega(k)]} \quad (1)$$

$$\omega(k) = \sum_{i=1}^{k} p_i \;\; ; \;\; \mu(k) = \sum_{i=1}^{k} i p_i \;\; ; \;\; \mu_T = \sum_{i=1}^{L} i p_i, \quad (2)$$

where **L** is the total number of gray levels, **i** is the i_{th} histogram bin and $\mathbf{p_i}$ is the normalized probability distribution obtained from the histogram of the image.

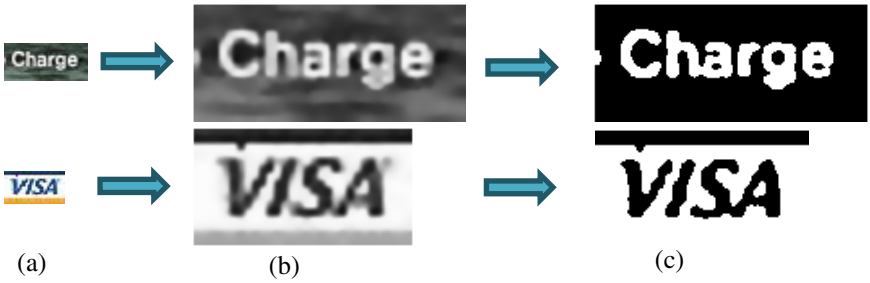


Fig. 3. (a) Original image (b) Rescaled and Bilateral filtered (c) Binarized image

3.4 Post-binarization

In post-binarization stage, we apply a programmed technique that does not require any human intervention. This technique makes sure that the binary image always has black foreground text on white background, regardless of foreground-background intensity. Next, to clean isolated, small and border noises, a modified version of shrink filter [15] is passed through binary image. Fig. 4 shows the output of this step.

Description of Inverting Logic
We consider pixels at each corner and each middle point of four boundaries.

$$P = \{1,1;\ \mathrm{h},\mathrm{w};\ 1,\mathrm{w};\ 1,\tfrac{\mathrm{w}}{2};\ \mathrm{h},1;\ \mathrm{h},\tfrac{\mathrm{w}}{2};\tfrac{\mathrm{h}}{2},1;\tfrac{\mathrm{h}}{2},\mathrm{w}\}, \quad (3)$$

where h and w are the height and width of the binary image. In set P, let W = the number of white points, B = the number of black points. If $B > W$, then the binary image is inverted.

Remove Isolated, Small and Border Noises

Performing a threshold on image often introduces a certain amount of error that can be corrected through a post processing step. This step is applied on the binary image which results from previous step in order to eliminate noises in the background and improve the quality of foreground text. For this purpose, we perform a modified version of shrink filtering [15]. An *NxN* sliding window is considered around each foreground pixel in the binary image. If number of foreground pixels in the sliding window is smaller than a threshold or this pixel belongs to borders (since text pixels are close to center), then it is turned into background. The process is described below.

Let $I_B(x,y)$ denoted the binary image at pixel (x,y) obtained from previous stage. For each border b, the following condition must be satisfied:

$$IF\ I_B(x,y) = 1 \quad AND\ \sum_{ky=y-w}^{y+w} \sum_{kx=x-w}^{x+w} I_B(kx,ky) < T_1 \quad OR \quad d\big((x,y),b\big) \leq T_2\ THEN\ I_B\ (x,y) = 0, \tag{4}$$

where $N = 2w+1$ is window size, T_1 denotes the threshold for pixel density in the sliding window, d is the Euclidean distance, T_2 is the border space. In our implementation, the value of parameters in turn is: w=2, T_1=T_2=12.

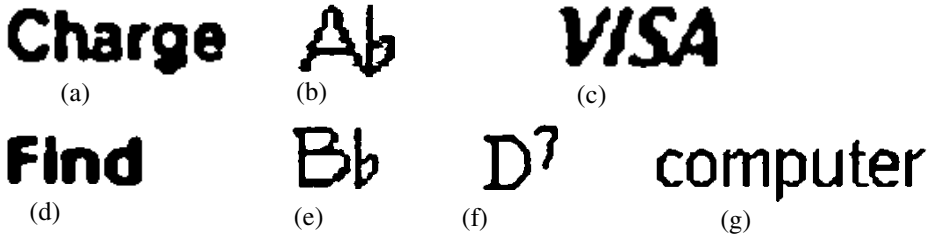


Fig. 4. Binary image after post-binarization

3.5 Optical Character Recognition

We develop a novel post-processor, which applies statistical knowledge of English language, and integrate it into Tesseract OCR open source engine [8]. The post-processor limits the character set used by Tesseract to the following characters:

{0123456789abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ.,}

Tesseract uses dictionary files for dealing with languages. These are all optional, and help Tesseract to decide the likelihood of different possible character combinations. In this implementation, the *eng.word-dawg* file contains 101,532 words of English language; *eng.freq-dawg*, which includes the most frequent words, has 98

words. The *unicharambigs* file is also useful to avoid possible ambiguities between characters or sets of characters. A part of *unicharambigs* file is described in *Table 1*.

Table 1. Sample lines from *unicharambigs* file

2 ' ' 1 " 1	A double quote (") should be substituted whenever 2 consecutive single quotes (') are seen.
1 m 2 r n 0	The characters 'rn' may sometimes be recognized incorrectly as 'm'.
3 i i i 1 m 0	The character 'm' may sometimes be recognized incorrectly as the sequence 'iii'.

The recognition result is improved after this step. Though a huge amount of training text for dictionary is necessary, once it's trained, the post-processor can correct errors in OCR output. This post-processor is built to find real words, which are in the dictionary, that have been misrecognized as similar-looking non-words. The dictionary is built from popular English words and ICDAR training set to use with OCR engine. Fig. 3 shows some words corrected by this post-processor.

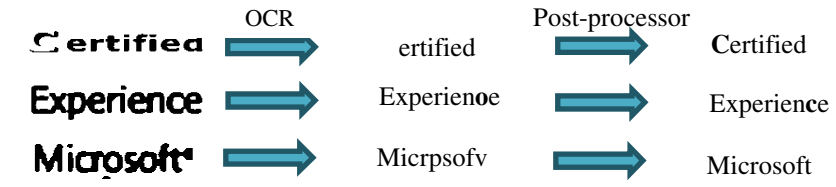


Fig. 5. OCR to post-processor. Examples of corrected words

4 Experimental Results

4.1 The Dataset

The born-digital dataset is created by ICDAR 2011 [1] competition for the Word Recognition task. It includes images extracted from different types of 412 HTML documents such as Web pages, spam and newsletter email. The test set has 918 images. These are cut from full images, with a border of 4 pixels to maintain the context. Besides, we also experiment our method on bounding box images extracted from full music sheet images.

4.2 Performance Evaluation Methodology

For each word image the competition requests a single transcription, which was compared to the ground truth transcription based on a simple edit distance metric

(equal weights for additions, deletions and substitutions). The edit distances calculated are normalized by the length of the ground-truth transcriptions and then summed over the 918 words of the test set. This is the primary metric they use for ranking. In addition to the edit distance, there's also completeness the percentage of words that were correctly recognized (no errors).

$$Correctly\ Recognized\ Words = \frac{number\ of\ correctly\ recognized\ words}{number\ of\ words\ in\ ground\ truth} \quad (6)$$

4.3 Experimental Results

Overall, our proposed method's performance is lower than methods from two other authors. Nevertheless, in few situations of skew or inhomogeneous text, our proposed method gets better results. Those cases haven't been presented here, due to the limited length of a conference paper.

Table 2. Comparisons methods of text recognition on ICDAR 2011 [1] data set

Participant	Method	Total Edit distance	Correctly Recognized Words
Yanghaojin [12]	VOCR	106.6	82.03 %
Deepak Kumar [10]	Power-law Transformation for Enhanced Recognition	108.7	82.9 %
MinhHieu	280513	186.6	72.88 %
Bolan Su [6]	proposed_abbyy	188.2	72.77 %
Chen Yang	TH-OCR 2007	189.1	61.98 %
Alvaro [11]	Alvaro Gonzalez	226.8	66.88 %
Baseline [9]	Baseline	231.2	63.94 %

5 Conclusions

In this paper, we propose an automatic method for recognizing text in born-digital images. After the image is filtered and binarized, we apply the logic to turn foreground text of binary image into black, as the requirement before feeding into OCR engine. Then, the binary image is cleaned from isolated, small and border noises. Later, the Tesseract OCR integrated with the post-processor can recognize the text from treated binary image.

With a combination of preprocessing, binarization and post-processor, which is different from existing methods, experimental results of the proposed method are better than several previous works, including [6], [9], [11], TH-OCR 2007. However, the method fails in certain cases (Fig. 7). This makes the recognition rate lower than [10], [12]. In future work, we would improve the binarization step for better binary image, mainly in case of difficult-to-binarize color images or there are touching characters.

Fig. 6. Failed cases

Acknowledgment. This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST)(2012-047759). Also this research was supported by the MSIP(Ministry of Science, ICT&Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (NIPA-2013-H0301-13-3005) supervised by the NIPA(National IT Industry Promotion Agency).

References

1. Karatzas, D., Robles Mestre, S., Mas, J., Nourbakhsh, F., Pratim Roy, P.: ICDAR 2011 Robust Reading Competition - Challenge 1: Reading Text in Born-Digital Images (Web and Email). In: 11th International Conference of Document Analysis and Recognition (ICDAR 2011), pp. 1485–1490. IEEE CPS (2011)
2. Leavit, N.: Vendors Fight Spam's Sudden Rise. IEEE Computer 40(3), 16–19 (2007)
3. Aradhye, H.B., Meyers, G.K., Herson, J.A.: Image analysis for efficient categorization of image-based spam e-mail. In: 8th Int. Conf. on Document Analysis and Recognition, pp. 914–918 (2005)
4. Fumera, G., Pillai, I., Roli, F.: Spam fitlering based on the analysis of text information embedded into images. In: 1st Int. Symposium on Information and Communication Technologies, pp. 291–296 (2003)
5. TH-OCR, `http://www.wintone.com.cn/en/Products/detail118.aspx`
6. Su, B., Lu, S., Phan, T.Q., Tan, C.L.: Character Extraction in Web Images for Text Recognition. In: International Conference on Pattern Recognition (2012)
7. Sauvola, J., Pietikainen, M.: Adaptive document image binarization. Pattern Recognition Journal 33(2), 225–236 (2000)
8. Tesseract-OCR, `http://code.google.com/p/tesseract-ocr/`
9. Commercial ABBY FineReader OCR, `http://finereader.abbyy.com/`
10. González, A., Bergasa, L.M.: A Text Reading Algorithm for Natural Images. Image and Vision Computing 31(3), 255–274 (2013)
11. Kumar, D., Ramakrishnan, A.G.: Power-law Transformation for Enhanced Recognition of Born-Digital Word Images. In: Conf. on Signal Processing and Communications, SPCOM (2012)
12. Yang, H., Quehl, B., Sack, H.: A Framework for Improved Video Text Detection and Recognition. Journal of Multimedia Tools and Applications (2012) ISSN: 1380-7501, ISSN: 1573-7721
13. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: International Conference on Computer Vision, pp. 839–846 (1998)
14. Otsu, N.: A threshold selection method from grey level histogram. IEEE Transactions on System, Man and Cybernetics 9, 62–66 (1979)
15. Schilling, R.J.: Fundamentals of Robotics Analysis and Control. Prentice-Hall, Englewood Cliffs (1990)

CICC to Support Location Based Service in Cloud Computing

Doohee Song and Kwangjin Park*

Department of Information and Communication Engineering, Wonkwang University, Iksan-shi, Chunrabuk-do 570-749, Republic of Korea
songdoohee@naver.com, kjpark@wku.ac.kr

Abstract. Recently, location-based services (LBS) which utilizes in mobile computing environment has been increased. The examples are Foursquare, Facebook places, Google buzz, and etc. After LBS are proliferated, the demand of various services has increased, resulting in bigger and more complex spatial data. The increase in the volume of data due to increases in the number of queries from clients and spatial queries overload servers and cause delays in query processing time. Thus we are using a cloud computing environment to store big data to overcome limited data spatial storage in more efficient way. The characteristics of cloud computing is to store data via the Internet based common computer (e.g., icloud, dropbox, etc) instead of a personal computer.

Then, cloud computing can process efficiently not only a personal spatial query but also in a large number of common spatial query processes. In addition, cloud computing has the other advantage. The advantage is that suppose each client sets a different mobile OS (e.g., Android, Apple OS, Window Mobile (MS), or others); these systems can be compatible within cloud computing. The purpose of this paper is going to introduce the simple general overview of spatial data method via cloud computing and propose the Common Issue Cloud Computing (CICC) for a spatial query type. An experimental result is used to verify the efficacy of our CICC.

Keywords: location based service, spatial query, cloud computing, common issues cloud computing (CICC), average query processing time.

1 Introduction

Dispreading, smartphone-related technologies are advancing at high speed. In general, information providing methods by server to client are largely divided into 2 methods [1, 2]. First, the on-demand is to send and receive information as 1 to 1 communication, and a server returns result values to client when a client requests query. 1 to 1 communication makes more effective to process query effectively when the number of queries is small and the size of data to be processed is also small. On the contrary to this, when the number of queries from clients is increasing or data size

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*, Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_13,

to be sent to clients is getting bigger, load of server is increasing, which in turn makes query time extended. The second one is 1 to multiple communication with server and clients for wireless broadcasting (Client is not able to send data to server).

The strength of this method is that it's possible to send common issues (e.g., weather, traffic information, popular sports) to unspecified multiple clients within communication radius of server. In other words, the load of server in wireless broadcasting is not increasing even if the number of queries from clients is increasing. However, the data amount to be sent to clients from server is increasing as number of common issues is increasing and broadcasting cycle is also increasing, which in turn makes query processing time longer. Therefore, we, hereby propose spatial query processing using cloud computing to solve above two problems.

Cloud computing is a method to process information using other computers which are connected via internet [3]. The strength of this technique is that people can use software through cloud computing even if there is no software in my smart phone. Moreover, there is an environment which makes it possible for multiple clients to access cloud computing depending on assigned rights from server not just for individuals. We build Common Issues Cloud Computing (CICC) for each common issue depending on queries requested by clients through cloud computing which multiple clients can use and then propose system model for clients which can process queries by selecting CICC depending on queries that they requested. Thus CICC is able to reduce load of server. Main contribution is following:

1. As server doesn't have to send duplicated query results through CICC, it is able to reduce load of server.
2. As server is able to identify objects using CICC-tree, effective searching pruning becomes available. Therefore, query processing time can be reduced.
3. We prove through experimental results that CICC proposed by us is better than existing methods.

The remaining of this paper is organized as follows: Section 2 reviews the previous related work; Section 3 explains the algorithm of our scheme and CICC; Section 4 shows the experimental result using dataset; finally, Section 5 concludes this paper with directions for the future.

2 Related Works

2.1 Spatial Query Processing Based on Various Kinds of Environments

As we mentioned from Section 1, an environment to process spatial query is on-demand and wireless broadcasting [4-8]. First of all, research for spatial query processing using wireless broadcasting is following. [4] searches nearest neighbor (NN) under wireless environment. Voronoi diagram is featured in a way that it forms Voronoi cell for each object and an object nearest those which conduct query within Voronoi cell is an object existing within Voronoi cell. However, the use of pre-computed Voronoi diagram causes higher cost as Voronoi diagram is reconstructed

for its update, which is ineffective. [5] proposes a technique to search continuous NN (CNN) under broadcasting environment combining R-tree [9]. This technique processes spatial query by constructing algorithm based on Hilbert Curves [10]. As Hilbert Curve expresses the sequence of grid in map linearly with major space filling curve, it is frequently used for wireless broadcasting environment. However, when Hilbert Curve recognizes its own location using Hilbert Curve, it should read grid of map by its order, which in turn increases access time. It's a weakness. Besides this, various researches using wireless broadcast are under processing [6, 7].

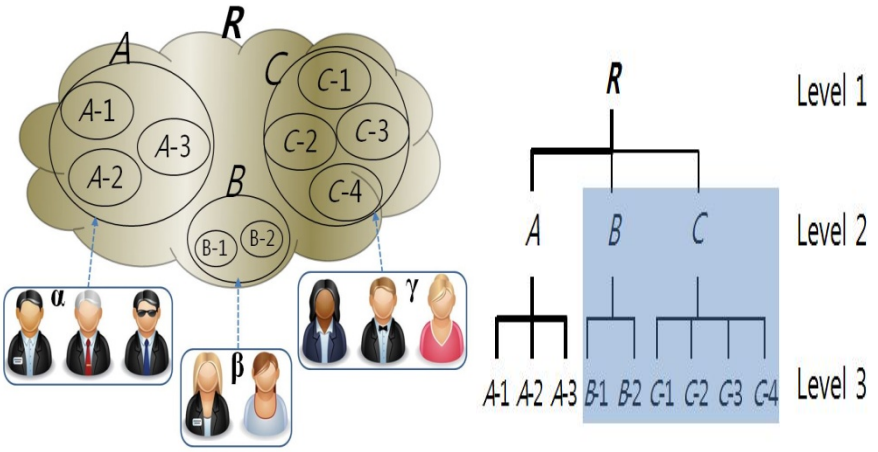
A research for on-demand spatial query process is following [8] proposes a technique to recognize the future location of object while moving object's direction and speed are uncertain. This technique is able to search NN effectively using time parameterized R-tree. However, as this technique processes queries based on uncertain direction and speed, it sets the up movement radius of object by itself. Therefore, in order to recognize the location of an object, this technique should consider all unnecessary data for objects.

[11] is defined it as " Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models". And there is a study on location-based mobile application in cloud computing [12]. [12] proposes a framework to protect privacy (e.g., individual location information) [13] when a user requests spatial queries. However, it didn't consider query processing time for actual spatial query processing.

3 Common Issue Cloud Computing (CICC)

Fig. 1 (a) shows a diagram for CICC. ***R*** indicates whole cloud computing while *A*, *B* and *C* are composed common issues within ***R*** (The structure of the group α, β, γ is changed based on the common issue of a server. Also, the common issue can be divided into more detailed based on the server's capacity). For example, ***R*** indicates all weather information in Asia while *A*, *B* and *C* indicate weather information in Japan, China and Korea respectively. In this case, the set of the group α finds weather in Korea. Also it is able to verify locations (e.g., Seoul or Busan, Korea) and other various information in Korea. The tree structure for information searching method is shown in Fig. 1 (b). Fig. 1 (b) shows a searching path (bold line) through which the set of the group α attempts to find a Korea weather. First of all, search ***R*** of CICC and then check *A*-1, *A*-2 and *A*-3 through *A*. A formula of average query processing time for CICC searching method is shown as following [2].

Average query processing time = $\frac{1}{2}$ (the number of CICC * data size (include index)) / transmission speed (include bandwidth, etc.) (1)



(a) CICC diagram (b) CICC-tree structure

Fig. 1. CICC diagram and CICC-tree structure

Fig. 1 is a tree structure for brief description of this paper. However, for actual tree structures, n trees are made not 3 trees: *A*, *B* and *C*. And the level of tree is also increasing up to n number. In this case, effective searching pruning is possible depending on level condition of client. This is expressed by algorithm as following.

Algorithm 1. CICC node search method

input: all node of CICC information (Node n; N_n)
output: result value of requested in the query
01: querier is select CICC-tree level x (e.g., R, A, A-1, · · ·(refer to Fig. 1))
02: From level x node (N_x) of CICC-tree search
03: **for** from search node of level x to final leaf node of (N_x) $\subset N_{result}$
04: **if** (N_x=N_{result})
05: result=N_x;
06: **end if**
07: **end for**
08: **return** result;

4 Experimental Result

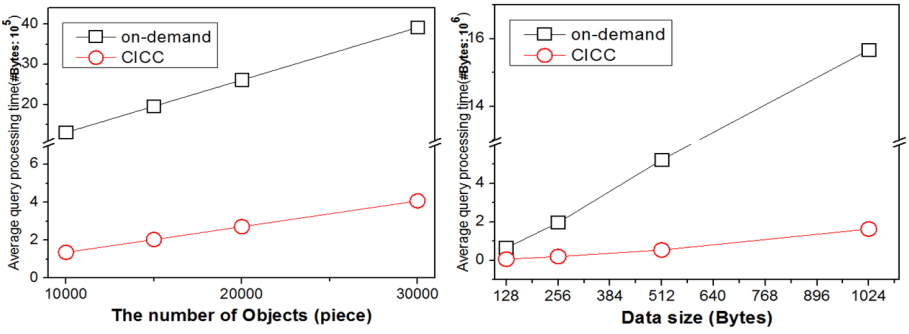
We assume that the single server (computer) uses clouding computing and performs and do query processing to a client. If the single server is increased by n, it can be applied into a multi server cloud computing environment.

Experiment environment compares existing on-demand technique with performance of CICC based on single channel (As on-demand and wireless broadcasting is different for their environment, their performances might be different.

Therefore, we exclude performance comparison for wireless broadcasting) All experiments are conducted by PC 2.9 Ghz CPU and 4 GB of main memory and C++. Set values for basic parameters to evaluate performance are shown as Table 1. The reason why y-axis in performance evaluation is used for data is that the average query processing time is varying depending on transmission speed.

Table 1. Experimental data set values

Parameter	set values
the number of CICC	(the number of client)/10
data size (Bytes)	128~1024 (basic : 256)
the number of client	10000~30000 (basic :10000)
CICC level n	1



(a) (b)

Fig. 2. The query processing time compared to the number of clients and data size

From the experiment environment in Fig. 2(a), the number of clients is increasing up to 10000, 15000, 20000 and 30000 when CICC level is 1 and data size is 256 Bytes. In this case, this experiment compares on-demand with the performance of CICC depending on the number of clients. After this comparison for each technique's performance, it's proved that the performance of CICC is improved by approximately 21% than on-demand. The reason why the graph of on-demand is increasing than CICC as the number of client is increasing is that query processing time is increasing due to increasing number of clients with 1 to 1 data transmission of server.

From the experiment environment in Fig. 2(b), data is increasing up to 128, 256, 512 and 1024 Bytes when the level of CICC is 1 and the number of clients is 10,000 people. In this case, the experiment compares on-demand with the performance of CICC depending on data size. After this comparison for each technique's performance, it's proved that the performance of CICC is improved by approximately 21.7% than on-demand. The reason why the graph of on-demand is increasing than CICC as the number of client is increasing is that query processing time from server is increasing due to increasing data when server sends query results to each client.

5 Conclusion

We conducted spatial queries using cloud computing to supplement weakness found from existing on-demand and wireless broadcasting environment. We could reduce the load of server as server doesn't send duplicated query result the proposed CICC. Moreover, as client verifies objects using CICC-tree, effective pruning is possible and consequently it can reduce query processing time. As the comparison with existing on-demand and CICC, it's proved that CICC proposed by us is better than existing technique. We will extend our study into really cloud computing experiments using the proposed system model in the future.

References

1. Gao, Y., Zheng, B., Chen, G., Li, Q., Guo, X.: Continuous visible nearest neighbor query processing in spatial databases. Very Large Data Bases Journal 20(3), 371–396 (2011)
2. Imielinski, T., Viswanathan, S., Badrinath, B.R.: Data on air: Organization and access. IEEE Trans. Knowledge & Data Eng. 9(3), 353–372 (1997)
3. Velte, T., Velte, A., Elsenpeter, R.: Cloud Computing: A Practical Approach. McGraw-Hill (2009)
4. Zheng, B., Xu, J., Lee, W.-C., Lee, D.L.: Grid-partition index: a hybrid method for nearest-neighbor queries in wireless location-based services. Very Large Data Bases Journal 15(1), 21–39 (2006)
5. Zheng, B., Lee, W.-C., Lee, D.L.: Search continuous nearest neighbors on the air. In: Proc. of the Mobile and Ubiquitous Systems: Networking and Services, pp. 236–245. IEEE Press, Boston (2004)
6. Park, K., Choo, H., Valduriez, P.: A scalable energy-efficient continuous nearest neighbor search in wireless broadcast systems. Wireless Networks 16(4), 1011–1031 (2010)
7. Mouratidis, K., Bakiras, S., Papadias, D.: Continuous Monitoring of Spatial Queries in Wireless Broadcast Environments. IEEE Transactions on Mobile Computing 8(10), 1297–1311 (2009)
8. Tao, Y., Papadias, D.: Time-parameterized queries in spatio-temporal database. In: Proc. of the ACM SIGMOD, pp. 334–345. ACM Press, Madison (2002)
9. Guttman, A.: R-trees: a dynamic index structure for spatial searching. In: Proc. of the ACM SIGMOD, pp. 18–21. ACM Press, Boston (1984)
10. Gotsman, C., Lindenbaum, M.: On the Metric Properties of Discrete Space-Filling Curves. IEEE Transactions on Image Processing 5(5), 794–797 (1996)
11. Mell, P., Grance, T.: NIST definition of cloud computing. National Institute of Standards and Technology (2009)
12. Chen, Y.-J., Wang, L.-C.: A security framework of group location-based mobile applications in cloud computing. In: Proc. of the Parallel Processing Workshops, pp. 184–190. IEEE Press, Taipei (2011)
13. Freni, D., Vicente, C.-R., Mascetti, S., Bettini, C., Jensen, C.S.: Preserving Location and Absence Privacy in Geo-Social Networks. In: Proc. of the ACM CIKM, pp. 309–318. ACM Press, Toronto (2010)

Method for Detecting Cars Cutting in to Change Lanes by Using Image Frames Played in Reverse

Chi-Hak Lee, Hyun-Woo Kim, Inwon Lee, Eun-Ju Lee, and Young-Mo Kim

School of Electronics Engineering, Kyungpook National University, Sankyuk-dong, Buk-gu, Daegu, Korea
ymkim@ee.knu.ac.kr

Abstract. In this paper, we suggest a method for tracking cars that violate traffic laws restricting cut in lane changes. We also present experimental results, which show that the tracking method can be implemented in unmanned instruments for detecting violating car by using the image frames played in reverse. Two kinds of camera were installed as a set at the Yang-Jae IC in Korea: A recognition-camera (R_c), and A tracking-camera (T_c). The R_c reads a plate number and the T_c tracks cars in a region of interest (ROI). The T_c determines whether or not the cars violate the law by analyzing image frames played in reverse. A plate number recognition algorithm was provided by KNU DILAB, and the KLT algorithm was used for tracking cars in the ROI. Our experimental findings show that the proposed method can be applied to unmanned systems for cars that illegally cut in to change in lanes.

Keywords: Backward-playing Images, Image Processing, Cut-in Lane Changing, Tracking Object.

1 Introduction

Traffic congestion in and around urban areas is a major problem throughout the world. The duration of peak-hour congestion is increasing by the day. Congestion adversely affects mobility, safety, and air quality, which in turn. Cause direct economic losses due to delays and accidents, as well as indirect economic losses due to environmental impact. In most cases, the capacity of existing roadway systems cannot be increased by adding additional lanes owing to space, resource, or environmental constraints. Travel along the entrance lanes leading into a city can be congested or reduced in speed during peak hours. In this paper, we find that at the Yang-Jae IC on the No. 1 Express Way heading south, traffic is congested or moves at speeds as low as 3 km/h, up to 10 km/h. Although many drivers turn into the entrance lane, some attempt to enter through restricted lanes in violation of the traffic law. This violation can lead to increased traffic congestion. Several methods have been used to control such cut-in lane changing in violation of the traffic law. The most common method involves having traffic officers administer penalties to traffic offenders. However, this is problematic because it requires a great deal of manpower

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_14,

and evidence is difficult to obtain. Consequently, efforts have been made to introduce image detecting strategies to solve this issue.

A significant amount of work has been carried out in the area of detecting moving vehicles [1]–[5]. Foreground detection schemes typically have the problem of detecting the shadows of moving objects as foreground. This problem has been addressed in previous work, (e.g., [6]–[9]). Several studies have been conducted on tracking and classifying vehicles [10]–[14]. In addition important traffic parameters have been estimated in other studies [15]–[17] by using image-processing techniques.

The aim of this paper is to automatically identify cut-in lane changing through the use of imaging techniques that do not require human assistance. The methods and results of an experiment, in which cars that illegally switch lanes are automatically identified through the use of backward image frame analysis, will be explained.

The automatic tracking system consists of two image processing devices: a camera that scans vehicles and another that tracks them. The recognition camera, that scans vehicles, R_c, sends real-time information to the tracking camera, T_c, which then searches for cars that illegally switch lanes through real-time backward tracking. The T_c rotates in a plane perpendicular to the ground in order to obtain as much information as possible for imaging the road, which has numerous vertical components. The performance evaluation in the current study, which was conducted at the entrance of the Yang-Jae IC, attained a success rate of 93%. This suggests that the proposed method can be used for effectively detecting cars that violate traffic laws.

The standards established for cut-in lane changing and the methods used to detect such offenses will be the main concerns of this study. We will also elaborate on the backward image-framing method used by the T_c to detect traffic offenders and the method of applying the KLT tracking algorithm, explain the pictures, and conclude the paper by discussing some arithmetic results.

The experimental results suggest that the proposed method should detect and manage such traffic offenses without any problems.

2 Paper Preparation

2.1 Criteria Considered to Determine Cut-in Lane Changing

There are no strict legal standards on the changing of lanes. Therefore, we have deemed it an offense for vehicles to change lanes in areas where lane changing is prohibited. A potential problem is that cars might be wrongly accused if detected during a non-congested period. Hence, investigating the number of cars that pass through the area and considering the average speed rates to determine whether a vehicle has committed a traffic offense will be important in the future.

The experiment was carried out at the Yang-Jae IC in Korea (Figure 1). Figure 2 is an image of a car that changed lanes illegally; Figure 3 shows the range of view for the T_c. Figure 4 shows an image of all cars detected by an internet protocol (IP) camera. The IP camera screen was installed Yang-Jae IC to verify and judge whether cars have committed an offence.

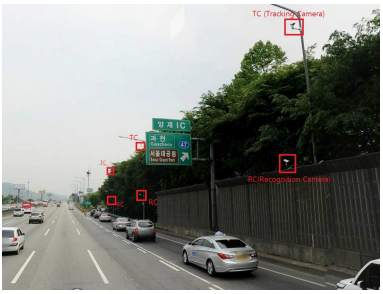


Fig. 1. Cameras at Yang-Jae IC

Fig. 2. Image recognized by R_c

Fig. 3. Image captured image by T_c

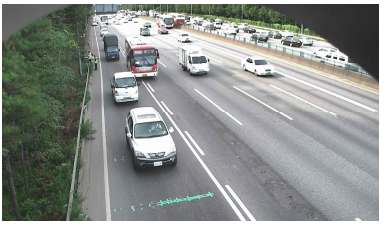


Fig. 4. Image captured by IP camera

2.2 Cut-In Lane Changing Situation

Figure 5 shows a schematic diagram that depicts lane changes considered legal or illegal in the present study. The vehicle that proceeds along the sequence of positions ⓓ →ⓒ →ⓑ →ⓐ (within the range of view of CAM2) is considered to have illegally changed lanes. The vehicle in view of CAM4, which proceeds along the sequence ④ →③ →② →①, changes lanes in an area where the lane divider is dotted. Therefore, the lane change is considered legal lane

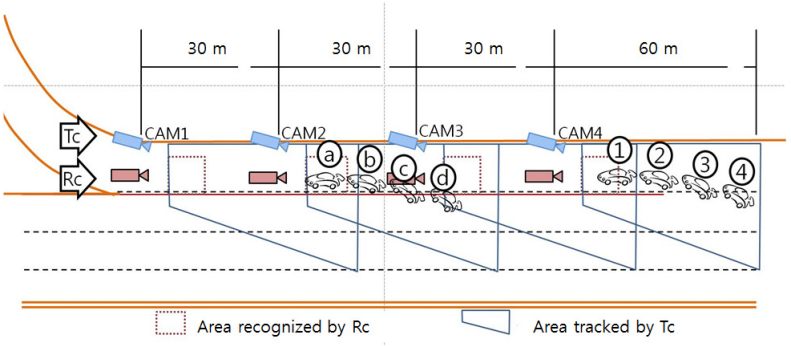


Fig. 5. Conditions of cut-in lane changing

3 Step to Detect Cut-in Lane Changing

3.1 Identifying Regions Where Cut-In Lane Changing Is Illegal

(1) R_c1 searches the number plate of the vehicle, which is saved as CAM2, CAM3 or CAM4 from the detector's number list that has passed the CAM1 region.
(2) If the number is not found in CAM2, CAM3 or CAM4, this indicates that the illegal cut-in lane change occurred in the region of CAM1.
(3) If the number exists in CAM2 and not in CAM3 or CAM4, the illegal cut-in lane change is considered to have in the CAM2 region.
(4) If the number is found in CAM2 and CAM3 but not in CAM4, the illegal change is considered to have occurred in the CAM4 region.
(5) If the number is found in all CAMs, the cut-in lane change is legal.

3.2 Tracking Offending Vehicles

(1) Confirm the region where the illegal lane change occurred.
(2) If the car changed lanes illegally in the region of CAM2(CAM1, CAM3), search for a unique feature of the car from the frame in ⓐ, which is closest to the camera.
(3) Search for the car with the unique features in frame ⓑ, which is just before ⓐ.
(4) Confirm the illegal lane change in frame ⓒ which is just before ⓑ.
(5) Obtain an image from ⓓ as evidence so that the direction of travel can be identified.

4 Image Matching Procedure

Figure 6 below shows an image, in which the locations of the R_c and the T_c. The T_c recognizes the vehicle in the matching state and tracks the vehicle as in Figure 7 to determine the final traffic offense.

Fig. 6. Image Matching – In

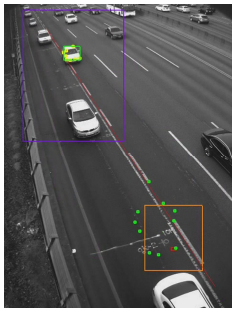


Fig. 7. Image Matching - Out

5 Tracking Algorithm

Although the world is three-dimensional (3D), images only contain a two-dimensional (2D) center, which produces 2D information. Therefore, 2D information makes up for the real 3D world. Figure 8 shows an example of how 3D information is projected onto a 2D plane. Optical flow is used to detect 3D movement, which is projected into 2D information.

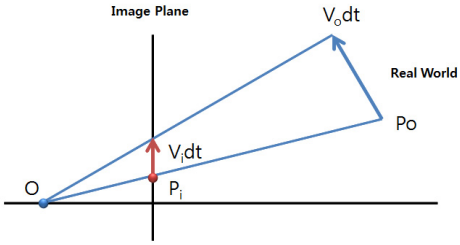


Fig. 8. Projection of 3D information onto 2D plane

The optical flow is used to track the object and to estimate the movement of segment, according to the algorithm suggested by Horn and Schunck. The movement and flow of the images are determined by two sequential images, with the least difference. The error is calculated with the following algorithm.

$$E^2(x, y) = (f_x u + f_y v + f_t)^2 + \lambda^2 (u_x{}^2 + u_y{}^2 + v_x{}^2 + v_y{}^2) \tag{1}$$

The first term of Eq. (1) indicates the change in darkness in a pixel across time and space between the two images. f_x, f_y, and f_t indicate functions of x (horizontal coordinate), y (vertical coordinate), and t (time), respectively. λ, which indicates a constant value, is used for gradual tracking. u_x, u_y, and v_x, v_y represent directional changes in u and v, i.e., the optical flow velocities along the x and y directions respectively. u and v, which are necessary for solving Eq. (1), are calculated as follows according to Horn and Schunck.

$$\begin{aligned} u &= u_{av} - f_x \frac{P}{D} \\ v &= v_{av} - f_y \frac{P}{D} \\ where &\begin{cases} P = f_x u_{av} + f_y v_{av} + f_t \\ D = \lambda^2 + f_x{}^2 + f_y{}^2 \end{cases} \end{aligned} \tag{2}$$

The optical flow is the sum of vectors *u* and *v*. The KLT algorithm, which is based on optical flow, does not change the value of the pixel's brightness even if the frame changes. Movement of the object between frames is slight, and dots near one another may belong to one object and be considered to move in the same direction. We also use an algorithm, based on the pyramid. Several Gaussian pyramids are created on the original image and the optical flow is calculated from the top pyramid. Subsequently, this calculation is repeatedly used to that the movement of the object is insubstantial [B]. Extracting unique features of the car through backward imaging techniques [C] has become easier.

Figure 9 shows the tracking of a lane-changing car using the restricted algorithm. The frames are played from the present to past, the largest image of the vehicle is at the start of tracking (bottom-left of). Therefore, the extraction of unique features is easily accomplished. The success rate of tracking the object increases even if the image of the vehicle decreases.

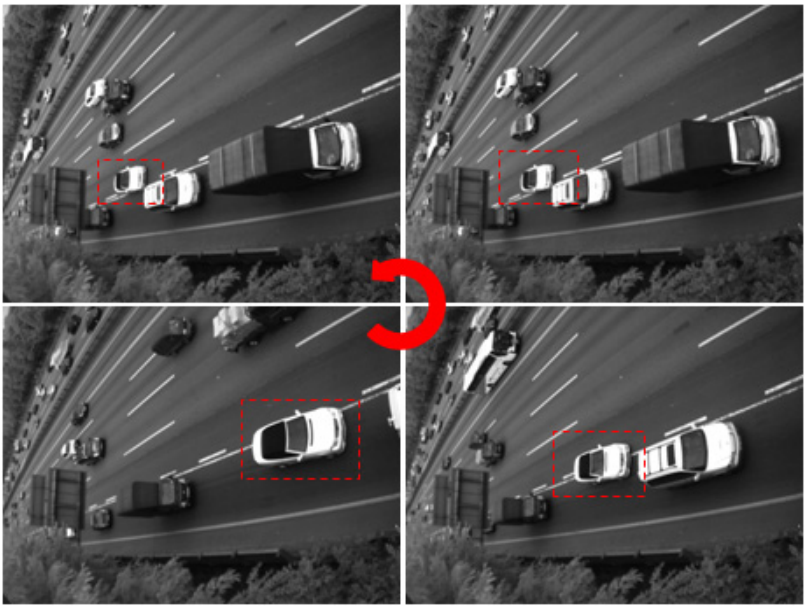


Fig. 9. Tracking a car cutting in to change lanes

6 Experiment Results

The entrance to the southbound lane toward Gwacheon on the Yang-Jae IC in Korea was used for the experiment, Cars were observed from 4:45 to 5:00 on the 8th of November 2012.

Table 1. Test result

Research Categories		Number	Percentage	Note
Total Number of Cars		1,263		Cars that passed the region
Cut-in Lane Changing		421	100%	Manual confirmation
Automatically regulated vehicles		392	93.10%	Success Percentage Rate
Failure of Automatic Regulation	Tracking Failures	14	3.30%	Disappearance of vehicles during tracking
	Unidentifiable Vehicles	15	3.60%	Temporary number plates, unclear images
	False Identification	0	0%	Legal cars falsely identified as illegal
	Subtotal	29	6.90%	Failure rate of automatic regulation

A total of 1,263 vehicles passed the experimented region, of these, 421, were identified by the human eye to be crossing lanes illegally. Three hundred ninety two were identified by the equipment this indicates an accuracy rate of 93.1%. The equipment failed to identify 14 cars, and 15 vehicles were unidentifiable because of temporary number plates or unclear images. There were no falsely accused vehicles.

7 Conclusion

The success rate of the experiment was 93.1% for 1263 cars. If we exclude the unidentifiable number plates, the success rate becomes 97.7%. Therefore, if clear regulations are set in place, the standards of the experiment indicate that an algorithm can be established to regulate traffic.

Future directions for research include experiments conducted in various weather conditions and the testing of machines to judge their reliability.

Acknowledgments. This research was supported by the MSIP(Ministry of Science, ICT & Future Planning), Korea, under the C-ITRC(Convergence Information Technology Research Center) support program (NIPA-2013/H0401-13-1005) supervised by the NIPA (National IT Industry Promotion Agency).

References

1. Nir, F., Stuart, R.: Image Segmentation in Video Sequences: A Probabilistic Approach (1997)
2. Gentile, C., Camps, O., Sznaier, M.: Segmentation for Robust Tracking in the Presence of Severe Occlusion (2004)
3. Dubuisson, M., Jain, A.: Contour extraction of moving objects in complex outdoor scenes. International Journal of Computer Vision 2, 83–105 (1995)
4. Chris, S., Grimson, W.E.: Adaptive background mixture models for real time tracking (1999)
5. Sun, Z., Zhou, W., Zhang, W.: Vehicle detecting in traffic scenes with introduction of subtractive clustering algorithm (2010)
6. Kilger, M.: A Shadow Handler in a Video-based Real-time Traffic Monitoring System (1992)
7. Prati, A., Mikib, I., Grana, C., Trivedi, M.M.: Shadow detection algorithms for traffic flow analysis: A comparative study (2001)
8. Yoneyama, A., Yeh, C.H., Kuo, C.C.J.: Moving Cast Shadow Elimination for Robust Vehicle Extraction based on 2D Joint Vehicle-Shadow Models (2003)
9. Kuxrrar, P., Sengupta, K., Lee, A.: A Comparative Study of Different Color Spaces for Foreground and Shadow Detection for Traffic Monitoring System (2002)
10. Koller, D., Weber, J., Huang, T., Malik, J., Ogasawara, G., Rao, B., Russell, S.: Toward robust automatic traffic scene analysis in real-time (1994)
11. Lipton, A.J., Haering, N.: Commode: An algorithm for video background modeling and object segmentation (2002)
12. Medioni, G., Cohen, I., Bremond, F., Hongeng, S., Nevatia, R.: Event detection and analysis from video streams (2001)
13. Kumar, P., Ranganath, S., Weimi, H.: Bayesian network based computer vision algorithm for traffic monitoring using video (2003)
14. Tao, H.T.H., Sawhney, H.S., Kumar, R.: Object tracking with Bayesian estimation of dynamic layer representations (2002)
15. Dailey, D.J., Cathey, F.W., Pumrin, S.: An algorithm to estimate mean traffic speed using uncaliberated cameras (2000)
16. Yamada, K., Soga, M.: A compact integrated visual motion sensor for ITS applications (2003)
17. Beymer, D., McLauchlan, P., Coifman, B., Jitendra, M.: A Real-time Computer Vision System for Measuring Traffic Parameters (1997)

An Efficient Video Hooking in Androidx86 to Reduce Server Overhead in Virtual Desktop Infrastructure

Tien-Dung Nguyen[1], Cong-Thinh Huynh[1], Hyun-Woo Lee[2], and Eui-Nam Huh[1,*]

[1] Computer Engineering, Kyung Hee University,
Global Campus, South Korea
[2] Electronics and Telecomunications Research Institute, Daejeon
{ntiendung,thinh,johnhuh}@khu.ac.kr,
hwlee@etri.re.kr

Abstract. Recently, Virtual Desktop Infrastructure (VDI) for mobile devices emerges as one of the key concept in Mobile Cloud Computing (MCC). Executing resource-intensive applications in virtual desktop on the cloud servers rather than on the mobile devices enables accessing any application from any location with any devices. Based on the remote display solution, the server captures its screen and delivers to client to display. However, with the increasing number of clients, resource management in cloud servers emerges as a key issue in order to enhance performance and reduce total cost. In this paper, we propose a video hooking method of remote display solution which intercepts at the surface flinger layer. By intercepting at this layer, we can disable the display function on the cloud servers, we thus reduce cpu and memory consumption on the server side. From the results of experiment, we show that our approach can reduce cpu and memory usage than others.

Keywords: Remote Display, VDI, Video Hooking, Androidx86.

1 Introduction

MCC emerges as a new mobile computing technology that leverages unified elastic resources of varied clouds and network technologies toward unrestricted functionality, storage, and mobility. MCC allows mobile devices to run heavier applications and access data faster, regardless of heterogeneous environments and platforms. With the aid of multiple processors and high-speed networks on the cloud, complicated computing and accessing big data can be executed powerfully and responsively. Since offloading computing tasks to the cloud, mobile devices do not need stringent requirements on powerful hardware. Even equipped with powerful CPUs and GPUs, mobile devices can spare the local processing power to handle other tasks such as user interface, interactive function, speech and gesture recognition, etc.

VDI is a technology based on multiple remote display solutions to achieve access to a virtual desktop on cloud. There are some approaches introduce remote display solutions that is adapted for MCC [3], [12]. Similar to solution on PC, complicated

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_15,

computing of mobile applications are executed on the cloud servers rather than on the individual mobile devices. Firstly, the demanded mobile application is configured and started on an installed-OS cloud server. The cloud server then renders screen and delivers pixel data to the mobile device. The installed-OS can be PC's OS (such as Microsoft Windows, Linux, etc.) or mobile's OS (such as Android, iOS, etc.). However, almost dedicated mobile applications which are designed only for mobile devices with multi-touching features are incompatible with PC's OS. Therefore, mobile's OS must be installed on cloud server to utilize the mass of mobile applications in the Android Market [7], Apple App Store [2].

In this paper, we choose Android [6] to install on cloud server, an open- source mobile OS initialed by Google. The main reason behind our choice is that Android OS is not only designed for mobile devices with an ARM processor, but also is being ported to the x86 platforms [1], [4]. During this paper, the scenario is described as follows: the cloud servers install Androidx86 on x86-based CPU, the clients use mobile devices on ARM-based CPUs and run mobile applications through remote display solution.

There are some methods to support remote display solution for mobile VDI such as RDP [9], VNC [10], android-vnc-server [11]. [13], [8] proposed the novel isolation technology and remote display protocol for mobile thin client to reduce the overhead of cloud server. [14] presented the hybrid display protocol for reducing both computing on cloud server and data transmission. However, these remote display protocols are basically developed on VNC technology which mainly reads pixel data from framebuffer (*/dev/fb0* in Linux OS [5]). The framebuffer *fb0* is only updated after the system executes drawing functions, saves data to buffer of graphic card and displays on the screen. Therefore, the reading at framebuffer *fb0* takes more delay and the server musts consume more CPU to display pixel data on server's screen. Suppose that there are a huge number of customers using mobile services, the reducing overhead of cloud server will increase the number of customers allocated in one server, and the cost is obviously reduced.

In this paper, we propose a remote display protocol that is capable of capturing pixel data in memory at surface flinger layer before saving into *fb0*, and prevent displaying on server's screen (off-screen mode). With this approach, not only the capturing process is faster, but also the cloud servers can save CPU/memory usage in the off-screen mode.

The rest of this paper is organized as follows. We introduce mobile cloud architecture in section 2. Section 3 presents video display in androidx86 platform. In section 4, we describe our proposed video hooking method. We finally show some performance results in section 5 and some conclusions in section 6.

2 Mobile Cloud Architecture

In this section, we introduce an overview architecture of mobile cloud supporting for mobile thin client. As Fig. 1 shows, the server infrastructure includes a number of modules. At first, the mobile device (or client) sends a request to authentication module to log in the system. Then the server selects applications from application

store, and provides the necessary information to the payment system. The information of the requested applications is sent to Task Manager module. In task manager module, it requires resource condition from QoS module and application profiles from Internal Data Repository. After analyzing all information, it decides task allocation on cloud server. Cloud Servers are physical machines equipped with high performance CPU and memory which are virtualized into multiple virtual machines (VMs). Each VM installs one mobile OS (Androidx86) and requested applications. Meanwhile, QoS module continuously measures the network condition as well as the cloud server resource to detect QoS violation. The internal data repository is used to save or load user sessions. Meanwhile the external data resource provides the external data for VMs whenever they need.

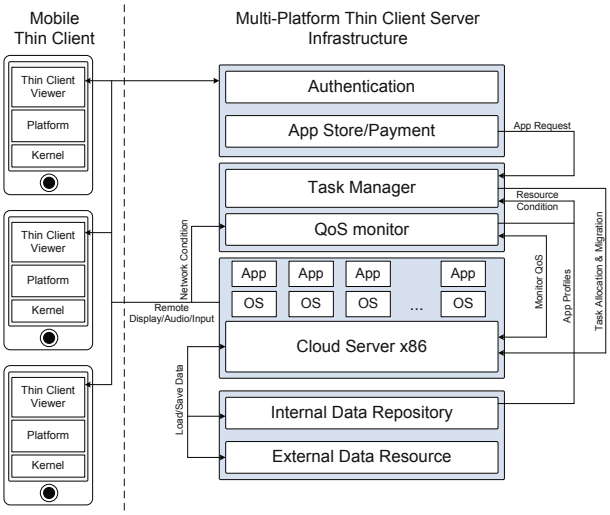
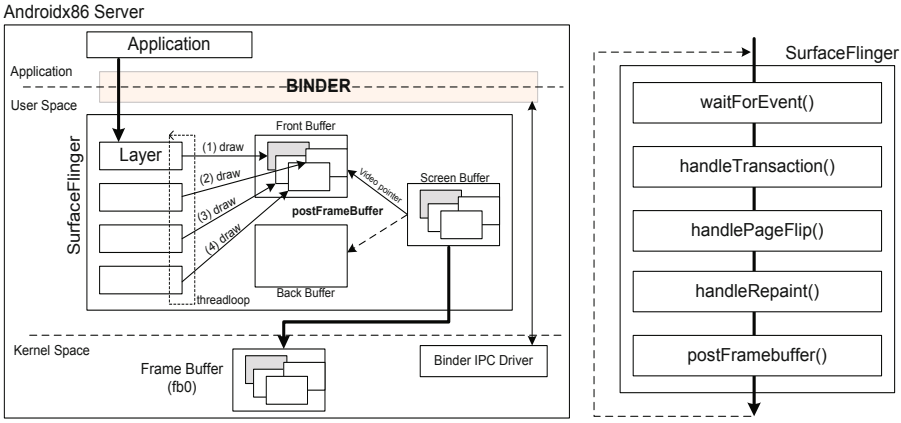


Fig. 1. Mobile Cloud Architecture

3 Surface Flinger in Androidx86

A surface is an object holding pixels that are being composed to the screen. Every window that can be seen on the screen (a dialog, full-screen activity, the status bar) has its own surface that it draws into. In android system, each application probably draws one or more surfaces. Surface flinger renders these surfaces to the final display in their correct Z-order. Suppose that we run an application to draw an object. Fig. 2(a) presents the flow of data from application to frame buffer.

As we mentioned earlier, android-vnc-server [11] (a version of vnc designed for androidx86) reads pixel data from framebuffer *fb0*. Thus, the server must save the raw pixel data from buffer in surface flinger to framebuffer *fb0* as well as enable the graphic card. This unnecessary operation consumes more CPU and increase delay in



(a) From Application to Frame Buffer (b) Surface Flinger Thread Loop

Fig. 2. Surface Flinger

capturing screen raw data to deliver to mobile device. In order to remove this redundancy, we capture data at Surface flinger module and avoid saving pixel data to graphic card and framebuffer *fb0*.

The surface flinger module is mainly defined in file *framework/base/services/surfaceflinger/SurfaceFlinger.cpp* (Androidx86 source). Its operation is based on the thread loop that is shown in Fig. 2(b). At first, the function *waitForEvent* is called to check the rendering thread's message queue is empty or not. The message queue is updated with any change in the screen, it means that the current frame must be different from the previous one. If the rendering thread's message queue is empty, the surface flinger will be in the sleep state for a short time and wake up to check the queue again. If the queue has message, the function *handle-Transaction* is invoked to handle all transactions or layers. It mainly calculates for each layer whether it is necessary to redraw. The function *handlePageFlip* is then called to compute the visible region of layers. With the given width and height of visible region, all the layers that are out of the considered layer will be eliminated. Then, the visible area of the current layer is indicated by subtracting the upper layers that overlap the current layer. When drawing, the layer only computes and draws according to its visible area with the corresponding visible area data. After computing the visible region, the function *handleRepaint* is called to compose all layers into the main surface. Finally, it pushes the raw data into the front/back buffer and saves to framebuffer *fb0*. The raw data is defined as four fields: Red (R), Green (G), Blue (B) and Alpha (A). The size of raw data RGBA is calculated by given height (h) and width (w). Since each pixel is represented by 8 bits, the size of one frame raw data is w*h*8*4 (bit).

4 Proposed Video Hooking Method

In this section, we present two methods to hook video data in surface flinger: the former method is implemented in application-space and the latter one is implemented

inside surface flinger (user-space). The reason we provide two methods is that we can flexible to manage the hooking operation based on the particular purposes.

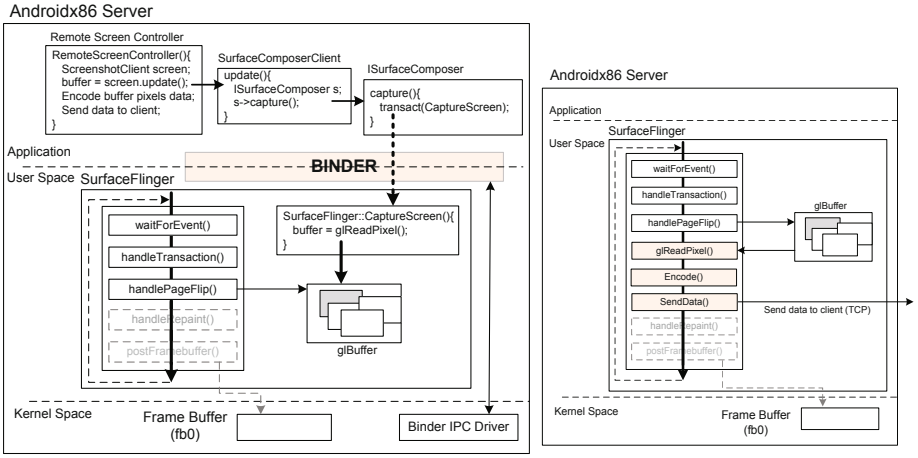
4.1 Off-Screen Video Hooking Method in Application-Space

Based on the thread loop in surface flinger operation, we implement the function *CaptureScreen*() to capture screen buffer after the main frame is composed and saved to buffer *glBuffer* (as Fig. 3(a) shown). The main function that we use to hook screen buffer is *glReadPixel*(). This function is defined in Opengles library, the functionality is to read a block of pixels from buffer and return the pixel data. The type of pixel data is RGBA that we explained in the previous section.

In the application space, we create function *RemoteScreenController*() to invoke function *CaptureScreen*() in surface flinger. After capturing the pixel data, we encode RGBA into image formatted by portable network graphics (PNG) and send to client through TCP/IP. In order to invoke function in user-space from application-space, we must define a surface composer client and call function in surface flinger through Binder IPC driver.

In the thread loop, we eliminate function *handleRepaint*() and *postFramebuffer*() whose functionalities are to deliver pixel data to graphic card (framebuffer *fb0* in kernel-space) and display on screen. By the elimination, the framebuffer *fb0* is freed and the screen will not be displayed as we expect.

In this method, we can easy to enable or disable the hooking operation without any effect on the kernel system.



(a) Off-Screen Video Hooking Method in Application-space

(b) Off-Screen Video Hooking Method in User-space

Fig. 3. Off-Screen Video Hooking Method

4.2 Off-Screen Video Hooking Method in User-space

In this method, we try to hook pixel raw data in the user-space rather than application-space. As shown in Fig. 3(b), we modify thread loop of surface flinger operation by adding three new modules. The first method is to capture pixel data by the function *glReadPixel*() which reads directly from *glBuffer*. This module actually is implemented in the previous method. However, since being in the surface flinger, we can invoke directly as opposed to using surface composer client and Binder IPC. The second module is encoding part which encodes the pixel data RGBA into PNG image. The images then are sent to client by TCP/IP in the third module. We also eliminate two last steps of thread loop (*handleRepaint*() and *postFramebuffer*()) to be not display on the server's screen.

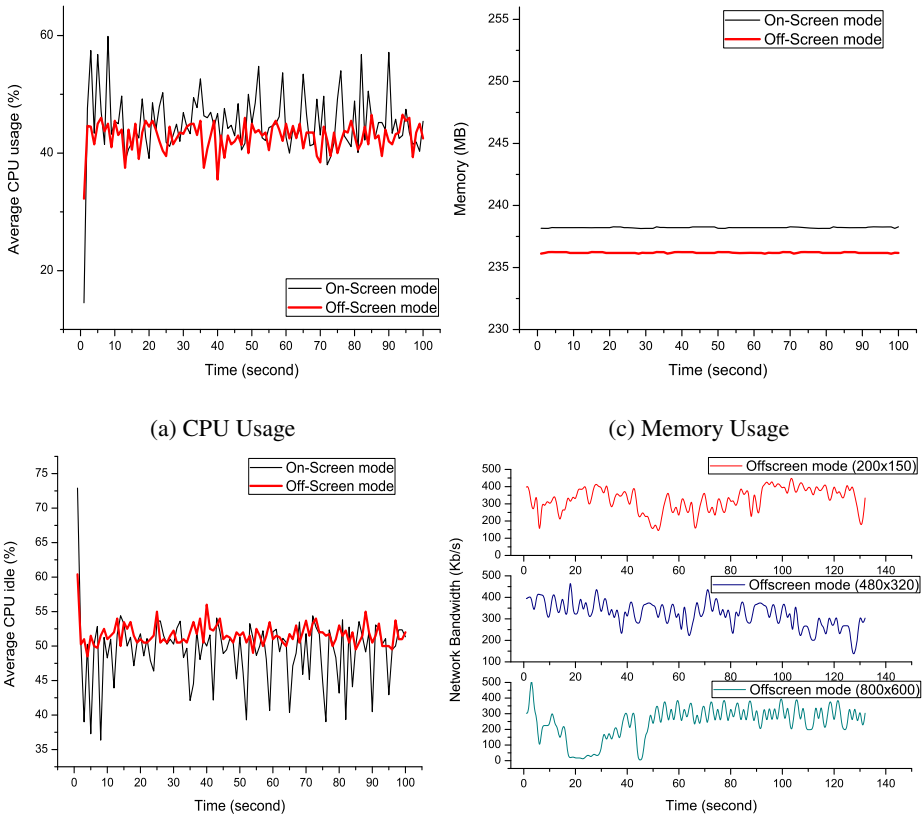
5 Experiments and Result

In the experiment, we compare the performance of our proposed off-screen mode with the on-screen mode protocol. The hardware/software environment according to which we conducted the experiment is described as follows. *Server*: CPU (AMD Athlon(tm) II X2 240 Processor 2.8 GHz), Memory (4GB), Network (100Mbps), OS (Microsoft Windows 7), Hypervisor (VirtualBox ver. 4.1. *VM configuration*: 2 Virtual CPUs, Base memory (1024 MB), Video Memory (12MB), Network (100Mbps), OS (Androidx86 ICS 4.0.4), application (JetBoy application).

Since the cpu usage/memory usage of our two proposed method is similar, we then choose one method to compare with the on-screen mode hooking method. The on-screen mode is the mode in which the server displays drawing on the screen while capturing the screen pixel and sending to client. The results are presented in Fig. 4. We compare CPU performance in Fig 4(a), 4(b). Comparing with on-screen mode, the off-screen mode can save 5% of cpu usage. Thus, the cpu idle is increased in the off-screen mode. Considering memory usage, our proposed method can reduce 2Mb/1024Mb (approximately 0.2%) as shown in fig. 4(c). Although the advantage of our proposed method on one client is quite small, it could achieve an impact on resource consumption as if a large number of clients are considered.

On the other hand, due to the similar PNG encoding, the network consumption of on- and off-screen mode is same. We therefore measure the network consumption of the on-screen mode with three different resolutions: 200x150, 480x320, and 800x600. Since the effect of resolution on the packet size of pixel data, the packet size of each resolution then also differs with each other. The lower resolution composes the frames whose packet size is less than the one composed by higher resolution. However, as Fig. 4(d) shows, the network consumption of three resolution modes are approximately 350Kbps. The reason why they almost consume similar network bandwidth is that the hooking method captures pixel data of lower resolution faster than the higher one. It means that the number of frames per second in lower resolution is greater than in the higher resolution, and they require more bandwidth to transmit to client. The quality of lower resolution is obviously better than the higher one.

Generally, hooking in off-screen mode, we can reduce cpu and memory consumption on the server side in order to increase the number of VMs allocated in one single cloud server.



(a) CPU Usage

(c) Memory Usage

(b) CPU Idle

(d) Bandwidth Usage

Fig. 4. Performance Comparison

6 Conclusion and Future Works

In this paper, we proposed the video hooking method which intercepts at the surface flinger layer of Androidx86. We also propose the off-screen mode to disable the display functions in server. From omitting to display on the server's screen, we can reduce cpu and memory consumption on the server side. However, the quality of video becomes worse when we use higher resolution due to the packet size of pixel data. In the future, we are going to optimize the hooking code to increase the number of frame rate so that the quality of video can be improved.

Acknowledgments. This research was supported by the ETRI, Korea, under the R&D program supervised by the ETRI (12RR1500). This research was also supported by the MSIP (Ministry of Science, ICT&Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (NIPA-2013-(H0301-13-4006)) supervised by the NIPA (National IT Industry Promotion Agency). Professor Eui-Nam Huh is corresponding author.

References

1. Android-x86: Run android on your pc (May 2013), http://www.android-x86.org/
2. Apple: Apple app store (May 2013), https://itunes.apple.com/us/genre/mobile-software-applications/id36?mt=8
3. Deboosere, L., Vankeirsbilck, B., Simoens, P., Turck, F.D., Dhoedt, B., Demeester, P.: Cloud-based desktop services for thin clients. IEEE Internet Computing 99 (2011) (PrePrints)
4. Done'stevez, A.A.F.: Android diskless system. Journal of Free Software and Free Knowledge 1(2) (2012)
5. FB0: The frame buffer device (May 2013), https://www.kernel.org/doc/Documentation/fb/framebuffer.txt
6. Google: Android developers (May 2013), http://developer.android.com/index.html
7. Google: Android market (May 2013), https://play.google.com/store
8. Nguyen, T.D., Choe, S., Huh, E.N.: An efficient mobile thin client technology supporting multi-sessions remote control over VNC. In: 2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE), vol. 3, pp. 155–159 (2012)
9. RDP: Remote desktop protocol (May 2013), http://en.wikipedia.org/wiki/Remote_Desktop_Protocol
10. Richardson, T., Stafford-Fraser, Q., Wood, K.R., Hopper, A.: Virtual network computing. IEEE Internet Computing 2(1), 33–38 (1998)
11. android-vnc server: VNC server for androidx86 (May 2013), http://code.google.com/p/android-vnc-server/
12. Simoens, P., De Turck, F., Dhoedt, B., Demeester, P.: Remote display solutions for mobile cloud computing. Computer 44(8), 46–53 (2011)
13. Song, B., Tang, W., Huh, E.N.: Novel isolation technology and remote display protocol for mobile thin client computing. In: Proceedings of the 6th International Conference on Ubiquitous Information Management and Communication, ICUIMC 2012, pp. 41:1–41:7. ACM, New York (2012)
14. Tang, W., Song, B., Kim, M.S., Dung, N.T., Huh, E.N.: Hybrid remote display protocol for mobile thin client computing. In: 2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE), vol. 2, pp. 435–439 (2012)

Algorithm for Detection of Four Lane Highway Accidents in CCTV Stream

In Jung Lee

Department of Computer Engineering,
Hoseo University, South Korea
leeij@hoseo.edu

Abstract. In this paper, an algorithm is proposed that detects highway Accidents in CCTV Stream. It is known that the flow of vehicle trace is a level spacing distribution on highway such as Wigner distribution. From this distribution we can derive a probability formation induced by distance of traffic flow graphs for each lane. Using this equation we can find abrupt state of vehicle flow comparing with normal flow. We shall show statistical results from some experiments in order to evaluate this algorithm.

Keywords: Accident detection system, detection abrupt signal, Wigner distribution.

1 Introduction

Means for detecting a vehicle accident can be seen in two ways. The one way is to detect by sensor and the other way is analyzing the camera images. In the method of analyzing the camera images, there are two ways. The one is detecting the accident by observed images, when the camera is viewed from individual accident vehicles. The other is finding abnormalities during the flow of traffic when the accident is occurred, at that time, the flow of traffic shall be changed.

In the sensor method, an accelerometer can be used in a car alarm application so that dangerous driving can be detected. It can be used as a crash or rollover detector of the vehicle during and after a crash. With signals from an accelerometer, a severe accident can be recognized. Vibration sensor will detect the signal or if a car rolls over, and Micro electro mechanical system sensor will detects the signal and sends it to controller. Then after conforming the location necessary action will be taken. If the person meets with a small accident or if there is no serious threat to anyone`s life, then the alert message can be terminated by the driver by a switch provided in order to avoid wasting the valuable time of the medical rescue team[1].

It is proposed an intelligent RFID traffic cone that is applied for vehicle accident detection and identification based on image compressing analysis and RFID detection tracking in an accident clamming system and traffic reporting system. This RFID technique deals with multi-vehicles, multi lane and multi road even or traffic junction area. It provides an efficiency time management scheme with enough correctly data reporting, in which a dynamic time schedule is worked out in real time for the driver or passengers of each accident situations. The accessing time operation of the RFID

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_16,

traffic cone system emulates the judgment of a traffic policeman on duty or user who may have the mobile nearby the RFID traffic cone that can be connected via by Wireless channels [2].

In Osaka and Kobe area, acknowledgement has been done by emergency calls, cameras and patrol cars, and takes about 8 minutes to acknowledge the accidents and about 22 minutes before an urgent car comes to at the spot where the accident occurs. Therefore, if you can detect accidents or stopped vehicles immediately, you can quickly alarm to the following cars and urgent cars, thereby preventing secondary accidents, responding to accidents quickly, and reducing jammed time. They have found an effective method for detecting accidents or stopped vehicles. At present, it has already developed a prototype system, 7 cameras at two locations are being used for detection. It is suggested a vision-based traffic accident detection system for automatically detecting, recording, and reporting traffic accidents at intersections [3].

It is proposed that a new framework for real-time automated traffic accidents recognition using histogram of flow gradient (HFG). This framework performs two major steps. First, HFG-based features are extracted from video shots. Second, logistic regression is employed to develop a model for the probability of occurrence of an accident by fitting data to a logistic curve. In case of occurrence of an accident, the trajectory of vehicle by which the accident was occasioned is determined. Preliminary results on real video sequences confirm the effectiveness and the applicability, and it can offer delay guarantees for real-time surveillance and monitoring scenarios [4].

They develop a real-time traffic accident detection system. This system helps us to cope with accidents and discover the causes of traffic accident by detecting the accident. It is gathered video data recorded at several intersections and used them to detect accidents at different intersections which have different traffic flow and intersection design. However, because the data gathered from intersections have incompleteness, uncertainty and complicated causal dependency between them, we construct probability-based networks which calculate based on the probability for correct accident detection. This system instantly sends the detected result to managers using accident alarm system [5].

In this paper, it is proposed an algorithm that tells the alarm to the following vehicle when the accident is detected by CCTV. There are two issues, the one is overcoming the error calculation by shadow, the other is a detecting abrupt signal from the 4 lane vehicle stream when accident has happened. Because, the shadow corrupt detecting accident.

2 Methodology of Overcoming Shadow

This system was tested under typical outdoor field environments at the test site, shown in Figure 1. We have gathered for traffic data, volume count, speed, and occupancy time, with 4 lanes as 2 lanes are upstream and the rests are downstream for three days. And we have evaluated the performance of our system at four different time periods, i.e., daytime, sunrise, sunset, and nighttime, for 30 minutes each.

And the performance of our video-based image detector system is qualified by comparing with laser detector installed on testing place as illustrated in Figure 1.

It would be more appropriate to establish the verification of the proposed system by comparing to a non VIPS measurement system.

The CCTV camera has a 640*480 resolution with 30fps, and the analog image sequences are transformed to digital 256 gray-levels via frame grabber board. The weather conditions during the test are clear two days, and one is rainy day. In the rainy day, we could not gather the baseline data from laser detector because the laser detector cannot operate in rainy day.

Fig. 1. The outdoor test site

In this section, we explain the basic idea behind the tracking algorithm developed in this research. Vehicle tracking has been based on the region-based tracking approach. For individual vehicle tracking the first step, acquisition image sequences and predetermining the detection zones at each lane. The second, we have conducted the background subtraction, deciding threshold for binary images. The background subtraction algorithm requires a relatively small computation time and shows the robust detection in good illumination conditions (26). The third step, morphology for small particles removal as noise, sets in mathematical morphology represent the shapes of objects in an image, for example, the set of all white pixels in a binary image is a complete description of the image. The next step, it is important to remove cast shadows due to extract the vehicle area exactly, we developed the new algorithm in this paper using by edge detection and vertical projections within the vehicle particles. And the fifth step generates the vehicle ID and labeling to each vehicle, and individual vehicle's bounding rectangle data, i.e., left, top, right, bottom coordinates. These particle data are saved into reference table which can be referred to next sequence frames.

In this system, the occlusion detection is easy relatively because of short length of detection zones, less than 15m. And we have considered only limited to explicit occlusion. The explicit occlusion cases have taken place several times during the field test, that is, multiple vehicles enter a scene separately into detection zone, and merge into a moving object region in the scene. In this case, we have maintained each vehicle ID continuously as referred to previous frame.

In the nighttime, diffused reflections on the road due to vehicle headlights pose a serious concern. Thus, we need to pre-process by reflection elimination to adjust the light parameters such as luminosity or brightness, contrast, intensity.

And finally, generated the traffic information and saved these data into table. The performance evaluation of our system is computed by mean absolute percentage error (MAPE) comparing with baseline data, generated by laser detectors. For the vehicle extraction exactly, we have to consider about background estimation, occlusion, cast shadow detection and elimination, and light condition processing at nighttimes. We have considered only limited to explicit occlusion. Our system has covered the four lanes with single camera. As the more system has to be processed, the less performance of system has been. The reason for that if we have included the implicit occlusion process, the performance evaluation marked low grade especially the calculation of velocity. The occlusion detection of this system is easy relatively because of short length of detection zones, less than 15m.

So, many algorithms of cast shadow elimination are proposed, the various cases are occurred in the real traffic flows, for example, dark or light gray shadows, shadow from trees or clouds. The proposed algorithms as mentioned before, have been applied to our experiment, the shadows cannot be extracted exactly as a result. Thus, we have developed the appropriate algorithm in our test site. The basic concept is that the shadow area has less edge because of no variance within shadow. On the other side hand, vehicle area has more edges relatively. Let B be a binary image plane and B_x be a set of number of vertical pixels which value is 1 at x. We define a function $Verti : B / x \rightarrow B_x$

$$by\ Verti\ (x) = \sum_y B_1(x, y)\ ,$$

Where, $B_1(x, y)$ is a pixel of which value is 1 at (x, y) and B / x is a projection of B into x. The graph of $y = Verti(x)$ is shown in Figure 2 after smoothing. In Figure 2, the distribution of edges from moving object area can be discriminated between vehicle and cast shadow. It shows the density of edges via vertical projection. And then discard fewer than 25% that is cast shadow area.

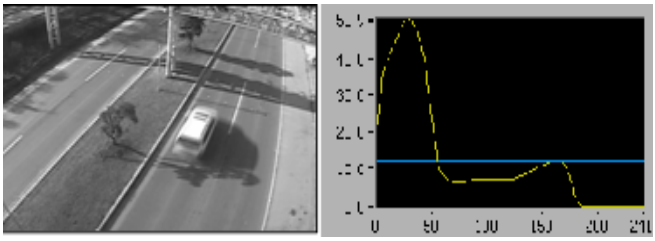


Fig. 2. Cast shadow detection process in our test site

We have tested under typical outdoor field environments at the test site. And have evaluated the performance of traffic information, volume count, speed, and occupancy time. We cannot obtain the baseline data from the radar detection on second day, because of rainy day, for the reason that radar detector have generated the diffused reflections from raindrops under rainy day, therefore the baseline data are incorrect. The traffic information can be obtained by aggregating count vehicles passing through the detection zones for one minute.

The evaluation basis, refer to mean absolute percentage error (MAPE), was calculated by following equation.

$$\text{MAPE}(\%)=\frac{\sum_{t=1}^{T}\frac{|B_t - M_t|}{B_t}}{n}\times 100$$

B_t, M_t is baseline data generated by laser detectors and measured data at time t respectively. T is total measuring time unit. Generally, the time unit T can be divided into four different time periods, i.e., daytime, sunrise time, sunset time, and nighttimes, the traffic information for time unit T is measured for 30 minutes each. The MAPEs of measuring for each time period are shown as Table 1.

Table 1. The mape(%) of measuring traffic information on the test site

Time	Volume Error Rate	Speed Error Rate	Occupancy Time Error Rate
First Day 15:00~15:30(daytime)	3.32%	1.69%	3.18%
3nd Day 07:25~07:55(sunrise)	5.73%	4.01%	7.84%
3nd Day 16:50~17:20(sunset)	4.15%	1.81%	3.01%

3 Four Lane Trajectories

The flow of vehicle can be represented by trajectories, and a time series can be figured from this trace when the flow of vehicle volume calculated for some time interval at each lane as Figure 3.

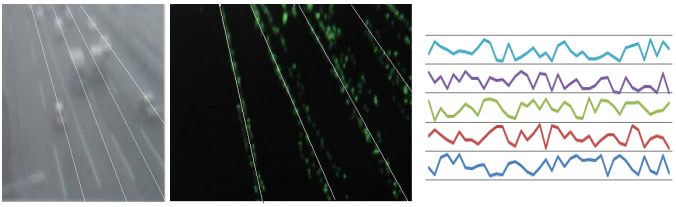


Fig. 3. Trajectories of vehicle and it's time series for each lane

The distribution of position from the bottom line is Wigner distribution (2) in this time series as Figure 4.

$$P(S)=\frac{\pi}{2}S\exp(-\frac{\pi}{4}S^2) \tag{2}$$

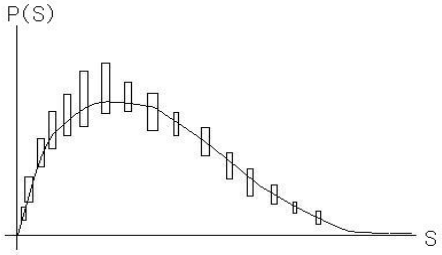


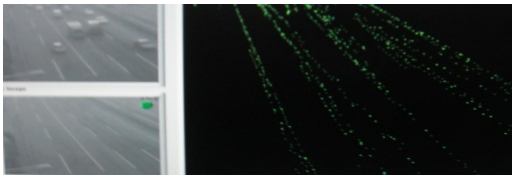
Fig. 4. Wigner distribution of time series

If the Generalized Calogero-Moser system is applied to this equation (2) after finding the eigenvalue of Hamiltonian, we can get the equation (3).

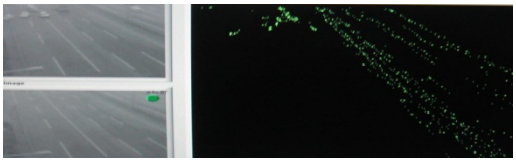
$$p(x_1, x_2, ..., x_n) = C \prod_{1 \le i, j \le n} | x_i - x_j |^v \tag{3}$$

Where, x_i is position value of each line. This equation (3) has maximum value when vehicle flow is same for every time at each lane for example each vehicle flow is same. But if one or two lane has no vehicle flow except other lane has normal flow, the value of equation (3) is abruptly changed. So, the detection system can be made by checking the value of (3) is abrupt value than previous value for some time. In this case we can choose C, v properly for accurate precession and time interval to decision abrupt.

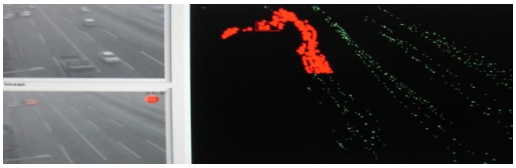
Since, as in Figure 5. , we can find some change of trace image (b) (c) and need to calculate abrupt signal, after that, delivered warning to manager when the calculated



(a) Vehicle trace has no change



(b) Vehicle trace is changed.



(c)Vehicle trace is changed again.

Fig. 5. Showing images from normal flow to abrupt flow and warning state

abrupt signal image like as Figure 6, in this case C, v handle the height of $P(S)$. If C increase, the noise increase and if v increase, than normal state and abrupt state did not distinguished.

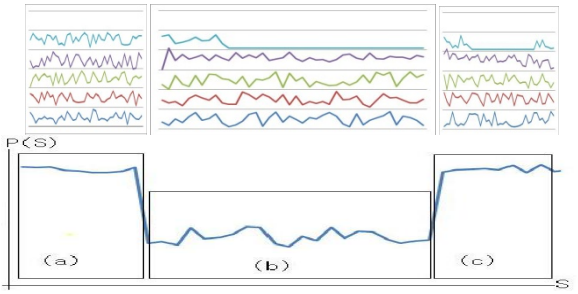


Fig. 6. (a) is an area of normal vehicle flow signal, (b) is an area of one lane flow is stopped and (c) is an detour area of flow

4 Experimental Result

The traffic information can be obtained by aggregating count vehicles passing through the detection zones for one minute. The more detailed measuring results, which are compared with aggregating one minute of baseline data and measuring volume counts for 30 minutes within each time period, are illustrated in Figure 7. as followings.

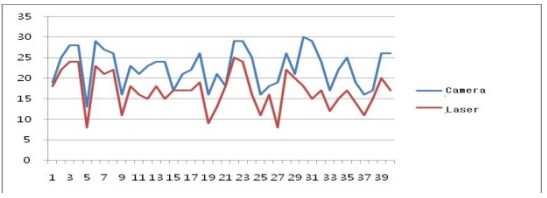


Fig. 7. Volume counts for 30 minutes within each time period

For evaluating accident detection system, we experiments from stored moving image, because accident image is not easy searched, in this case detection time interval is changed per each experiments as shown 10 case in Table 2.

Table 2. Experimental results by changing detection time where "O" is detected case "X" is missed case

Case / Time	1	2	3	4	5	6	7	8	9	0
0.5min	x	x	x	x	o	o	x	o	o	o
1min	o	o	o	o	o	o	o	o	o	o
1.5min	o	o	o	o	o	o	o	o	o	o
2min	o	x	o	o	x	o	x	o	x	x

From this data we take two values as $C = 2.3,\ v = 1.2$ is properly good as in Figure 8.

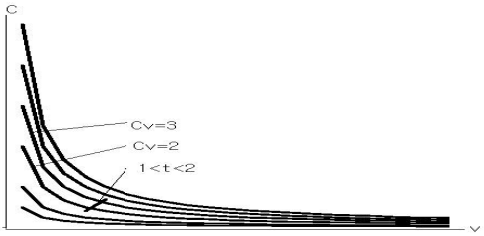


Fig. 8. When $1 \le t \le 2$, $2 \le Cv \le 3$ induced from experimental data formatting toolbar

5 Conclusions

There are two issues, the one is overcoming occlusion by shadow and the other is accident detection system in this paper. The two issues are strictly related because occlusion disturbs accurate accident detection system. In order to eliminate shadow, we have developed the new algorithm using analysis of edge distribution of vehicle and shadows. If the shadow was not erased, than the volume for each lane would not be calculated correctly. Accordingly, the accident detection would not be accurate. In the present work, we have known the flow of vehicle trace has Wigner distribution in the level statistics when we represent each trace avoidable. From this distribution a probability was derived by representing different of position for each lane. Using this equation we could find abrupt state of vehicle flow comparing with normal flow. In this situation, the detection time interval was experimentally good for 1-1.5minutes. In future works, we shall calculate the optimized detection time interval for every accident on high way.

References

1. Goud, V., Padmaja, V.: Vehicle Accident Automatic Detection and Remote Alarm Device. International Journal of Reconfigurable and Embedded Systems (IJRES) 1(2), 49–54 (2012)
2. Kantawong, S.: Tanasak Phanprasit International. Journal of Information Engineering 2(3), 106–115 (2012)
3. ISuge, A., Takigawa, H., Osuga, H., Soma, H., Morisaki, K.: Vehide Navigation & information Systems Conference Proceedings, pp. 45–55. IEEE (1994)
4. Sadeky, S., Al-Hamadiy, A., Michaelisy, B., Sayed, U.: Real-time Automatic Traffic Accident Recognition Using HFG. In: 2010 International Conference on Pattern Recognition, pp. 3348–3352. IEEE (2010)
5. Hwang, J.-W., Lee, Y.-S., Cho, S.-B.: Hierarchical Probabilistic Net-work-based System for Traffic Accident Detection at Intersections. In: 2010 Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing, pp. 211–217. IEEE (2010)

A Petri Net Design toward Prolonging Operational Lifetime of Ad Hoc Networks under Flooding Attack

Fuu-Cheng Jiang[1,*], Hsiang-Wei Wu[2], Chu-Hsing Lin[1],
I-En Liao[2], and Ching-Hsien Hsu[3]

[1] Dept. of Computer Science, Tunghai University, Taichung, Taiwan
[2] Dept. of Computer Science and Engineering,
National Chung-Hsing University, Taichung, Taiwan
[3] Dept. of Computer Science and Information Engineering,
Chung Hua University, Hsinchu, Taiwan
admor@thu.edu.tw

Abstract. The mobile ad hoc networks (MANETs) are vulnerable to flooding attacks launched through compromised nodes or intruders. One crucial type of flooding attacks called RREQ flooding appears to be inevitably proliferated in wireless network. For the RREQ flooding attack, attackers would launch massive RREQ packets with out-of-domain IP address being its destination node. The proposed approach can suppress redundant RREQ packets by the co-operation of destination node and neighbor nodes under one-hop range of attacking node. Much of useless RREQ packets can then be avoided to transmit and receive by legal nodes. In other words, the massive saving of radio energy in MANET can prolong the lifetime of the MANET. To model the proposed approach for qualitative analysis, a Petri Net has been designed to model all relevant system aspects in a concise fashion. On quantitative viewpoint, relevant network simulations have been conducted to validate the proposed scheme approaching the practical scenario. The experimental results reveal that a significant improvement level on lifetime elongation for the MANET can be achieved.

Keywords: Petri net, AODV, Flooding attack, Mobile ad hoc network.

1 Introduction

A mobile ad hoc network (MANET) can be applied in medical emergencies, during natural catastrophes, for military applications and to conduct geographic exploration [1-2]. Mobile and wireless devices belonging to a MANET are usually called mobile nodes. These nodes are characterized by high mobility, low power, limited storage, limited transmission range and finite energy budget without recharging gears. Based on a hop-by-hop routing scheme, AODV (Ad hoc On-Demand Vector) routing protocol offers quick adaption to dynamic link, low processing and memory overhead [3]. When a source node needs a route to a destination, it disseminates a route request (RREQ) message to its neighbors. Each node receiving the message creates a reverse

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_17,

route to the source. This message is flooded until the information needed is complete by either reaching the destination or reaching an intermediate node that has a valid route (in its routing table) to the destination. In conventional wired-networking environment, flooding attacks were once notorious for firing pervasive Denial-of-Service (DoS) attacks or/and Distributed DoS (DDoS) attacks on crucial servers of worldwide enterprise and institutions [4].

On receiving RREQ packets for the first time, any legal node in AODV-based MANET has the obligation to re-disseminate the message. Even in the scenario without malicious attackers, there are considerable negative impacts incurred from rebroadcasting regulation. Some obvious and unavoidable impacts include redundant rebroadcast, contention, and collision [5]. Aimed to such storm attack scenarios, Yi, et al. [6] coined a new phrase in MANETs: RREQ flooding attack. By the RREQ flooding attack, attackers would issue massive RREQ packets with out-of-domain IP address being its destination node. To explore possible solutions for prior flooding attacking issues, we propose a feasible and cost-effective approach called Jointly Bivariate Defensive System (JBDS) to mitigate the impact from flooding attacks.

2 Related Work

To resist RREQ flooding attack in MANETs, Yi et al. [6] proposed an FAP scheme to resist against RREQ flooding attacks. By FAP, one scheme composed of "Neighbor Suppression" had been suggested to resist the RREQ flooding attack. With the scheme of "Neighbor Suppression", they change the processing rule of FIFO (default) for the arriving EERQs from neighbor nodes. Upon the standpoint of intermediate node, the processing priority of an RREQ from a specific node is in reversely proportional to the arriving frequency from that specific node in a unit time interval. Some drawbacks regarding the "Neighbor Suppression" scheme are elaborated as follows.

Firstly, the exact definition of RREQ frequency threshold is not even mentioned in their work. The denial threshold of stop forwarding RREQs for the neighbor nodes cannot be conducted clearly. Secondly, for each intermediate node, their work did not give any feasible approach to learn of the status of neighbor nodes. Thirdly, once the node has been suppressed due to excessive RREQs from some unstable transiently but legal node, it is impossible to recover RREQ relay service back forever for some legal nodes. Finally since each RREQ priority depends on its sender's firing frequency, each node must record every RREQ it receives and reserve the space to hold the priority value after calculating frequency. Inspired by these observations, an effective and feasible approach, called JBDS, has been proposed and described in details hereafter.

3 Proposed Approach to Alleviate RREQ Storm Attack

For the sake of clear presentation, it is assumed that the nodes communicate via a single shared bi-directional wireless medium and all radio transmissions and receptions are omni-directional. All nodes in MANET can operate in promiscuous mode without using any tamper proof hardware. We refer to nodes that are one-hop away (i.e., in the direct range of transmission) as "Direct-Hop Nodes" (*DHNs*) and the covering area under a node's transmission range as the node's *hamlet*. The terms *DHN* and *hamlet* will be used

throughout the following paragraphs. The proposed approach to alleviate RREQ flooding attacks in MANET is termed as RREQ Priority Assessment Index (PAI) scheme, and abbreviated as RREQ_PAI scheme hereafter. The RREQ_PAI scheme configures two pieces of data structures: DS_1 and DS_2, and are shown as follows:

DS_1: Field format of DHN Table (DHNT):

DHN IP Address	AOL (Active or Lost)	RREQ_PAI

DS_2: Field format of RREQ_PAI

Source IP Address	RREQ Numbers	Timestamp	RREP Numbers	RREQ_PAI

Numerically, to assess the malicious tendency for each DHN around a legal node, we use RREQ_API=1, 2and 3 to specify the corresponding levels of normal, greylist and blacklist respectively. Level 3 is the strongest tendency (blacklist) which this DHN is trustless and regarded as an attacker. Any legal node should discard all packets received and refuse further forwarding service from this attacker's packets. Level 2 (greylist) is the moderate tendency implying that this DHN is considered to a potential attacking suspect. Level 1 (normal) implies that the DHN is treated as a legal node and the relay service on it will be conducted normally. The level of priority and PAI setting are summarizes and demonstrated in Table 1. The RREQ_PAI can be upgraded or downgraded dynamically according to arriving profile of packets from each DHN. For resisting RREQ flooding attack, the reception frequency of RREQ is monitored and used to compare it with two threshold parameters: Max_Threshold and Min_Threshold. The level of RREQ_PAI is hence graded by the comparison result. Whenever a grading variation occurs on RREQ_PAI field in DS_2, the RREQ_PAI field in DS_1 is also simultaneously updated to maintain consistency on data structure.

Table 1. Level of Priority and RREQ_PAI Setting

Level of priority (RREQ_PAI)	Malicious tendency	Forwarding service by router nodes
RREQ_PAI = 1	Normal	Router node holds packets and forwards them per threshold policy of RREQ_RATELIMIT
RREQ_PAI = 2	Greylist (Moderate)	This DHN is treated as a gray node. The relay service is conducted with downgrading/upgrading scheme.
RREQ_PAI = 3	Blacklist (Strongest)	Forwarding service is rejected on this DHN and discards its packets silently.

4 A Petri Net Design

4.1 Petri Net Concepts

Petri Nets (PNs) are a graphical mathematical modeling tool application to many systems [8]. They are a promising tool for describing and studying information

processing systems that are characterized as being concurrent, asynchronous, distributed, parallel, nondeterministic, and/or stochastic. A PN is identified as a particular kind of bipartite directed graph populated by three types of objects. They are places, transitions, and directed arcs connecting places and transitions. A PN is a 5-tuple, PN ={P, T, I, O, M_0} [9] where:

$P = \{p_1, p_2, \ldots, p_m\}$ is a finite set of places, where $m > 0$;
$T = \{t_1, t_2, \ldots, t_n\}$ is a finite set of transitions with $P \cup T \neq \emptyset$ and $P \cap T = \emptyset$, where $n > 0$;
I: $P \times T \rightarrow N$ is an input function that defines a set of directed arcs from P to T, where $N=\{0, 1, 2\ldots\}$;
O: $T \times P \rightarrow N$ is an output function that defines a set of directed arcs from T to P.
M_0: $P \rightarrow N$ is the initial marking

A transition t is enabled if each input place p of t contains at least the number of tokens equal to the weight of the directed arc connecting p to t. When an enabled transition fires, it removes the tokens from its input places and deposits them on its output places. PN models are suitable to represent the system that characterizes event-driving, choice, concurrency, conflict, and synchronization. In graphical representation, places are drawn as circles, transitions as bars or boxes. In modeling, using the concept of conditions and events, places represent conditions, and transitions represent events. A transition (an event) has a certain number of input and output places representing the pre-conditions and post-conditions of events, respectively. PNs have been widely accepted for modeling of discrete event systems for qualitative analysis [10].

4.2 PN Modeling of the First Defensive Wall in JBDS Scheme

In JBDS, the defensive task is conducted via two-layer defensive scheme. The first defensive wall (FDW) is executed by DHNs around node originating RREQs using information stored in its DHN Table. The corresponding PN notations for states (places) and transitions (events) are defined in PN_Notation 1 and PN_Notation 2 respectively. In Figure 1, state P2 would be moved to P3 as the transition t3 occurs, which implies the IP address of arriving RREQs is absent in DHN Table. Hence those RREQ packets originated from non-registered suspicious node can be discarded by the FDW.

PN_Notation 1 There are seven transitive states (places) for a router node during modeling proposed JBDS system:
P1: Buffer for arriving RREQ packet.
P2: DHN Table where the IP address of incoming RREQ is present or absent.
P3: Buffer holding RREQ packets under processing.
P4: DHN Table where the PAI state is in one of three basic states.
P5: RREQ_PAI = 1 for the entry node originating RREQ (Normal).
P6: RREQ_PAI = 2 for the entry node originating RREQ (Greylist).
P7: RREQ_PAI = 3 for the entry node originating RREQ (Blacklist).

PN_Notation 2 Given the set of transitive states (places) B={P1, P2, …,P7}, we define the state transitions (events) for a router node during modeling proposed JBDS system:
t1: Transition 't1' is enabled when RREQ packet arrives in place 'P1'.
t2: Transition 't2' is enabled when non-registered RREQ packet has been removed.
t3: The IP address of arriving RREQ is absent in DHN Table.
t4: The IP address of arriving RREQ is present in DHN Table.
t5: RREQ_PAI state is 2 (moderate). t6: RREQ_PAI state is 3 (strongest). t7: RREQ_PAI state is 1 (normal).
t8: Enabled when time duration D2*[b] has been elapsed and no RREQ reception beyond D2.
t9: Forwarding RREQs, and RREP response after D1*[a] time duration.
t10: Forwarding RREQs, and it is found that (1/Max_Threshold)< ΔT*[c] ≦(1/Min_Threshold).
t11: Checking ΔT, and it is found that ΔT≦(1/Max_Threshold).
t12: Forwarding RREQs, no RREP response after D1 time duration.
t13: Checking ΔT, and it is found that ΔT≦(1/Max_Threshold).
t14: Enabled when D2 has been elapsed and no RREQ reception beyond D2.
*a: D1= 2× NET_TRAVERSAL_TIME ; *b: D2= 8× NET_TRAVERSAL_TIME
*c: ΔT = time interval between two arriving successive RREQs.

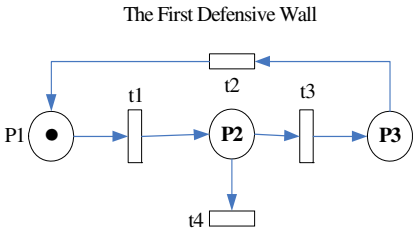


Fig. 1. PN modeling of the FDW in JBDS system

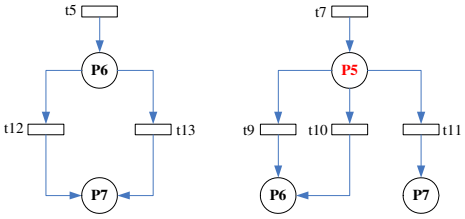


Fig. 2. Preliminary PN model for downgrading operational flow

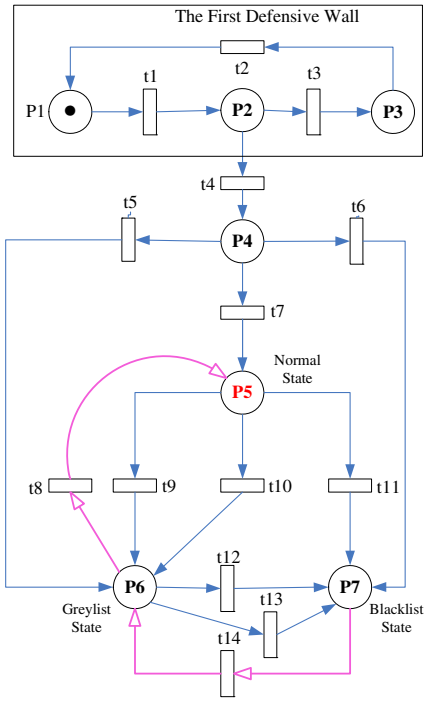


Fig. 3. A Petri Net Design of proposed JBDS system

4.3 Preliminary PN Modeling of RREQ_PAI Component in JBDS System

To shed light on crucial downgrading features in RREQ_PAI component step by step, a preliminary PN modeling of RREQ_PAI Component is designed and illustrated in Figure 2. The corresponding PN notations for states (places) and transitions (events) are defined in PN_Notation 1 and PN_Notation 2 respectively. Two downgrading scenarios would occur when RREQ_PAI =2 (place 6) is downgraded to RREQ_PAI=3 (place 7) and RREQ_PAI=1 (place 5) is downgraded to RREQ_PAI=2 or RREQ_PAI=3. If the reception frequency of RREQs exceeds (1/Max_Threshold), it implies that a storm flooding scene emerges and a urgent action should be taken to reject forwarding immediately. Hence the state must be downgraded to state RREQ_PAI=3 (place 7) whenever state was in either state P5 or state P6. This is the design origin for both transitions (events) t11 and t13 in Figure 2. The right diagram in Figure 2 illustrates the downgrading operational flow starting from the normal level (P5) to other two levels (P6, P7). The transition t10 (event) is triggered when checking time interval (ΔT) between these two successive RREQs, and it is found that $(1/\text{Max\_Threshold}) \leqq \Delta T < (1/\text{Min\_Threshold})$.

According to the disciplines in RFC 3561 standard (Section 6.6), there are two types of RREP generation. A node generates a RREP if either (1) it is itself the destination (i. e., RREP generation by the Destination), or (2) it has an active and valid route to the destination (i.e., RREP generation by an intermediate node). Hence, if this RREQ is not an attack one, i.e., it is destined to a legally existing destination node, an RREP should be responded not beyond proper latency, which is called "NET_TRAVERSAL_TIME" in terminology of RFC3561 standard. Practically, the allowable and proper responsive latency is configured to be D1 which is defined to be twice of NET_TRAVERSAL_TIME for our proposed downgrading policy. If the node cannot receive the responsive RREP in D1 from an existing destination node or an intermediate node with an active route, it can be reasonably concluded that this RREQ is a fraudulent one. Hence, the level of malicious tendency should be downgraded to the next worse state (P6 or P7) as well. This is the design rationale for both transitions (events) t9 and t12 in Figure 2.

4.4 Composed PN Model of the Proposed JBDS System

The composed PN model for the proposed JBDS System is illustrated in Figure 3, which consists of 7 places and 14 transitions and corresponding notations of the PN model are described in PN_Notation 1 and PN_Notation 2 respectively. To prevent transient, unstable accidents which two successive RREQs might be issued in too short interval by legal nodes, an upgrading policy is designed and integrated into the proposed JBDS system. Desilva et al. [11] once concluded that over a short period of time, RREQs from legal and malicious nodes are not easy to distinguish. However over a long duration, attackers can be identified since legal nodes send a high rate of RREQs for short duration, but attackers do so at all times. Based on their observation, an easy upgrading policy is proposed to curb bogus RREQs from attackers without hurting legal nodes. Because if once any unstable legal node has been downgraded to RREQ_PAI = 2 (place 6) or 3 (place P7), it has no chance to deserve proper grading assessment forever apparently. Hence, we design an upgrading policy into the proposed JBDS system. For

upgrading policy, the RREQ_PAI can be upgraded to its next better level if the DHN has not received any RREQ packets beyond time duration D2. The corresponding operational flow on upgrading part is expressed in terms of (pink) arcs with blank arrowheads in Figure 3. The moderate level (place P6) can be upgraded to the normal level (place P5) via the condition t8 and also the strong attacking level (place P7) can be upgraded to the next better level (place P6) if the condition t14 occurs.

5 Network Simulation and Performance Evaluation

To observe the RREQ attacking profile for each attacker, we trigger the attacking pulses against the network by attackers one by one on timing sequence. In each simulation, the normal path discovery process starts at the 50^{th} time units in NS2 simulation platform due to need on system stabilization. Thus, five RREQ attackers trigger their own flooding attack according to individual time instants expressed in the vector: [Attacker 1, Attacker 2, Attacker 3, Attacker 4, Attacker 5] = [100, 140, 180, 220, 260]. Two contours are obtained by recording and calculating the total number of RREQs existing in the network as shown in Figure 4. Each data point in Figure 4 is the average of 10 runs for each condition with the same topology. In Figure 4, the X axis and Y axis are configured to be the time units in the NS2 simulation platform and total number of RREQ packets detected throughout the simulation respectively.

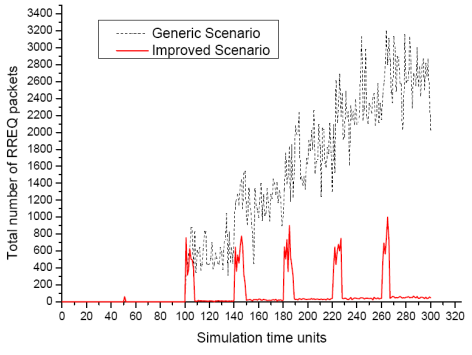


Fig. 4. Pattern on quantities of RREQ packets

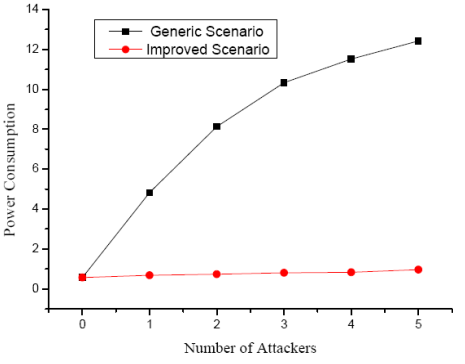


Fig. 5. Lifetime elongation via alleviation of power consumption

From the viewpoint of power-saving approach, one of major design issues is to manage power consumption efficiently and to increase the operational lifetime of the network as possible. In Figure 5, the curve of “Generic Scenario" (black box markers) represents the average power consumption pattern without the proposed defensive scheme. As the number of attackers is increased, the flooding impact on the power consumption goes higher and higher as shown along the contour: "Generic Scenario" of Figure 5. The contour of “Improved Scenario" (red ball markers) shows the average power consumption using the proposed defensive scheme. Obviously, the power consumption has been improved considerably. To perceive the improvement degree quantitatively, the average improvement ratio would reach about 90.09%,

which implies the average power consumption can be alleviated considerably by the amount of 90% of generic scenario. Such an improvement ratio exemplified in Figure 5 manifests the excellent power-saving toward prolonging operational lifetime of AODV network under flooding attack.

6 Conclusion

The proposed approach can suppress redundant RREQ packets by the co-operation of destination node and neighbor nodes under one-hop range of attacking node. On qualitative analysis, a Petri Net design is provided for in-depth understanding system profile. On quantitative viewpoint, relevant network simulations have been conducted to validate the proposed scheme approaching the practical scenario. Based on two kernel parameters, total number of RREQ packets and average power consumption, the experiment results using NS2 simulator are obtained. With almost 90% improvement ratio on average power consumption in the exemplified scenario, the proposed power-saving approach indeed provides a feasibly cost-efficient technique to enhance the longevity of the Ad Hoc network under flooding attack.

References

1. Corson, S., Macker, J.: Mobile ad hoc networking (MANET): routing protocol performance issues and evaluation considerations. IETF RFC 2501 (January 1999), ftp://ftp.rfc-editor.org/in-notes/rfc2501.txt
2. Perkins, C.E., Royer, E.M.: Ad hoc on-demand distance vector routing. In: The 2nd IEEE Workshop on Mobile Computing Systems and Applications (WMCSA 1999), New Orleans, LA, February 25-26, pp. 90–100 (1999)
3. Perkins, C.E., Royer, E.M., Das, S.: Ad hoc on-demand distance vector (AODV) routing. IETF Experimental RFC 3561 (July 2003), ftp://ftp.rfc-editor.org/in-notes/rfc3561.txt
4. Gu, Q., Liu, P., Chu, C.-H.: Analysis of area-congestion-based DDoS attacks in ad hoc networks. Ad Hoc Networks 5, 613–625 (2007)
5. Tseng, Y.-C., Ni, S.-Y., Chen, Y.-S., Sheu, J.-P.: The Broadcast Storm Problem in a Mobile Ad Hoc Network. Wireless Networks 8, 153–167 (2002)
6. Yi, P., Dai, Z., Zhong, Y., Zhang, S.: Resisting Flooding Attacks in Ad Hoc Networks. In: IEEE ITCC 2005, vol. 2, pp. 657–662 (2005)
7. Li, S., Liu, Q., Chen, H., Tan, M.: A New Method to Resist Flooding Attacks in Ad Hoc Networks. In: IEEE WiCOM 2006, pp. 1–4 (September 2006)
8. Murata, T.: Petri Nets: Properties, Analysis and Applications. Proceedings of The IEEE 77(4) (April 1989)
9. Lee, J.-S., Hsu, P.-L.: Implementation of a Remote Hierarchical Supervision System Using Petri Nets and Agent Technology. IEEE Transactions of Systems, MAN, and Cybernetics – Part C: Applications and Reviews 37(1) (January 2007)
10. Davidrajuh, R., Lin, B.: Exploring airport traffic capability using Petri net based model. Expert Systems with Applications 38, 10923–10931 (2011)
11. Desilva, S., Boppana, R.V.: Mitigating Malicious Control Packet Floods in Ad Hoc Networks. IEEE Communications Society / WCNC (2005)
12. The Network Simulator Ns2 (2010), http://www.isi.edu/nsnam/ns/

Clustering Method Using Weighted Preference Based on RFM Score for Personalized Recommendation System in u-Commerce

Young Sung Cho[1], Song Chul Moon[2], Seon-phil Jeong[3], In-Bae Oh[4], and Keun Ho Ryu[1]

[1] Department of Computer Science, Chungbuk National University, Cheongju, Korea
[2] Department of Computer Science, Namseoul University, Cheonan-City, Korea
[3] Computer Science and Technology, DST, BNU-HKBU United International College
[4] Chungbuk Health & Science University, Chungbuk, Korea
youngscho@empal.com, moon@nsu.ac.kr, spjeong@uic.edu.hk, iboh@chsu.ac.kr, khryu@dblab.chungbuk.ac.kr

Abstract. This paper proposes a new clustering method using the weighted preference based on RFM(Recency, Frequency, Monetary) Score for personalized recommendation in u-commerce under ubiquitous computing environment which is required by real time accessibility and agility. In this paper, using an implicit method without onerous question and answer to the users, not used user's profile for rating, it is necessary for us to extract the most frequent purchase items from the whole purchase data and to calculate the weighted preference of item for customer in order to reduce customers' search effort, to reflect frequently changing trends by emphasizing the important items and to improve the rate of recommendation with high purchasability. To verify improved better performance of proposing system than the previous systems, we carry out the experiments in the same dataset collected in a cosmetic internet shopping mall.

Keywords: RFM Analysis, Collaborative Filtering, k-means Clustering.

1 Introduction

Along with the advent of ubiquitous computing and the spread of intelligent portable device such as smart phone, PDA and smart pad has been amplified, a variety of services and the amount of information has also increased. It is becoming a part of our common life style that the demands for enjoying the wireless internet are increasing anytime or anyplace without any restriction of time and place[2],[4]. The recommendation system helps customers to find items easily and helps the e-commerce companies to set easily their target customer by automated recommending process. Therefore, customers and companies can take some benefit from recommendation system. The possession of intelligent recommendation system is becoming the company's business strategy. A recommendation system using data mining technique based on RFM to meet the needs of customers has been actually

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*, Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_18,

processed the research[1-5]. It is crucial to have different weights for different transactions in order to reflect their different importance and to adjust the weighted preference based on RFM score for recommendation by emphasizing the important transactions. We can improve the performance of recommendation through a new clustering method using the weighted preference based on RFM score. The next section briefly reviews the literature related to studies. The section 3 is described a new method for personalized recommendation system in detail, such as system architecture with sub modules, the algorithm for proposing system, and the procedure of processing the recommendation. The section 4 describes the evaluation of this system in order to prove the criteria of logicality and efficiency through the implementation and the experiment. In section 5, finally it is described the conclusion of paper and further research direction.

2 Related Works

2.1 RFM Analysis

RFM method is generally known in database marketing and direct marketing. It is easy for us to recommend the item with high purchasability using the customer's score and the item's score. The RFM score can be a basis factor how to determine purchasing behavior on the internet shopping mall, is helpful to buy the item which they really want by the personalized recommendation. One well-known commercial approach uses five bins per attributes, which yields 125 cells of segment. The following expression presents RFM score to be able to create an RFM analysis. The RFM score will be shown how to determine the customer as follows, will be used in this paper. The variables (A, B, C) are weights. The categories (R, F, M) have five bins.

$$\text{RFM score} = A \times R + B \times F + C \times M \quad (1)$$

The RFM score is correlated to the interest of e-commerce[2]. It is necessary for us to keep the analysis of RFM to be able to reflect the attributes of the item in order to find the items with high purchasability. In this paper, we can make the task of preprocessing for clustering purchase data to join customer's data using the weighted preference based on RFM score to recommend the item they really want exactly.

2.2 Collaborative Filtering

Collaborative filtering comes from the method based on other users' preferences. It is one of the filtering methods, which is associated with the interests of a user by collecting preferences or taste information from many users. There are two kinds of method. One is the explicit method, used user's profile for rating. The other is the implicit method, which is not used user's profile for rating data. The implicit method without onerous question and answer to the users, is not used user's profile for rating but is used user's web log patterns or purchased history data to show user's buying patterns so as to reflect the user's preference. There are some kinds of the method for recommendation, such as collaborative filtering, demographic filtering, rule-base

filtering, contents based filtering, the hybrid filtering which put such a technique together and association rule and so on in data mining technique currently. The explicit method can not only reflect exact attributes of item, but also still has the problem of sparsity and scalability, though it has been practically used to improve these defects[5].

2.3 k-Means Clustering

Clustering is the process of organizing objects in a database into clusters. It involves classifying or segmenting the data into groups based on the natural structure of the data. Clustering techniques [6,7] fall into a group of undirected data mining tools. The principle of clustering is maximizing the similarity inside an object group and minimizing the similarity between the object groups. Clustering algorithm is a kind of customer's segmentation methods commonly used in data mining, can often use to k-means clustering algorithm. K-means is the most well-known and commonly, used partition methods are the simplest clustering algorithm. In the k-means algorithm, cluster similarity is measured in regard to the mean value of the objects in a cluster, which can be viewed as the cluster's center of gravity. This algorithm uses as input a predefined number of clusters that is the k from its name. Mean stands for an average, an average location of all the members of a particular cluster. The euclidean norm is often chosen as a natural distance which customer a between k measure in the k-means algorithm[8]. The a_i means the preference of attribute i for customer a.

$$d_{a,k} = \sqrt{\sum_i (a_i - k_i)^2} \tag{2}$$

There are two part of k-means algorithm. The 1st part is that partition the objects into k clusters. The 2nd part is that iteratively reallocate objects to improve the clustering. The system can use euclidean distance metric for similarity. In this paper, we can do clustering the customers' data using K-means algorithm[9] to segment customers and finally forms groups of customers with different features. Through analyzing different groups of customers' data, we try to do the recommendation for the target customers of internet shopping mall efficiently.

3 Our Proposal for a Personalized u-Commerce Recommendation System

3.1 System Architecture

We can depict the system configuration concerning the recommendation system with clustering method using the weighted preference based on RFM score under ubiquitous computing environment which is required by real time accessibility and agility. This system had four agent modules which have the analytical agent, the recommendation agent, the learning agent, the data mining agent in the internet shopping mall environment. We observed the web standard in the web development,

so developed the interface of internet to use full browsing in mobile device. As a matter of course, we can use web browser in wired internet to use our recommendation system. We can use the system under WAP in mobile web environment by using feature phone as well as using the internet browser such as safari browser of iPhone and Google chrome browser based on android so as to use our system by using smart phone.

3.2 Clustering Method Using the Weighted Preference Based on RFM Score

In this section, we can describe a new clustering method using the weighted preference based on RFM score. We can extract the most frequent purchase data from the whole purchase data to join the customer information for pre-processing so as to be possible to recommend the items with efficiency. Then, the system can create the cluster with neighborhood user-group using the task of preprocessing for clustering purchase data to join customer's information using the weighted preference based on RFM score, as having input vectors demographic variables: the code of classification such as age, gender, occupation, region, and RFM variable. The system can take the preprocessing task which is able to extract the most frequent purchase data from the whole purchase data, to apply the rate of weight based on the quantity item by each rank of the RFM score and then create the cluster of purchase data sorted by item category, joined the cluster of user information called by customer DB, neighborhood user group[5]. As you can see above, the system can use the extracted purchase data(sale_dat3) with a lot of purchasing counts, between the score is more than 19 points and the score is less than 40 points, calculate the rate of weight based on the quantity item by each rank of the RFM score. After that, the system can apply weight to the preference of item, then adjust the results of preference by emphasizing the important item by each rank of the RFM score. The procedural algorithm of preprocessing for k-means clustering using the weighted preference based on RFM score (KCWP) is depicted as the following Table 1.

Table 1. The procedural algorithm of preprocessing for KCWP

Step 1 : The RFM score of customer is computed so as to reflect the attributes of the customer, consists of three attributes(R, F, M), each attribute has five bins divided by each 20%, exact quintile.
Step 2 : The system can aggregate the quantity of purchased data by each interval customer's RFM scores, which is aggregated counts of distribution from the whole data, make the rate of weight.
Step 3 : The system can calculate the rate of weight based on quantity item with each rank of RFM score
Step 4 : The system can scan whole database(sale) and calculate the weighted preference, weighted by each rank of RFM score.
Step 5 : The system can recommend the item with high purchasability according to the preference in the cluster selected by the code of classification reflected demographic variable and customer's RFM score.

The following algorithm of KCWP is encoded by pseudo code using k-means Clustering.

Table 2. Clustering method using the weighted preference based on RFM Score

Input : Item Category Code Table(CCT), Customer-Item Category-Preference(UCP) Matrix, weighted_tbl
Output : Feature Vector, Purchase data neighborhood Group

begin
1. Classify the purchase DB joined customer DB with the Feature Vector as the basis of demographic *variable and customer's RFM scores;*
// the Feature Vector which has customer's score, age, gender, occupation
2. For(each brand item in whole CCT)
Compute the preference of brand item in item category
Endfor;
2.1 Compute the average of item preference(Pref_UC(u,c)) by Group function as an aggregative function, it is normalized at first;
2.2 For(each Item Category
Compute the average of preference of item category based on CCT
Endfor;

2.3 Adjust the result of weighted item preference, finished, sorted by CCT, using purchase data classified by the Feature Vector before clustering;
// V is the set of all the item preferences that M in CCT

$$V = (V_1,\ V_2, V_3 \ldots,\ V_m)$$

$$V_i = \sum_k (Pref\_UC(u_i, c_k * weight_k / M)) /$$

$$\sum_i (\sum_k (Pref\_UC(u_i, c_k * weight_k / M)))$$

3. Compute the average of the weighted item preference(Pref_UC(u,c)) by Group function as an aggregative function, it is normalized at the second.

4. Create the cluster of neighborhood by using K-means algorithm ;
// The neighborhood cluster is created by using k-means algorithm
End;

3.3 The Procedural Algorithm for Recommendation

The system can search the information in the cluster selected by using the code of classification and RFM score. It can scan the preference of brand item in the cluster, suggest the brand item in item category selected by the highest preference as the

average of brand item. This system can create the list of recommendation with TOP-N of the highest preference of item to recommend the item with purchasability efficiently. This system can recommend the items with efficiency, are used to generate recommending item according to the basic the weighted preference through an clustering method using k-means algorithm. It can recommend the associated item to TOP-N of recommending list if users want to have the cross-selling or up-selling. This system takes the cross comparison with purchase data in order to avoid the duplicated recommendation which it has ever taken.

4 The Environment of Implementation and Experiment and Evaluation

4.1 Experimental Data for Evaluation

This system proposes a new clustering method(KCWP) using the weighted preference based on RFM score under ubiquitous computing environment. In order to do that, we make the implementation for prototyping of the internet shopping mall which handles the cosmetics professionally and do the experiment. It is the environment of implementation and experiment below.

- OS : Windows XP SP2,
- Web Server: Apache 2.2.14 / WAP 2.0
- Server-Side Script : JSP/PHP 5.2.12
- Client-Side Script : XML/WML2.0/ HTML5.0/CSS3/JAVASCRIPT
- Database : MySQL 5.1.39
- J2SDK(1.7.0_11)
- MySQL JDBC
- jQuery Mobile
- jakarta-tomcat (5.0.28)

We have carried out the implementation and the experiment for proposing system through system design, we have finished the system implementation about prototyping recommendation system. It could be improved and evaluated to new system through the result of experiment with the metrics such as precision, recall, F-measure as comparing proposing system with other previous system(KCIP) using clustering algorithm of item preference based on RFM and existing system.

4.2 Experimental Data for Evaluation

We used 319 users who have had the experience to buy items in internet shopping mall, 580 cosmetic items used in current industry, 1600 results of purchase data recommended in order to evaluate the proposing system[4]. In order to do that, we make the implementation for prototyping of the internet shopping mall which handles the cosmetics professionally and do the experiment. We have finished the system implementation about prototyping recommendation system. We'd try to carry out the experiments in the same condition with dataset collected in a cosmetic internet

shopping mall. It could be evaluated in MAE and Precision, Recall, F-measure for the recommendation system in clusters. It could be proved by the experiment through the experiment with learning data set for 12 months, testing data set for 3 months in a cosmetic cyber shopping mall[4]. The 1st system of clustering method using weighted preference based on RFM score, is proposing system called by "proposal", the 2nd system is other previous system(KCIP) using clustering algorithm of item preference based on RFM[5] called by "previousl", the third system is existing system based on the whole data called by "existingl".

4.3 Experiment and Evaluation

We can make the task of clustering of item category based on purchase data for preprocessing under ubiquitous computing environment. The proposing system's overall performance evaluation was performed by dividing the two directions. The first evaluation is mean absolute error(MAE). The mean absolute error between the predicted ratings and the actual ratings of users within the test set. The mean absolute error is computed the following expression-3 over all data sets generated on purchased data.

$$MAE = \frac{\sum_{i=1}^{N} |\epsilon_i|}{N} \tag{3}$$

N represents the total number of predictions, ε represents the error of the forecast and actual phase i represents each prediction.

Table 3. The result for table of MAE by comparing proposal system with existing system

	P_count	Proposal	KCIP	Existing
MAE	50	0.24	0.47	0.65
	100	0.13	0.23	0.32
	300	0.05	0.07	0.08
	500	0.03	0.05	0.06

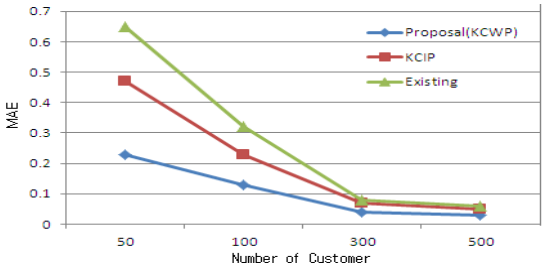


Fig. 1. The result for the graph of MAE by comparing proposal system with existing system

The next evaluation is precision, recall and F-measure for proposing system in clusters. The performance was performed to prove the validity of recommendation and the system's overall performance evaluation. The metrics of evaluation for recommendation system in our system was used in the field of information retrieval commonly[10].

Table 4. The result for table of precision, recall, F-measure for recommendation ratio by each cluster

Cluster	Proposal(KCWP)			KCIP			Existing		
	Precision1	Recall1	F-measure1	Precision2	Recall2	F-measure2	Precision3	Recall3	F-measure3
C1	53.10	61.65	57.06	56.98	91.44	65.90	56.98	50.89	50.21
C2	62.50	47.50	53.98	100	27.27	42.86	38.97	15.18	20.88
C3	54.49	63.99	58.86	48.79	55.70	48.41	48.79	31.32	35.64
C4	55.83	60.00	57.84	49.36	52.53	48.09	49.36	29.54	35.06
C5	41.57	80.95	54.93	55.50	23.93	32.19	44.26	21.81	27.65
C6	45.95	61.25	52.51	52.49	38.37	41.95	52.49	34.98	39.75
C7	45.99	74.65	56.92	50.41	47.40	45.24	50.41	43.21	43.10
C8	55.55	61.31	58.29	50.93	37.23	40.03	50.93	36.60	39.64
C9	100	95.24	97.56	47.41	27.27	32.60	47.41	26.81	32.26
C10	50.00	50.00	50.00	43.60	37.23	38.17	43.60	36.60	37.82
C11	41.67	71.43	52.63	67.18	20.69	31.17	46.53	18.32	25.10
C12	16.67	100	28.58	67.23	62.50	60.94	67.23	55.34	57.10

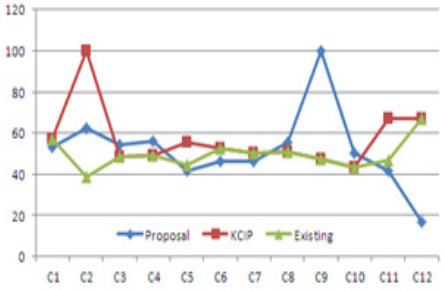


Fig. 2. The result of recommending ratio by precision

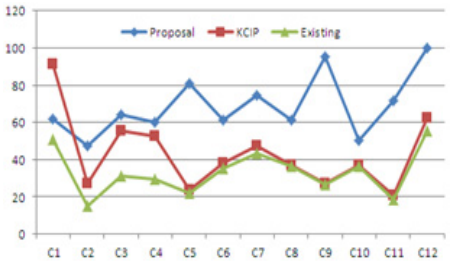


Fig. 3. The result of recommending ratio by recall

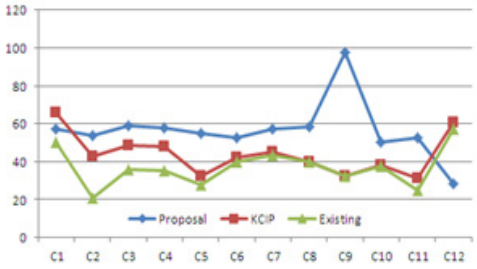


Fig. 4. The result of recommending ratio by F-measure

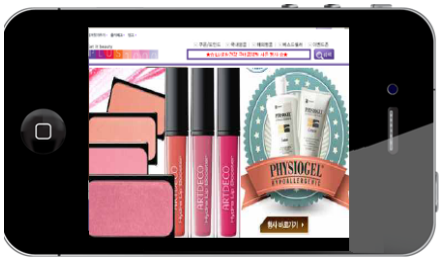


Fig. 5. The site of recommendation of cosmetics

Above Table 4 presents the result of evaluation metrics (precision, recall and F-measure) for recommendation system. The new clustering method(KCWP) is improved better performance of proposing system than the previous systems. Our proposing system with the method using item preference is higher 28.73% in recall, higher 16.02% in F-measure even if it is lower 1.28% in precision than the system(KICP). As a result, we could have the recommendation system to be able to recommend the items with high purchasability. The figure 5 is shown in the site of recommendation of cosmetics on a smart phone. The new clustering method is better performance than the previous method although it is lower in precision.

5 Conclusion

Recently u-commerce as an application field under ubiquitous computing environment required by real time accessibility and agility, is in the limelight[4]. We proposed a new clustering method using the weighted preference based on RFM score for recommendation system in u-commerce in order to improve the accuracy of recommendation with high purchasability. The previous system(KICP) did not reflect the importance of a item, and do not consider these dynamic changes in different items in the retail market basket data analysis. It is crucial to have different weights for different items and adjust the results of preference by emphasizing the important quantity item by each rank of the RFM score.We have described that the performance of the proposing system with new clustering method is improved better than the system (KICP) and existing system. To verify improved better performance of proposing system, we carried out the experiments in the same dataset collected in a cosmetic internet shopping mall. It is meaningful to present a new clustering using the weighted preference of item, reflected quantity item by each rank of RFM score for recommendation system in emerging data under ubiquitous computing environment. The following research will be looking for ways of a personalized recommendation using the task of clustering method based on ART neural network to integrate various cluster for the efficiency and scalability.

Acknowledgements. This work[1)] was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST) (No. 2012-0000478) and this paper[2)] was supported by funding of Namseoul University.

References

1. Cho, Y.S., Jeong, S.P., Ryu, K.H.: Implementation of Personalized u-commerce Recommendation System using Preference of Item Category based on RFM. In: The 6th International Conference on Ubiquitous Information Technologies & Applications, pp. 109–114 (December 2011)
2. Cho, Y.S., Moon, S.C., Noh, S.C., Ryu, K.H.: Implementation of Personalized recommendation System using k-means Clustering of Item Category based on RFM. In: 2012 IEEE International Conference on Management of Innovation & Technology Publication (June 2012)
3. Cho, Y.S., Moon, S.C., Ryu, K.H.: Personalized Recommendation System using FP-tree Mining based on RFM. In: KSCI, vol. 17(2) (February 2012)
4. Cho, Y.S., Moon, S.C., Ryu, K.H.: Mining Association Rules using RFM Scoring Method for Personalized u-Commerce Recommendation System in emerging data. In: Kim, T.-h., Ramos, C., Abawajy, J., Kang, B.-H., Ślęzak, D., Adeli, H. (eds.) MAS/ASNT 2012. CCIS, vol. 341, pp. 190–198. Springer, Heidelberg (2012)
5. Cho, Y.S., Moon, S.C., Jeong, S.P., Oh, I.B., Ryu, K.H.: Clustering Method using Item Preference based on RFM for Recommendation System in u-Commerce. In: Han, Y.-H., Park, D.-S., Jia, W., Yeo, S.-S. (eds.) Ubiquitous Information Technologies and Applications. LNEE, vol. 214, pp. 353–362. Springer, Heidelberg (2012)
6. Hand, D., Mannila, H., Smyth, P.: Principles of Data Mining. The MIT Press (2001)
7. Collier, K., Carey, B., Grusy, E., Marjaniemi, C., Sautter, D.: A Perspective on Data Mining. Northern Arizona University (1998)
8. Shin, M.-S.: An Alert Data Mining Framework for Intrusion Detection System. Journal of Korea Academia-Industrial cooperation Society 12(1) (2011)
9. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning – Data Mining, Inference, and Prediction. In: Springer (2001)
10. Herlocker, J.L., Kosran, J.A., Borchers, A., Riedl, J.: An Algorithm Framework for Performing Collaborative Filtering. In: Proceedings of the 1999 Conference on Research and Development in Information Research and Development in Information Retrival (1999)

Motor Primitive Generation Framework for NAOs in Ubiquitous Applications

Yunsick Sung[1], Kyungeun Cho[2], Young-Sik Jeong[2], and Kyhyun Um[2,*]

[1] Department of Game Mobile Contents, Keimyung University, Daegu, Republic of Korea
[2] Department of Multimedia Engineering, Dongguk University, Seoul, Republic of Korea
khum@dongguk.edu

Abstract. Robots currently being developed for ubiquitous applications contain diverse kinds of system components to smoothly control themselves and the mutual synchronization among one another to achieve their predefined goals. Given that motor primitives determine a robot's capabilities, research on how to define and generate such primitives is crucial to understanding the core techniques of robotic control. One method that generates motor primitives of greater variety than other methods is based on learning by demonstration. However, the generated motor primitives are similar when compared by a Euclidean-distance algorithm. This necessitates a mechanism that can accurately compare two motor primitives. In this paper, a NAO framework that generates motor primitives and compares them accurately is proposed. In an experiment in which the framework was applied to a NAO, 1487 candidate motor primitives were generated and 40 were utilized.

Keywords: Human-Robot Interaction, Q-learning, Motor Primitive, NAO.

1 Introduction

In ubiquitous computing environments, diverse kinds of services can be provided by using humanoid robots. Such robots typically perform their goals by executing predefined motor primitives. Therefore, the methods that define the motor primitives affect the resultant services of the humanoid robots.

Various approaches to generate motor primitives are available. For example, one method uses the Maximin-selection algorithm [1]. The advantage of this method is that the types of motor primitives it generates are more diverse than those of other methods. However, each motor primitive is configured based on different movements. Hence, the difference between motor primitives needs to be compared accurately to increase the quality of the generated motor primitives.

In this paper, a framework is proposed to collect data from, and generate motor primitives for, NAO robots. The generated motor primitives are then updated to, and executed by, the NAOs. The rest of the paper is organized as follows. Motor primitive generation approaches are described in Section 2. Subsequently, the structure of a

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*, Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_19,

motor-primitive framework is proposed in Section 3, and an example of motor primitive generation is shown in Section 4. Finally, in Section 5, we offer concluding remarks on our approach.

2 Related Work

Recently, motor primitives have been established via learning by manipulating robots directly instead of manually defining human-like motor primitives. In this section, we discuss approaches to establish learned motor primitives based on learning by demonstration.

One line of research related to the automatic generation of motor primitives concerns task generation by dividing the observed consecutive motor primitives [2]. The positions at which the motor primitives are divided are determined by a cost function that calculates the movement variants in the motor primitives. The tasks of the robot are also generated automatically by dividing series of spatio-temporal data by the SEG-2 [3], taking the sum of all squared values of joints into consideration. Expressing the spatio-temporal data by spatial coordinates gives rise to a space problem. Hence, the amount of the space is reduced by a method that uses the Isomap algorithm [4]. Moreover, the kinematic centroid segmentation algorithm is also used to divide consecutive data. In another approach, consecutive movements are divided by using the Maximin-selection algorithm [1]. This approach generates all possible series of movements and selects some as representative motor primitives.

In this paper, we propose a method that improves the Maximin-selection algorithm to compare motor primitives accurately, and a framework that generates motor primitives including an improved comparison method. We then apply the framework to the humanoid robot, NAO.

3 Humanoid Robot Framework

The proposed framework comprises a server and NAOs (Fig. 1). Each user controls a NAO by touching the sensors of the arms and head and grabbing and moving parts of the NAO. A NAO has two modes of operation: one for learning, and the other for performance. The user touches the arms to make the NAO walk, stand, and execute motor primitives, and the head to change the mode of operation. To make the NAO walk, we use the embedded walking module. The data for all joints of the NAO are then sent to the server, which receives and preprocesses the data from multiple NAOs and then generates motor primitives for the NAOs. The generated motor primitives are downloaded from the server and then uploaded to the NAOs by using the motor primitive uploader.

A movement in the proposed framework is defined as follows. m_j is the j*th* measured movement, $f_{j,i}$ is the i*th* joint of m_j, and d_j is the duration of m_j. Each movement consists of ρ joints and a movement duration, as shown in Eq. (1). The movement sender in a NAO defines a new movement by measuring all joints of the NAO and then transfers the results to the movement receiver in a server.

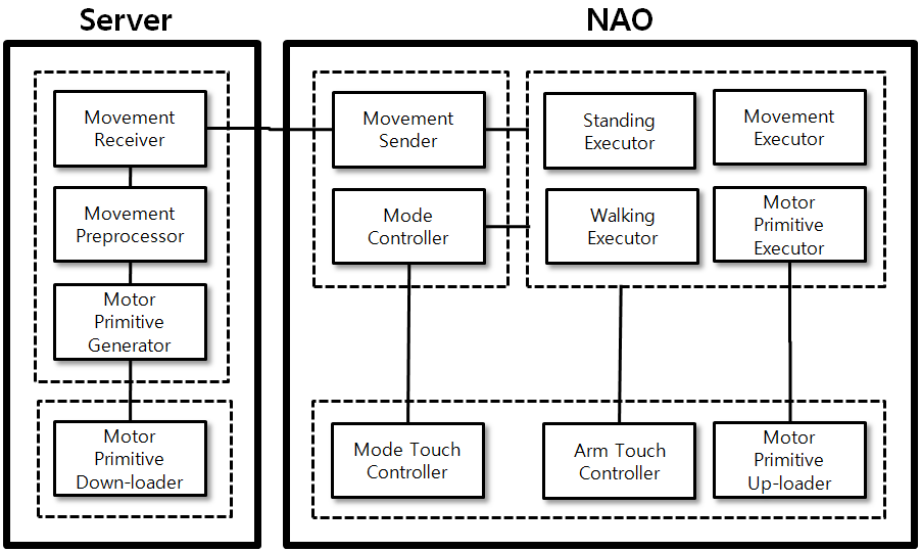


Fig. 1. Configuration of the proposed framework

$$m_j = <f_{j,1}, f_{j,2}, ..., f_{j,\rho}, d_j> \quad (1)$$

The movement preprocessor receives the transferred movements, compares them with the corresponding previous movements, and then calculates the difference between the two. If the difference is smaller than ζ, the later movements are eliminated. The motor primitive generator generates the motor primitives of NAOs by combining the movements of the movement preprocessor. A motor primitive is defined as follows.

$$p_k = \{m_{j'}, m_{j'+1}, ..., m_{j''}\},\ 1 \le j' \le j'' \le n \quad (2)$$

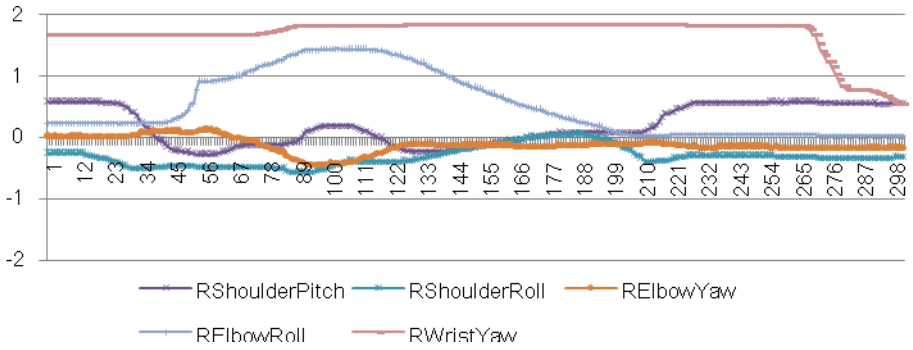
Here, p_k is the k*th* generated motor primitive; n is the number of transferred movements; and j' and j'' are greater than and equal to 1 and smaller than or equal to n, respectively. Therefore, each motor primitive is defined by multiple movements. To generate motor primitives, the Maximin-selection algorithm is used [7]. However, we use the Levenshtein-distance algorithm [5] instead of the Euclidean-distance algorithm to compare two motor primitives accurately. Given that Euclidean-distance algorithm compares the pairs of two movements of different motor primitives one by one, it is impossible to compare the two motor primitives that have different number of movements.

4 Experiment

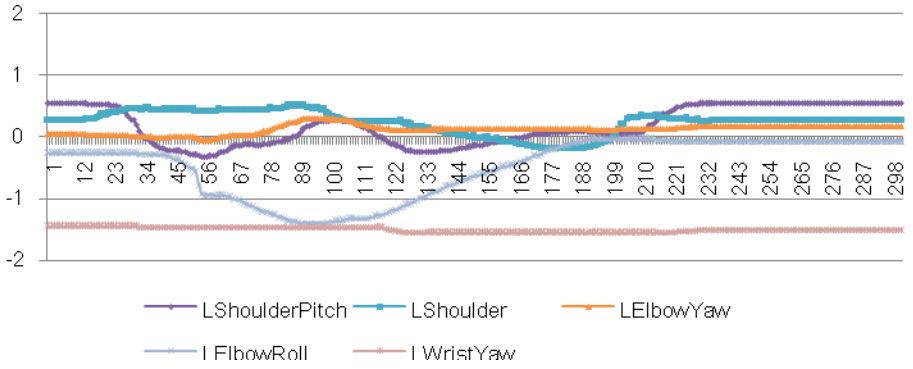
In this experiment, we verified the process of motor primitive generation based on the proposed framework. A NAO has 25 joints with values between –2 and 2. A subject manipulates the NAO to make it learn the movements for lifting a cup, as shown in Fig. 2.

Fig. 2. Lifting a cup for manipulation

Fig. 3 shows the movement of joints when the subject manipulates the NAO. Each arm has five joints for its shoulder, elbow, and wrist.



(a) Movements of the right arm



(b) Movements of the left arm

Fig. 3. The movement of NAO joints

The NAO then divides the consecutive movements and generates diverse motor primitives. Over the course of the experiment, 1487 candidate motor primitives were generated, after which 40 were selected by using the Maximin-selection and Levenshtein-distance algorithms.

5 Conclusion

In this paper, we proposed a framework for NAOs that collects movements and generates motor primitives. To improve the previous method based on a Euclidean-distance algorithm, the Levenshtein-distance algorithm was used to compare two motor primitives and increase the quality of the generated motor primitives. In experiments to test proof of concept, a NAO learned cup-lifting movements during a 30-s period and then generated 40 motor primitives.

Given that the size of NAOs is too small to provide services in real ubiquitous applications, the research to apply the proposed framework to human-sized humanoid robots is further required. Then the generated motor primitives can be executed by diverse kinds of algorithms such as Q-learning, HTN, and Bayesian probability. In the future, we will investigate methods to execute the generated motor primitives.

Acknowledgments. This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC(Information Technology Research Center)) support program (NIPA-2013-H0301-13-4007) supervised by the NIPA(National IT Industry Promotion Agency). And this work was also supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (2012R1A1A2009148).

References

1. Sung, Y., Cho, K.: An Actions Generation Method of Virtual Character using Programming by Demonstration. Journal of Korea Game Society 11(2), 141–149 (2011)
2. Koenig, N., Mataric, M.J.: Behavior-based segmentation of demonstrated task. In: International Conference on Development and Learning (2006)
3. Fod, A., Mataric, M.J., Jenkins, O.C.: Automated Derivation of Primitives for Movement Classification. Autonomous Robots, 39–54 (2002)
4. Jenkins, O.C., Matarić, M.J.: Deriving Action and Behavior Primitives from Human Motion. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2002), pp. 2551–2556 (2002)
5. Sung, Y., Cho, K.: An Expanded Levenshtein Distance Algorithm for Action Similarity Measurement of Virtual Characters. In: Proceedings of International Conference on Computer and Application (CCA 2012), Olympic Parktel, Soul, Korea, March 30-31, p. 149 (2012)

A Patient Status Classification Method for Metabolic Syndrome Care Based on Service Level Agreements

Sangjin Jeong[1,2], Chan-Hyun Youn[3], and Yong-Woon Kim[3]

[1] Dept. of Information and Communications Engineering, KAIST, Daejeon, Korea
[2] Protocol Engineering Center, ETRI, Daejeon, Korea
{sjjeong,qkim}@etri.re.kr
[3] Dept. of Electrical Engineering, KAIST, Daejeon, Korea
chyoun@kaist.ac.kr

Abstract. Chronic disease is a long term disease that requires life time care. Physicians need to keep tracking patients' status over time including routine medical examinations. There are a lot of medical resources and it is inadequate for physicians to review the huge number of medical information for a chronic patient. Thus, the results from these examinations and patients need to be classified according to patients' service level agreements. In this paper, to efficiently manage chronic disease patients, we propose a patient status classification method for chronic disease care based on service level agreements. We evaluate the proposed method using data obtained from the third Korea National Health and Nutrition Examination Survey among non-institutionalized civilians in the Republic of Korea, which was conducted by the Korean Ministry of Health and Welfare in 2005.

Keywords: healthcare, chronic disease, metabolic syndrome, service level agreement.

1 Introduction

Advances in patient caring and monitoring technologies have allowed physicians to track a patient's physiological state more closely and more accurately. In addition, these technologies enable out-of-hospital health monitoring. With the increasing amount of electronic medical data, system assisted medical decision should be adopted to effectively provide health care services. One of mostly popular systems for medical services is health information systems. Health information systems are being used to support key medical care procedures and to make medical decision and prescription, to manage patient's health conditions or even for hospital administration, respectively. In general, the traditional health information systems have been developed to generate raw patient examination results and have not provided expert advice for managing a patient's specific condition. Improved rules of a series of concerning preventive care tests and procedures, testing, drug therapies, and hospital stay status are used in a clinical practice process to manage a disease over time. However, the physicians usually overlook to examine the recommendations, or they

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_20,

hardly cooperate with the clinicians who disagree with the recommendations. Most of clinical expert systems have not directly integrated into the care process to provide suggestions about patient management when medical staffs visit a patient. To solve these service problems, it entails enormous cost to pay for human labor, to spend time and share information with other remote physicians and so on. The traditional systems that were available during the patient visit are used to disrupt the routine care process by requiring the clinician to enter additional data into the computer [1].

To efficiently manage chronic disease patients, we propose a patient status classification method for chronic disease, particularly metabolic syndrome care based on service level agreements. The proposed method classifies the status of patients' tier having metabolic syndrome using areal similarity degree analysis model and chronological distance function proposed by Jeong et al. [2][3]. The proposed method is evaluated using data obtained from the third Korea National Health and Nutrition Examination Survey among non-institutionalized civilians in the Republic of Korea, which was conducted by the Korean Ministry of Health and Welfare in 2005 [4].

2 Architecture of Patient Status Classification Method (PSCM) for Disease Care Service Levels Support

Systems which assist medical decision-making in hospitals are known as the health information systems. This systems support disease care procedures, medical decision-making, writing of prescriptions, and management of patients' health status. Among the health information systems, Jeong et al. proposed a preliminary Patient Status Classification Method (PSCM) to provide preliminary chronic disease diagnosis functionality [5]. The PSCM method has been further elaborated to quantify and predict the risk of having chronic disease, particularly metabolic syndrome (MS) in the future [6]. In this section, we describe an extended PSCM method with ambient visualization support for metabolic syndrome. Figure 1 shows the extended PSCM model with ambient visualization support. The extended model supports visualization of chronic disease status using ambient visualization interface. This model can aid physicians to deliver better healthcare services to their patients and better analyses of the patients' diseases, as it makes the workload manageable. The extended PSCM model for patients with chronic diseases offers automatic medical service procedures in the form of an effective medical information visualization system. It reduces the workload by offering readily available data. The PSCM process contains three parts: the Patient Tier Classifier, the Disease & Complications Identifier, and the Health Risk Quantification [5]. Our visualization system can be divided into an ambient information system and patient device. This visualization system can help a physician to manage each individual patient better and more efficiently than the conventional PSCM system. As the proposed visualization systems are extended to mobile devices by means of 'mobile interaction' technology and a 'widget interface', it can eventually generate more effective interaction with the health information system. Also, through the mobile widget on a device and with the order communication system interface, we can use data in conjunction with the medical diagnosis system as

a progressive medical service. Additionally, the individual patient receives an examination at their respective hospitals. The results of the examination are filtered to find whether or not the person who underwent the examination is a patient. The actual inspection ratio and the results of the Personal Examination Result set and the Disease database can then be utilized.

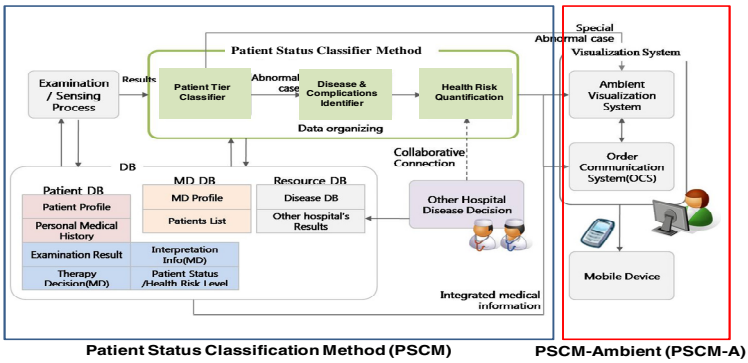


Fig. 1. The model system architecture for the extended PSCM supporing disease care service levels

To determine patient tier, we process patient status decision as shown in Fig 2. The results by Disease and Complications Identifier module can be generated at each individual hospital. The results obtained from the remote hospitals can be integrated with web-based system and can be calculated with the mapping module in the Health Risk Quantification module. In this step, the medical information analysis is shown in the health information system, and then physician can use the data at the examining time [7].

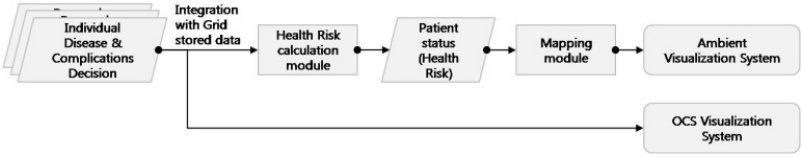


Fig. 2. Health risk quantification process

After a physician reviews all needed information from the health information system, they make a decision about the patient status. The patient status defined by physicians is to visualize on the health information system and ambient visualization system. The patient's condition can be mapped and compared to that of the normal levels in manual and then check it to see if the patient is at the risk of the chronic illnesses out of the normal levels and the mapping threshold are shown in Table 1. This risk is based on the patient, and this judgment depends on the physician's decision.

3 Determination of Thresholds for Patient Tier Classification for Service Levels Agreement Using Sensitivity Level

This section describes the determination of patient tier classification thresholds based the sensitivity level of the metabolic syndrome risk quantification (*ASD*; *Areal Similarity Degree*) model and the temporal change analysis ($d_N(2)$; *Chronological Distance Function between two medical examination result*) model proposed by Jeong et al. [2][3]. Figure 3 shows the three service levels based on patient tier classification thresholds. In the figure, the green graph indicates that the current medical report has not been changed since the previous report. The blue and the red graph indicate that there are minor and major change between the current and the previous reports, respectively. Thus, a physician can easily recognize the temporal change on patient's disease status with ambient interface. In the basic service level, the sensitivity level α is determined as TH_1 to indicate major change of patient disease status. We chose 0.5 for the value of TH_1, so when the distance value $d_N(2)$ of the current report value and the previous one is greater than or equal to 0.5, it is notified to the physician that the patient disease status was changed since the last examination. When the $d_N(2)$ is less than 0.5, the examination results are not notified to the physician. For the standard service level, sensitivity level α is determined as TH_2 to indicate moderate and major change of patient disease status. We chose 0.75 for the value of TH_2. So, when $d_N(2)$ is greater than or equal to 0.25, the patient's status change is notified to the physician. For the premium service level, the sensitivity level α is determined as TH_3 to indicate changes including minor change of patient disease status. We chose 0.95 for the value of TH_3. So, when $d_N(2)$ is greater than or equal to 0.05, the patient's disease status change is notified to the physician.

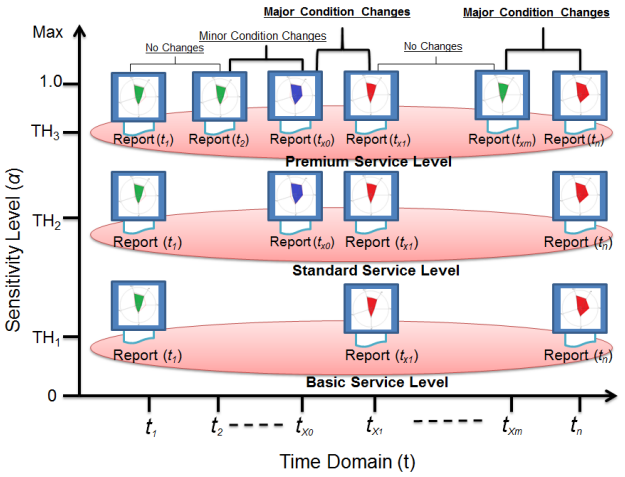


Fig. 3. Sensitivity for temporal change of disease status according to different values of α

According to the disease care service levels of metabolic syndrome shown in Fig. 2, we classify at-home services into three levels according to the patients' disease tiers, basic for *TIER(1)* and *TIER(2)*, standard for *TIER(3)*, and premium for *TIER(4)*, respectively. Table 1 lists the criteria for patients tier classfication. We consider metabolic syndrome only in this paper, so the Table 1 describes criteria for metabolic syndrome patients. However, the criteria may be applicable to other chronic diseases, if risk quantification models for other chronic diseases are established. Each patient's tier is determined based on the proposed tier classification criteria. The high risk patient is classified into *TIER(4)* group. A patient with medium risk is classified into *TIER(3)* group, while low or very low risk patient is classified into *TIER(1)* and *TIER(2)*, respectively.

Table 1. Patient's Tier Classification based on Metabolic Syndrome Risk Thresholds

Patient TIER (Threshold)	*TIER(1)* ($ASD < TH_1$)	*TIER(2)* ($TH_1 \leq ASD < TH_2$)	*TIER(3)* ($TH_2 \leq ASD < TH_3$)	*TIER(4)* ($TH_3 \leq ASD$)
Patient Health Risk	Very Low	Low	Medium	High

To establish the values of the thresholds for patient tier classification, we performed metabolic syndrome patient tier classification analysis using a large number of clinical data. The analysis is based on data obtained from the third Korea National Health and Nutrition Examination Survey (KNHANES III) among non-institutionalized civilians in the Republic of Korea, which was conducted by the Korean Ministry of Health and Welfare in 2005. This survey was a nationwide representative study using a stratified, multistage probability sampling design for the selection of household units. The survey consisted of the following 4 components: the Health Interview Survey, the Health Behavior Survey, the Health Examination Survey, and the Nutrition Survey [4]. A total of 34,145 individuals from these sampling frames were included in the health interview survey; among them, 25,161 subjects aged over 20 years were identified as potential participants in our study. We excluded those with incomplete data for the standardized analysis. This resulted in a final analytical sample of 5,355 subjects (2276 male, 3079 female), aged over 20 years. A total of 5,355 subjects (aged over 20 years) were included in this paper. The proportion of female subjects was higher than male (57.47% vs. 42.53%). The mean ages of the male and female subjects were 47.22±14.61 and 46.99±15.62 years, respectively. The average BMI was 23.99±3.10 and 23.52±3.38 kg/m2 for male and female subjects, respectively. The percentage of subjects with diabetes mellitus in the male group was higher than that in the female group (6.90% vs. 4.48%). Also, 22.98% of the male subjects had hypertension, whereas 14.68% of the female subjects did. In this paper, we have classified total subjects into two subject groups by gender and further classified each subject group into three sub-groups by age: young-adult (from 20 to 39 years old), middle-aged (from 40 to 64 years old), and old-aged (more than 65 years old), respectively. Therefore, we use a total of six sub-groups for the evaluation of our proposed risk quantification model [2]. To perform in-depth analysis regarding the determination of ASD thresholds, we further divided each sub-group into four Cases, as listed in Table 2. Since the objective of patient tier classification is to categorize chronic disease patients according to disease risk, we

chose thresholds for patients tiers based on the incidence of MS disease, i.e., the percentages of patients over given ASD values. Table 3 shows the criteria for determining tiers' thresholds.

Table 2. Detailed Sub-cases of Each Subject-group [2]

Case 1	A subject whose ASD value exceeds ASD threshold and having MS
Case 2	A subject whose ASD value exceeds ASD threshold and NOT having MS
Case 3	A subject whose ASD value does NOT exceed ASD threshold and having MS
Case 4	A subject whose ASD value does NOT exceed ASD threshold and NOT having MS

Table 3. Criteria for Determining Thresholds of Patients Tiers

Patient TIER	*TIER(1)* ($ASD < TH_1$)	*TIER(2)* ($TH_1 \leq ASD < TH_2$)	*TIER(3)* ($TH_2 \leq ASD < TH_3$)	*TIER(4)* ($TH_3 \leq ASD$)
Criteria for ASD values	Incidence of MS(Metabolic Syndrome) patients is			
	Less than 50%	Less than 75%	Less than 95%	Greater than or equal to 95%

Figure 4 shows the percentages of MS subjects over ASD values for each male subject.

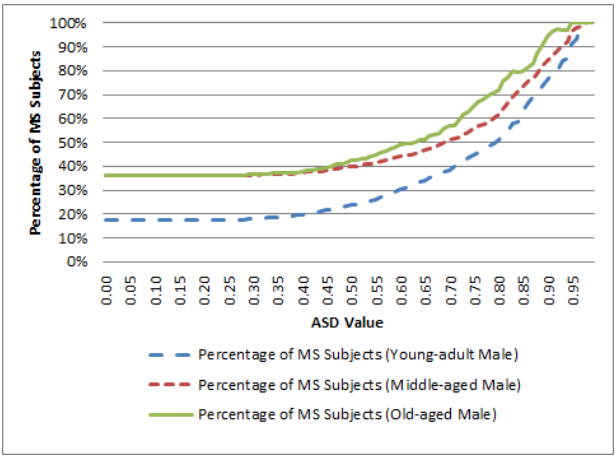


Fig. 4. ASD thresholds for TIER classification of male subjects

Figure 5 shows the percentages of MS subjects over ASD values for each female subject. Table 5 lists the determined ASD thresholds. To sum it up, the rate of MS is higher in female than in male for all ages, especially women are more susceptible to MS as they get older.

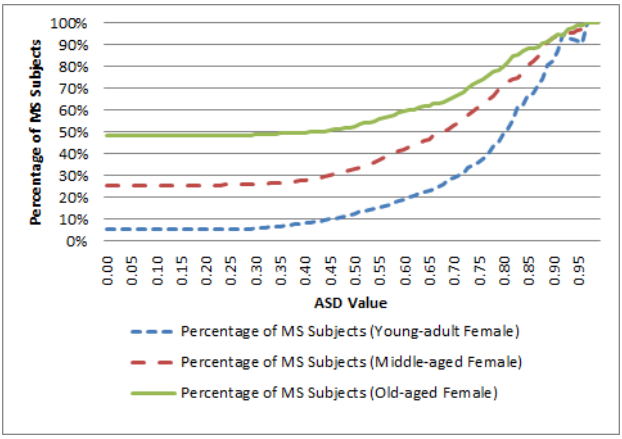


Fig. 5. ASD thresholds for TIER classification of female patients

As described in Table 3, we chose ASD thresholds according to the percentages of MS subjects. Table 4 and Table 5 list the determined ASD thresholds for male subjects and female subjects, respectively.

Table 4. ASD Thresholds for Male Subjects Sub-groups

Patient tier	TIER(1)	TIER(2)	TIER(3)	TIER(4)
Young-adult	$ASD<0.80$	$0.80\leq ASD<0.90$	$0.90\leq ASD<0.97$	$0.97\leq ASD$
Middle-aged	$ASD<0.69$	$0.69\leq ASD<0.86$	$0.86\leq ASD<0.95$	$0.95\leq ASD$
Old-aged	$ASD<0.63$	$0.63\leq ASD<0.81$	$0.81\leq ASD<0.91$	$0.91\leq ASD$

Table 5. ASD Thresholds for Female Subjects Sub-groups

Patient tier	TIER(1)	TIER(2)	TIER(3)	TIER(4)
Young-adult	$ASD<0.81$	$0.81\leq ASD<0.89$	$0.89\leq ASD<0.97$	$0.97\leq ASD$
Middle-aged	$ASD<0.68$	$0.68\leq ASD<0.84$	$0.84\leq ASD<0.92$	$0.92\leq ASD$
Old-aged	$ASD<0.43$	$0.43\leq ASD<0.77$	$0.77\leq ASD<0.93$	$0.93\leq ASD$

The gender differences on the MS related factors have been discussed in many literatures. Among them, Regitz-Zagrosek et al. reviewed on gender differences in the metabolic syndrome. They discovered the gender difference of the components in MS like glucose intolerance pattern, different lipid accumulation pattern in male and female, and its morphological change for postmenopausal women. These factors

affects higher incidence of MS in female group. Also the high development rate of MS in postmenopausal women can be related to sex hormone. The sex hormone which determines the physical and functional characteristics in male and female is thought to be the factor that affects the gender different glucose intolerance and lipid metabolism. Among them, it was revealed that estrogen has an important role in energy homeostasis and metabolic syndrome in both men and women from the studies using estrogen deficient animal models and in estrogen deficient men. A deficiency of estrogen like menopause and/or aging leads to higher incidence of metabolic syndrome in female than in male [8-9]. The distributions of frequency counts and percentage of MS subjects over *ASD* value shown in this sub-section indicate difference pattern among gender and age groups, which complies with the literatures. Therefore, we can claim that our proposed method effectively representing the risk of MS disease.

4 Conclusions

To efficiently manage chronic disease patients, we proposed a patient status classification method for chronic disease, particularly metabolic syndrome care based on service level agreements. The proposed method classified the status of patients' tier having metabolic syndrome using areal similarity degree analysis model and chronological distance function. We evaluated the proposed method using data obtained from the third Korea National Health and Nutrition Examination Survey among non-institutionalized civilians in the Republic of Korea, which was conducted by the Korean Ministry of Health and Welfare in 2005. The evaluation results showed that our proposed method could classify patients' metabolic syndrome risk status.

Acknowledgment. This research was supported by the ICT Standardization program of MKE (The Ministry of Knowledge Economy).

References

1. March, R.: Delivering on the promise of personalized healthcare. Personalized Medicine 7(3), 327–337 (2010)
2. Jeong, S., Jo, Y.M., Shim, S.-O., Choi, Y.-J., Youn, C.-H.: A Novel Model for Metabolic Syndrome Risk Quantification Based on Areal Similarity Degree. Under Review on IEEE Transactions on Biomedical Engineering (2013)
3. Jeong, S., Youn, C.-H.: A Method for Identifying Temporal Progress of Chronic Disease Using Chronological Clustering. In: Under Review on 15th International Conference on E-Health Networking (2013)
4. Korea Centers for Disease Control and Prevention (KCDC). The Third Korea National Health and Nutrition Examination Survey, KNHANES III (2005), `http://knhanes.cdc.go.kr/knhanes/index.do`
5. Bae, S., Lee, M.K.: Definition and Diagnosis of the Metabolic Syndrome. Journal of the Korean Medical Association 48(12), 1157–1164 (2005)

6. Jeong, S., Youn, C., et al.: An Integrated Healthcare System for Personalized Chronic Disease Care in Home-Hospital Environments. IEEE Transactions on Information Technology in Biomedicine 16(4), 572–585 (2012)
7. Hunt, K., Resendez, R.G., Williams, K., Haffner, S., Stern, M.: National Cholesterol Education Program versus World Health Organization metabolic syndrome in relation to all-cause and cardiovascular mortality in the San Antonio Heart Study. Circulation 110, 1251–1257 (2004)
8. Regitz-Zagrosek, V., Lehmkuhl, E., Weickert, M.: Gender differences in the metabolic syndrome and their role for cardiovascular disease. Clinical Research in Cardiology 95(3), 136–147 (2006)
9. Simpson, E.R., et al.: Estrogen—the Good, the Bad, and the Unexpected. Endocrine Reviews 26(3), 322–330 (2005)

FakePIN: Dummy Key Based Mobile User Authentication Scheme

Siwan Kim, Hyunyi Yi, and Jeong Hyun Yi*

Department of Computer Science and Engineering,
Soongsil University, Seoul, Korea
{kimsiwan,hyunyiyi,jhyi}@ssu.ac.kr

Abstract. Recently, smartphones have been used to store and manipulate a large amount of personal information. Hence, the importance of user authentication such as password setting has increased. Previous techniques to prevent a shoulder-surfing attack have considerable advantages with respect to security because they are mostly based on graphical features. They also have shortcomings in that the length of the password has to be memorized or that authentication takes a long time; hence, their usability is less than that of a text-based password technique. In this paper, we propose a dummy-key based password authentication scheme, called FakePIN, to assure good usability as well as to prevent shoulder-surfing, guessing, and smudge attacks.

1 Introduction

With the increase in the number of smartphones, users are able to plug into a variety of application services [1] regardless of time and place. However, if a smartphone is lost and with it important information, it may cause serious problems. Hence, user authentication such as password setting has increasingly become more important for the secure management of information stored in smartphones [2], [3]. The 4-digit PIN (Personal Identification Number) used in mobile devices has the advantage of easy and swift authentication but is vulnerable to the guessing attack that is capable of identifying the password within 10,000 trials. To cope with this vulnerability, hybrid password techniques between graphics and text are being proposed. For example, Android smartphones use a pattern-based graphical password input technique. The pattern lock is an authentication technique that makes the decision pattern on a 3×3 grid consisting of contact points. Because 389,112 different patterns can be constructed, this technique is about 38 times more secure than the 4-digit PIN based authentication. However, fingerprints (smudges) on the touch panel cause the movement of the user's fingers to be exposed. Therefore, smartphones are vulnerable to the so-called smudge attack [4]. In addition, smartphones are also vulnerable to the shoulder-surfing attack whereby an attacker glances over a user's shoulder and

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_21,

records the password. Camera and type-sound recording can also be used to execute this attack. To address this vulnerability, several studies various studies [5], [6], [7], [8], [9] are being conducted to develop graphic-based techniques that are not only resistant to shoulder-surfing attacks but also convenient to use. So far, most studies have focused on techniques for desktops with large display and ATM machines. However, these techniques require the memorization of a long password and authentication takes a long time. Thus, in this paper, a new text-based scheme is proposed that aims to prevent smudge and shoulder-surfing attacks, but is still convenient to use under mobiles conditions. The proposed scheme adopts a dummy key (faked one) as a user input value. The authentication technique, called "FakePIN [10]," consists of constructing a final password by internally combining the input value and a pre-set value.

This paper proceeds as follows. Section 2 deals with the proposed FakePIN scheme. Section 3 analyzes the security of FakePIN. Section 4 presents the usability test results of the proposed scheme. Lastly, section 5 presents the conclusion.

2 Proposed Scheme

Usually the text-based password schemes have high usability, but are very vulnerable to shoulder-surfing attacks and recording attacks. To circumvent these vulnerabilities, various password authentication techniques are introduced. Although these techniques are secure against shoulder-surfing attacks, authentication takes a long time or the user is required to remember a large amount of information. Thus, these techniques have a low usability. In addition, mobile devices such as smartphones have a limited amount of display room, so it is difficult to apply password techniques that require a large amount of display room. The method suggested in this paper has the advantages of the existing text-based password techniques. Thus, we propose FakePIN as a text-based password technique that is suitable for mobile devices.

FakePIN is composed of a password consisting of an alphanumeric text and a direction. For the direction, the user can select Up, Down, Left, or Right. In this paper, this secret information is called "direction password." The direction password involves the secret of allowing a key to be input in the direction the direction password is pointed in the original password. In other words, the user input value in the presence of a shoulder-surfing attack is a dummy key value (faked one). The original password is acquired by internally combining the input value and the previously-set direction password.

Fig. 1 shows an authentication example using FakePIN. Suppose that the original password has been set to <1234> and the direction password to <Up.> Then, during the user authentication step a virtual keypad is presented and the password must be input. The user presses <7> in the upper direction <Up> as a value input for <1.> Likewise, for <2> the user presses <6> for the upper direction <Up,> <5> for <3,> and <8> for <4.> Then, the user authentication process is finished. In addition, the keypad letters are randomly rearranged for each authentication, so a new input value is used as an authentication trial value. Since the faked dummy key value is different

for every authentication, an attacker fails in user authentication even when using the input password acquired through shoulder-surfing. This solves the problems with the existing text-based password techniques and, at the same time, strengthens the security. In FakePIN, the password is available in the form of alphabets and signs as a password, so the password used in the previous PIN method and that in the alphanumeric password are applicable as they are. Thus, it offers high compatibility with text-based systems and can be easily integrated with an intuitive and simple user interface. Therefore, the usability level is also high.

Fig. 1. FakePIN Authentication Phase. The User Presses the Digits <7658> for the Actual Password <1234> in Case of the Direction <Up. >

3 Security Analysis

In this section, we analyze the security of FakePIN and previous password based authentication techniques.

3.1 Password Space

The size of the password space is defined as the number of all possible passwords in the given techniques. Table 1 shows the password space comparison for the existing techniques where M indicates the number of possible letters and images to form a password, K the number of direction passwords, and N The length of the password.

Table 1. Password Space Comparison

	Password Construction	No. of Trials for Guessing	No. of Possible Passwords
DAS	*N*-digit number	*N*-digit number	10^N
PIN-Entry	*N*-digit number	Select black or white in every number and enter *N* times.	10^N
Dementor-SGP	User image, *N*-piece (including hole image)	Path from user image to hole image	$P(M-1,N)$
Passfaces	Face image, *N*-piece	Face image, *N*-piece	$P(M,N)$
M.TransKey	Number/character /symbol, *N*-digit	Number/character/symbol, *N*-digit	M^N
FakePIN	*K*-direction, number/character /symbol, *N*-digit	Number/character/symbol, *N*-digit	$K \times M^N$

In general, a 4-digit PIN has 10,000 password spaces. The number of possible passwords in FakePIN is $K \times M^N$. This value is higher than M^N, which is the standard number of possible alphanumeric passwords when using a smartphone for financial activities. In other words, this means that the guessing attack on FakePIN is as difficult as when financial services are being conducted over a smartphone. In the password configuration of FakePIN, numbers, alphabetic capital/lowercase letters, and signs are all used. In addition, the direction password is used in the form of the alphabetic password. The password space is *K* times as large as that of M.TransKey.

3.2 Guessing Attack Resistance

The chance that a guessing attack succeeds depends on the size of the password space. The larger the size of *M* and *N*, the smaller the chance that a guessing attack is successful. As for techniques such as Dementor-SGP, Passfaces, and others, it is possible to increase the size of *M* so that the security against the guessing attack becomes high. If the password of FakePIN consists of only numbers, it offers the same degree of security relative to the guessing attack.

Table 2. Resistance to Guessing and Recording Attacks

	Guessing Attack Probability	Recording Attack Probability
DAS	10^{-N}	1
PIN-Entry	10^{-N}	1
Dementor-SGP	$1 / P(M-1,N)$	$1 / M$
Passfaces	9^{-N}	1
M.TransKey	M^{-N}	1
FakePIN	M^{-N}	$1 / K$

However, if the password consists of numbers, alphabetic capital/lowercase letters, and signs, it does not affect the attacks, so it can be acquired in a guessing attack with a probability of M^{-N}. This offers higher security than DAS and PIN-Entry i.e., a 4-digit method, and has the size *M*, which is the same as that of alphanumeric

passwords used for payment services over smartphones. The size of N is also available for more than 4 digits, so it has a high resistance against a guessing attack. Hence, FakePIN has the same high security as the alphanumeric password relative to the guessing attack. Table 2 shows the comparison of existing techniques and FakePIN in terms of the probability of success of the guessing attack.

3.3 Shoulder-Surfing Attack Resistance

From analyzing the shoulder-surfing attack, it can be seen that the existing techniques, except for Dementor-SGP, allow the original user passwords to be input as they are, so they are very vulnerable to the shoulder-surfing attack. On the contrary, FakePIN uses a dummy password as a user input value, not the original password, and the value is different for every authentication. Therefore, the user password is not exposed to the attacker. If an attacker remembers not only the value used by a user but also the values of, the probability of success of the shoulder-surfing attack becomes high. However, the amount of information to be remembered for a one-time shoulder-surfing attack to be successful is quite large. The requirement to remember the values of Up, Down, Left, and Right add up to $4 \times N$. If the password is 8 digits long, then 32 digits must be remembered, all at the same time. This is very difficult to do without the help of an exterior device.

3.4 Recording Attack Resistance

To acquire the actual password, it is necessary to record all the user authentication processes and to combine the user input value with the adjacent four values of Up, Down, Left, and Right. Therefore, the success probability for a one-time recording attack is $1 / K$. Thus, FakePIN is secure against one-time recording attacks. However, if an attacker acquires two sets of information for the same password and performs an intersection attack to acquire the intersection of these two sets, then the attacker is in a position to discover the actual password. Thus, in an environment where the attacker records the whole authentication process using a camera, FakePIN is not secure from recording attacks for more than two times. Table 2 shows the comparison of cases acquired through one-time recording attack on FakePIN.

3.5 Smudge Attack Resistance

The pattern lock provided by Androids smartphones offers good security compared to the existing PIN method, and is widely used due to its convenience in inputting the password. But the trace of pattern left on a touch screen can be checked with the naked eye. In addition, if a camera, lighting, and a microscope are used, the chance of password exposure becomes very high. As for FakePIN, the keypad has randomly-arranged letters when user authentication is done. Thus, although a smudge may be left, its type will be different every time. For this reason, the password cannot be inferred from a smudge attack.

4 Experiments

In the design of user authentication techniques, it is always required to evaluate the balance between the usability and security of the user authentication process. To this end, we conducted a usability- measuring test based on the GOMS model and implemented the proposed scheme to smartphones

4.1 GOMS Model Test

To calculate usability evaluation data before a user test, a GOMS model was applied. Based on this, an objective comparison of FakePIN and the existing techniques is now available. As the GOMS model, Cog Tool 1.1.6 was used. The CogTool is a universal UI proto-typing tool as seen in Fig. 2. This tool is used to objectively predict and compare an application's execution time by defining each technique's function and operation on the story board and applying the user's operation time.

Fig. 2. FakePIN Analysis Using the CogTool

A test was conducted to measure individually the password setup time and the user authentication time of FakePIN and the existing techniques. Table 3 shows the comparison of the password setup time and the user authentication time.

Table 3. GOMS Test Results

	Password Setup Time (sec)	Authentication Time (sec)
DAS	12.46	13.18
PIN-Entry	12.46	18.71
Dementor-SGP	15.03	11.81
Passfaces	18.33	11.62
M.TransKey	23.48	11.33
FakePIN	15.46	10.90

4.2 User Test

To conduct a usability test and survey for FakePIN and the other techniques, 21 men and 9 women (30 in total) were screened. The group consisted of 10 users in their 10s, 10 in 20s, 5 in 30s, and 5 older than 40. In the usability test, each user was assigned a phone but were not told which authentication system was implemented. Then, the usability test was conducted 10 times in total; 5 times for measuring the password setup time and 5 times for the authentication time. As for usability measurement, the log data stored in the test equipment was used to collect the information on an error rate that occurred in both the password setup and the user authentication test cases. Table 4 shows the average password setup time, user authentication time, and average error rate.

Table 4. User Test Results

	Password Setup Time (sec)	Authentication Time(sec)	Error Rate (%)
DAS	14.67	11.56	8.7
PIN-Entry	18.80	21.18	23.3
Dementor-SGP	33.33	9.02	5.3
Passfaces	111.17	14.55	15.50
M.TransKey	26.17	9.27	13.30
FakePIN	25.63	14.13	4.70

5 Conclusion

We proposed FakePIN, a text-based password technique, which is more secure and usable than the text-based password used in smartphones. The proposed scheme replaces the original password with a password consisting of alphanumeric characters and one direction among Up, Down, Left, and Right, as an additional secret value. The proposed scheme is secure against shoulder-surfing, guessing, and smudge attacks, while not requiring any additional techniques or complex calculations. In particular, the proposed scheme showed higher security against the guessing attack, along with the alphanumeric password. It offers much higher security against the shoulder-surfing attack than the existing techniques, because an input value changes every time and there is much information to be remembered. In addition, the usability test showed that the user authentication time was a little slower than the other techniques, but also showed the lowest authentication error rate due to the authentication error rate of the simple and intuitive touch method. Thus, it was found that the proposed scheme as a whole had authentication time similar to those of all previously existing techniques and showed high usability as well.

Acknowledgment. This research was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2013R1A1A2013041).

References

1. Luo, H., Shyu, M.: Quality of Service Provision in Mobile Multimedia - a Survey. Human-centric Computing and Information Sciences 1(5) (2011)
2. El Kettani, M.D., En-Nasry, B.: MIdM: an Open Architecture for Mobile Identity Management. Journal of Convergence 2(2), 25–32 (2011)
3. Chuan, D., Lin, Y., Linru, M., Yua, C.: Towards a Practical and Scalable Trusted Software Dissemination System. Journal of Convergence 2(1), 53–60 (2011)
4. Aviv, A.J., Gibson, K., Mossop, E., Blaze, M., Smith, J.M.: Smudge Attacks on Smartphone Touch Screens. In: USENIX Security Symposium, WOOT (2010)
5. Jermyn, I., Mayer, A., Monrose, F., Reiter, M.K., Rubin, A.D.: The Design and Analysis of Graphical Passwords. In: USENIX Security Symposium (1999)
6. Weinshall, D.: Cognitive Authentication Schemes Safe against Spyware. In: IEEE Symposium on Security and Privacy (2006)
7. Wiedenbeck, S., Waters, J., Birget, J.: Design and Evaluation of a Shoulder-Surfing Resistant Graphical Password Scheme. In: Advanced Visual Interfaces (2006)
8. Hayashi, E., Christin, N., Dhamija, R., Perrig, A.: Use Your Illusion: Secure Authentication Usable Anywhere. In: Symposium on Usable Privacy and Security (2008)
9. Takada, T.: FakePointer: An Authentication Scheme for Improving Security against Peeping Attacks using Video Cameras. In: International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (2008)
10. Yi, J.H., Ma, G., Yi, H., Kim, S., Ji, H.: Method and Apparatus for Authenticating Password of User Device using Dummy Key. Patent App. No.: 10-2011-0128302, Korea (2011)

Image Compression System Using Colorization and Meanshift Clustering Methods

Taekyung Ryu, Byung Gook Lee, and Suk-Ho Lee*

Dept. Software Engineering, Dongseo Univ.,
San69-1, Jurye-2-Dong, Sasang-Gu,
Busan 617-716, Korea
petrasuk@gmail.com

Abstract. We propose a color image compression system which uses the colorization method used in the computer graphic society for automatic colorization and the meanshift clustering method. The encoder makes use of the meanshift clustering algorithm in automatically selecting the representative pixels(RP) from the original image from which the colored image is reconstructed in the decoder. Using the basin of attraction of the modes obtained by the meanshift clustering as the RP, the compression rate becomes high and the reconstructed image has good visual quality. Experimental results show that the proposed colorization method can provide good visual quality and reduce the smearing artifact which often occurs in other colorization based compression algorithms.

Keywords: Colorization, Meanshift Clustering, Image Compression, Color Image.

1 Introduction

Colorization is a method used in the computer graphic society to automatically colorize a grey image with only few color information [1]. The color information of these pixels is propagated to neighboring pixels by colorization methods making the whole image colorized. Colorization based coding refers to the color compression technique based on the use of colorization methods [2]-[5]. Colorization based coding utilizes the fact that the required number of pixels having color information is small. The encoder chooses the pixels required for the colorization process, which are called RP (representative pixels) in [5], and maintains the color information only for the RP. The position information and the color values are sent to the decoder only for the RP. Then, the decoder restores the color information for the remaining pixels using colorization methods. The main issue in colorization based coding is how to extract the RP such that the compression rate and the quality of the restored image become good.

In this paper, we propose a meanshift clustering [6] based RP selection method. The rationale behind the usage of the meanshift clustering is that the meanshift

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_22,

clustering has the capability of finding local maxima of the probability density function in a certain feature space. These local maxima are called 'Basin of Attractions of the Modes' (BAM) in [6]. When using the BAM as the RP, the RP have a main effect in the colorization process on the regions which approximately correspond to the regions clustered by the meanshift clustering. Therefore, it becomes possible to reconstruct all the colors to a sufficient level in the decoder. With the proposed method, there is no need to use any geometric methods, and the final RP set does not vary on manual initial settings, as is the case with other colorization based compression methods.

2 Previous Works

In this section, we introduce previous works which are required to understand the proposed scheme.

2.1 Colorization Method

Colorization is a technique used in the computer graphic society to automatically colorize a grey-scale image using only a few pixels having color information[1][6]. In [1], Levin *et al* proposed a colorization method based on optimization. The colorization process is performed by minimizing the following cost function with respect to $\mathbf{u}$:

$$J(\mathbf{u}) = \|\mathbf{x} - A\mathbf{u}\|^2. \tag{1}$$

Here, $\mathbf{u}$ is the solution vector, and $\mathbf{x}$ is the vector which contains the color values only at the positions of the RP, and zeros at the other positions. The vectors $\mathbf{u}$ and $\mathbf{x}$ are all in raster-scan order. Furthermore, $A = I - W$, where I is an $n \times n$ identity matrix, n is the number of pixels in $\mathbf{u}$, and W is an $n \times n$ matrix containing the w'_{rs} weighting components. The w'_{rs} weighting components are

$$w'_{rs} = \begin{cases} 0 & \text{if } r \in \Omega \\ w_{rs} & \text{otherwise,} \end{cases}$$

where

$$w_{rs} \propto e^{(y(r)-y(s))^2/2\sigma_r^2}. \tag{2}$$

Here, Ω denotes the set of the positions of the RP, σ_r^2 is a small positive value, and w_{rs} is the weighting component between the pixels at the r's and the s's positions, where $s \in N(r)$, and $N(r)$ is the 8-neighborhood of the r's pixel. Furthermore, $y(r)$ and $y(s)$ are the luminance values at the r's and the s's positions in the luminance channel, respectively. The minimizer of (1) can be explicitly computed as

$$\mathbf{u} = A^{-1}\mathbf{x}. \tag{3}$$

The obtained vector $\mathbf{u}$ used together with the luminance channel produces the colorized image.

2.2 Colorization Based Compression Framework

Colorization based compression makes use of the fact that there exists correlation between the luminance channel and the chrominance channels in the color image. Only the luminance channel is compressed by conventional compression standards such as the JPEG standard. For the encoding of the chrominance channels, first, the RP are extracted using the information of the luminance channel. These are then encoded and sent to the decoder. In the decoder, the chrominance channels are reconstructed from the RP and the luminance channel using colorization methods.

2.3 Meanshift Clustering

The meanshift method is a gradient ascent method which finds the local highest density of a data set via the use of the meanshift. By performing this meanshift procedure for all data points and clustering the data points which trajectories of gradient ascend lead to the same mode, the data points can be clustered. In this manner, by putting the color image into a five dimensional feature space, where three components represent the Y, Cb and Cr components and two components represent the x and y spatial coordinates of the pixel, the image can be segmented by a five dimensional meanshift clustering.

3 Proposed Method

The proposed scheme is based on the usage of the BAM(Basin of Attraction of the Modes) of the meanshift clustering [7] as the RP. First, we explain the rationale for the use of the BAM as the RP. Then, we explain the proposed scheme in details.

3.1 Rationale for the Use of the BAM as the RP

As explained in the previous section, the meanshift procedure is a gradient ascent method which finds the local maxima of the probability density function in a certain feature space. These local maxima are called 'Basin of Attractions of the Modes' (BAM) in [6]. To explain why we want to use the local maxima as the RP, we again examine the colorization process. In fact, (3) is actually a nonlinear diffusion equation in matrix form. The matrix A acts in fact as a weighted Laplacian operator in matrix form. Therefore, the matrix A^{-1} acts as an inverse operator of A, i.e., as a weighted 'diffusion' operator (except at the RP), where the weights are determined from the luminance channel. In other words, the matrix A^{-1} diffuses (or propagates) the colors of the RP in $\mathbf{x}$ to nearby pixels, resulting in the whole colorized image. Each color of the colorized pixels (except the RP) is the blended result of all the color propagations starting from every RP in $\mathbf{x}$. Since the BAMs refer to the feature vectors which have highest local densities in a local region, the BAMs serve as good starting points of the diffusion of the color components in the local region.

Figure 1 shows the usage of the BAM as the RP for reconstructing the colors in the decoded image. The BAM as obtained in Fig. 1(a) are used as the RP for colorization (Fig. 1(b)) to give the final reconstructed result (Fig. 1(c)).

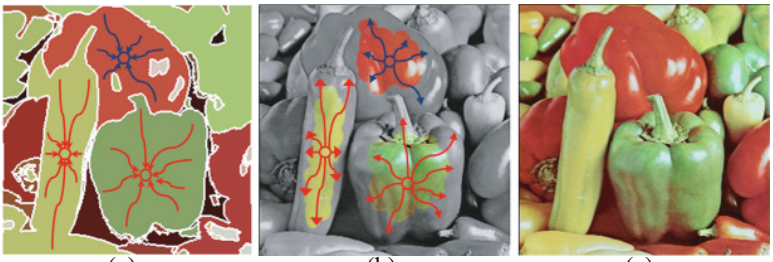


Fig. 1. Showing the concept of using the BAM obtained by the meanshift clustering as the RP for colorization. (a) obtaining the BAM by the meanshift procedure (b) using the BAM as the RP for colorization (c) colorized image.

However, the BAM (RP) obtained by a single meanshift procedure with constant kernel bandwidth are not enough to reconstruct a colored image with good quality. Regions with large variances in the color values need more RP than those with small variances to reconstruct the original color values with high fidelity. A single meanshift procedure with rather large bandwidth produces a number of RP which are enough for rather smooth regions, but not enough for regions with large variation. Therefore, we need to perform further meanshift procedures with smaller bandwidths in regions with large variations to obtain more RP. These steps have to be done iteratively until a sufficient RP set is obtained. The following section explains these steps in detail.

3.2 Proposed Encoder

Figure 2 shows the overall diagram of the proposed encoder. The main steps of the iterative meanshift approach are as follows: we first perform a meanshift clustering on the image with a kernel having a rather large bandwidth. Then we evaluate each segmented region by comparing the values of the reconstructed and the original color components in this region. If the color components are similar enough, the BAMs are assigned as the RP for this region, and no further meanshift procedure is performed for this region. If not, a further meanshift procedure is performed inside this region with a smaller bandwidth, and the BAMs of the resulting sub-regions are added to the RP set. Then, each sub-region is evaluated again to decide if any further meanshift procedure is needed. Thus, additional RP are added to the RP set if the variances between the original and the reconstructed color components are large in the previous step.

We explain the encoding steps in details. We denote by c_i a certain cluster (or segmented region) segmented by the meanshift segmentation, and denote by C_{set}

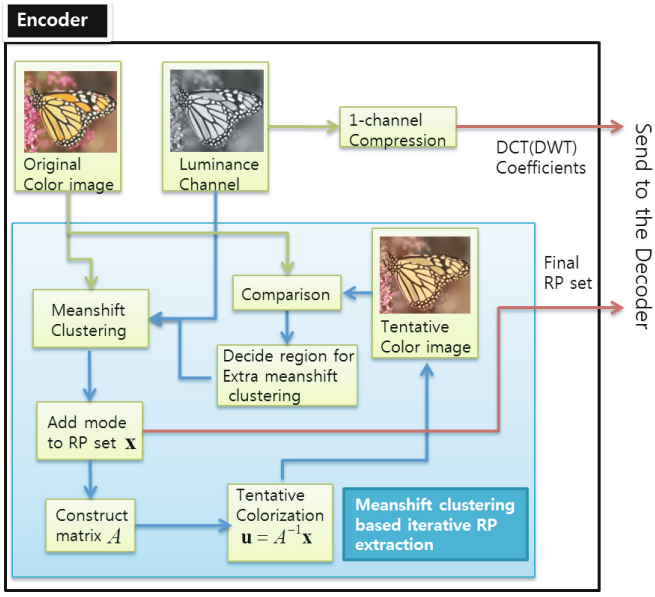


Fig. 2. Overall diagram of the proposed encoder

the set of clusters for which a further meanshift step has to be performed. We denote by C_{fin} the set of clusters for which no further meanshift segmentation is required. The procedure is done in the following steps:

1. Perform a meanshift based segmentation [6] on the whole image with predefined parameters (Hs, Hr, M), where Hs and Hr are the spatial and the range bandwidths of the meanshift kernel, respectively, and M is the minimal required number of pixels to be clustered as a cluster.
2. Register all the resulting clusters c_i to C_{set}, such that $c_i \in C_{set}, \forall_i$. For every cluster c_i in C_{set}, extract the BAM and assign them as the RP.
3. Setting the BAMs as the RP, reconstruct a tentative image by the Levin's colorization algorithm.
4. Evaluate the error between the values of the reconstructed and the original color components in every cluster c_i. If the error is large for that cluster, this means that additional RP have to be added in this cluster. We evaluate the error by comparing the error Err_{seg} measured in c_i with a predefined threshold value Thr. If $Err_{seg} < Thr$, register the cluster c_i to C_{fin} and process the next cluster. Otherwise, go to step 5. Here, the error metric is defined as $Err_{seg} = max(Cb_{err}, Cr_{err})$, where the function $max(a, b)$ takes the maximum of the two arguments, and $Cb_{err} = \frac{1}{K_{c_i}} \sum_{k=1}^{K_{c_i}} |f_{Cb}(\mathbf{r}_{c_i}) - \hat{f}_{Cb}(\mathbf{r}_{c_i})|^2$ and $Cr_{err} = \frac{1}{K_{c_i}} \sum_{k=1}^{K_{c_i}} |f_{Cr}(\mathbf{r}_{c_i}) - \hat{f}_{Cr}(\mathbf{r}_{c_i})|^2$, where f_{Cb} and f_{Cr} denote the original Cb and Cr color components, $\hat{f}_{Cb}$ and $\hat{f}_{Cr}$ denote the reconstructed color components, and K_{c_i} is the number of pixels in the

cluster c_i. $\mathbf{r}_{c_i}$ denotes the position vector for a pixel in the cluster c_i. The error Err_{seg} measures the mean error in the segmented region.

5. Perform mean shift clustering for every cluster $c_i \in C_{set}$ and $c_i \notin C_{fin}$ with the parameters set to $\left(\frac{Hs}{2}, \frac{Hr}{2}, \frac{M}{2}\right)$. The resulting sub-clusters will be registered as new clusters of C_{set} by step 2. Go to Step 2 and repeat steps 2-5, until the predefined iteration number is met or all clusters are registered to C_{fin}. If the predefined iteration number N_{set} is met, all the remaining clusters are forced to be registered to C_{fin}.
6. In the end of the algorithm, the RP in the clusters in C_{fin} become the desired set of RP.

3.3 Proposed Decoder

As in other colorization based compression methods, the decoder receives the compressed luminance channel and the RP set and reconstructs the color image using colorization methods. Here, we use the Levin's colorization method. In the decoder, the meanshift segmentation can be performed again on the decoded luminance channel. The segmented result will be the same as in the encoder since the same luminance channel is used. The segmentation result will then provide the decoder with information on the segmented regions. Then, the colorization effect of the RP can be restricted within the segmented region. In other words, the matrix A in the decoder can be constructed as in (3), i.e., $A = I - W$, however, this time the weights are weighted with the characteristic functions obtained by the meanshift segmentation:

$$\tilde{w}_{rs} = \begin{cases} 0 & \text{if } r \in \Omega \\ \gamma_{rs} w_{rs} & \text{otherwise,} \end{cases} \tag{4}$$

where γ_{rs} is the characteristic funtion of the meanshift segmented region containing the pixels r and s, i.e., it has the value 1 in the meanshift clustered region containing the pixels r and s, and is zero outside the region. Using the new matrix A, colors in other regions will not affect each other as happens often with the original Levin's colorization method.

4 Experiments

We performed the proposed scheme with $N_{set} = 4$, and let the initial parameters be $Hs = 14, Hr = 8, M = 40$. We compared the reconstructed results between the randomly selected RP encoded, the method of Cheng *et al.*[2] and the method of Ono *et al.* [5]. The sampling format used in the schemes is 4:2:0, which means that for the test image which has a 256×256 size, we recover the color components for a 128×128 size. The number of RP obtained with the proposed scheme is about 200. The same number of RP was used in the random RP using scheme. For the methods of Cheng *et al.* and Ono *et al.* the number of RP varies depending on the initial setting. We tried to make the size of files similar for comparison.

Table 1. Summarization of the comparison of the file size (KB), PSNR, SSIM values between the different colorization based coding methods

Image	Method	File Size	PSNR	SSIM(Cb)	SSIM(Cr)
Pepper	Proposed	0.684(KB)	25.4778	0.872	0.815
	Ono *et al*	0.724(KB)	22.04	0.781	0.757
	Cheng *et al*	0.7(KB)	23.49	0.872	0.796
Cap	Proposed	0.876(KB)	35.25	0.979	0.981
	Ono *et al*	0.992(KB)	32.5515	0.918	0.923
	Cheng *et al*	1.020(KB)	34.91	0.971	0.970
Cloth	Proposed	0.684(KB)	20.638	0.823	0.729
	Ono *et al*	0.752(KB)	17.605	0.623	0.542
	Cheng *et al*	0.7(KB)	19.675	0.814	0.679

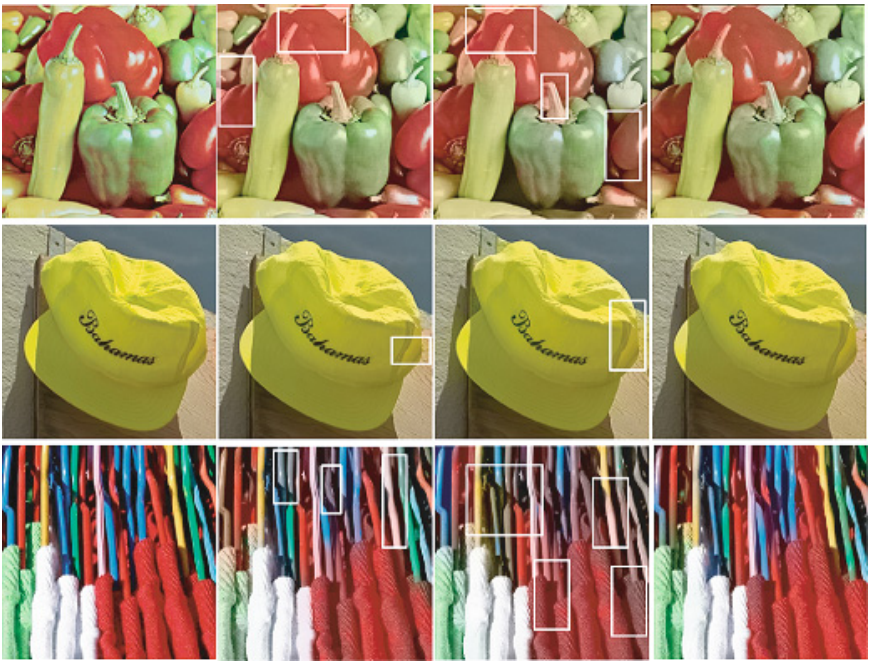


Fig. 3. Comparison of the encoded images of the proposed method and the JPEG standard. Column 1 : Original, Column 2 : Ono's method, Column3 : Cheng's method, Column 4 : Proposed.

Figure 3 and Table 1 compares the proposed scheme with the methods of Cheng *et al.* and Ono *et al.* The methods of Cheng *et al.* and Ono *et al.* reveal much color permeation as can be observed especially in the regions indicated with a white box in Fig. 3. Due to the false colors, the SSIM and the PSNR values become low. Compared with these methods, the proposed scheme shows less color permeation, and therefore, has higher SSIM and PSNR values. The reason that the proposed scheme shows less color permeation is due to the fact that the final RP set is closer to the optimal RP set, and also due to the fact that the colorization is restricted within the meanshift clustered regions by (4).

5 Conclusion

In this paper, we proposed a colorization based compression based on the usage of the meanshift clustering. Due to the close relationship between the 'Basin of Attraction of the Modes' (BAM) in the meanshift procedure and the RP in the colorization based coding method, the adoption of the BAM as the RP results in an RP set which gives good reproduction result of the color image in the decoder. Furthermore, the performance of the proposed scheme does not depend on manual setting of the initial RP set.

Acknowledgments. This work was supported by the Basic Science Research Program (2013R1A1A4A01007868) through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology.

References

1. Levin, A., Lischinski, D., Weiss, Y.: Colorization using Optimization. ACM Transactions on Graphics 23, 689–694 (2004)
2. Cheng, L., Vishwanathan, S.V.N.: Learning to Compress Images and Videos. In: Proc. ICML, vol. 227, pp. 161–168 (2007)
3. He, X., Ji, M., Bao, H.: A Unified Active and Semi-supervised Learning Framework for Image Compression. In: Proc. IEEE CVPR 2009, pp. 65–72. IEEE Press, Miami (2009)
4. Miyata, T., Komiyama, Y., Inazumi, Y., Sakai, Y.: Novel Inverse Colorization for Image Compression. In: Proc. Picture Coding Symposium, Chicago, pp. 1–7 (2009)
5. Ono, S., Miyata, T., Sakai, Y.: Colorization-based Coding by focusing on Characteristics of Colorization Bases. In: Proc. Picture Coding Symposium, Nagoya, pp. 11–17 (2010)
6. Yatziv, L., Sapiro, G.: Fast image and video colorization using chrominance blending. IEEE Trans. Image Processing 15(5), 1120–1129 (2006)
7. Comaniciu, D., Meer, P.: Mean Shift - A Robust Approach toward Feature Space Analysis. IEEE Trans. PAMI 24, 603–619 (2002)

Efficient Data Protection Scheme in Hybrid Clouds

Der-Kuo Tung[1], Wei-Hsiu Chen[2], and Chiang-Lung Liu[1]

[1] Department of Electrical and Electronic Engineering, Chung Cheng Institute of Technology, National Defense University, Taiwan ROC
{clliu,dktung}@ndu.edu.tw
[2] School of Defense Science, Chung Cheng Institute of Technology, National Defense University, Taiwan ROC
dtchentw@gmail.com

Abstract. The amount of digital data has rapidly increased at a rate of double per year since 2009. Adopting cloud computing storage services will consequently be an alternative approach to gain cost flexibility. Data encryption and access control are common data protection methods in cloud storages. However, traditional full data encryption in clouds can ensure privacy preserving but cause the low analysis efficiency problem. The emerging requirements of data analysis in clouds are balanced between data security and data analysis capacity. An efficient data protection scheme in hybrid clouds is proposed to improve data protection efficiency, rise up data availability and decrease the maintenance cost of cloud services. Data classification and selective data protection mechanisms are two key phases in this paper. According to the functional analysis and comparison results, The proposed scheme is therefore more suitable for efficiency data protection in hybrid clouds than traditional methods.

Keywords: Hybrid Clouds, Value at Risk, Anonymity, Selective Protection.

1 Introduction

The amount of digital data has rapidly increased at a rate of double per year since 2009 from IDC report [1]. High maintenance and management cost of data storage and archiving is therefore much higher than before and has become an unavoidable challenge for enterprises competitions. Adopting cloud computing storage services, a pay-per-use consumption, will consequently be an alternative approach to gain cost flexibility [2]. Typically, these are highly scalable resources delivered over the Internet to multiple companies, which pay only for what they use. Furthermore, large and complex data sets, so-called Big Data, are collected on the cloud services from everywhere and are processed using different data processing methods.

Data encryption and access control are common data protection methods for Big Data in the cloud [3-5]. [6] proposed a fully encrypt and partially decrypted method to protect data in public clouds. However, it would cause heavy workloads and

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*, Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_23,

degrading performance when numerous users access different encoded data at the same time.

The emerging requirements of Big Data analysis are balanced between data security and data analysis capacity. In [7], present fully encrypted cloud data often disturb the efficiency and effectiveness of data retrieve and query. Gentry [8] proposed a full homomorphic encryption method to realize search and calculation of encrypted cloud data without decryption to avoid data leakage. But, the full homomorphic encryption method is uneasy to implement. Sood [9] proposed a combined approach to classify data via (Sensitivity Rating, SR) in advance. Then, the approach built up an index of data and stored data with encrypted form in different section to provide keyword search in Cloud. However, whole encryption of data in the processing of data transmission and storage will degrade its' analysis performance.

An efficient data protection scheme in hybrid clouds is proposed to improve data protection efficiency, rise up data availability and decrease the maintenance cost of cloud services. Data classification and selective data protection mechanisms are two key phases in this paper. This paper is structured as follows: section 2 introduces main techniques of our scheme. In the section 3, a proposed scheme to provide cloud data security and utilization capacity is discussed in detail. Section 4 provides functional analysis and comparison with other previous methods. Section 5 concludes this paper.

2 Preliminaries

2.1 Anonymity Technique

The main goal in privacy preserving data mining (PPDM) is to design algorithms for trimming out from the original data in some way, so that the private data and private knowledge remain private even after the mining process [10-14]. Some PPDM techniques taxonomy in [11], including perturbation, blocking and swapping decrease the integrity levels and availability degrees of original data for privacy preserving. On the contrary, anonymity technique of PPDM can not only prevent original data from compromising other person's privacy, but also keep the trimmed data useful and available[12-14]. The anonymity technique is also used in this paper to gain the data protection. The details of anonymity technique start with the attributes of relational tables in a relational database system are generally classified into four categories [15].

- Identifier (ID) attribute can uniquely indentify a person, such as ID number, name, and cell phone number.
- Quasi-identifier (QID) attribute may reveal private information, such as zip-code, and race with the aid of external sources.
- Sensitive attribute (SA) contains sensitive information and need to be protected.
- Other attributes are not mentioned above.

The ID attributes will be certainly removed for private privacy information leakage prevention. And the QID attributes have potential mining values when they are cross-referenced with any attributes among different tables. Anonymity technique is used to

deal with the disclosure problem of QID and SA. Generalization and suppression are two common anonymity methods used in this paper. The former is to reduce precious values of any QID and sensitive attributes. And the latter is to remove or replace original attribute values with the highest generalization values. Then, the completeness of privacy data will be removed for sensitive information protection, while the truthfulness of the data is still preserved. The joint use of generalization and suppression helps in maintaining as much information as possible in the process of k-anonymity. The question is whether it is better to generalize, loosing data precision, or to suppress, loosing completeness

2.2 K-Anonymity

K-anonymity technique is a variant of anonymity technique proposed in [12-14] for microdata protection. K-anonymity demands that every tuple in the microdata table released be indistinguishably related to no fewer than k respondents. K-anonymity divides data into different groups and then replace original QID attributes values with the number of k same values in each group. The probability of sensitive attributes being identified is therefore decreased from one to 1/k through interactive inference.

Generalization consists in substituting the values of a given attribute with more general values. To this purpose, the notion of domain (i.e., the set of values that an attribute can assume) is extended to capture the generalization process by assuming the existence of a set of generalized domains. Each generalized domain contains generalized values and there exists a mapping between each domain and its generalizations. For instance, ZIP codes can be generalized by dropping, at each generalization step, the least significant digit; postal addresses can be generalized to the street (dropping the number), then to the city, to the county, to the state, and so on [12-14].

For each domain, the generalized domain represents with ordered hierarchy, called domain generalization hierarchy, denoted by DGH_D. A value generalization relationship associates with each value in domain and represents with hierarchy of tree, denoted by VGH_D. The measurement of K-anonymity can be defined with precision metric *Prec* (as Eq. 1 shown) [13]. $PT(A_1, A_2, \ldots, A_{NA})$is a table, and *RT* is a generalization of *PT*. Each DGH_A represents total height of the domain generalization hierarchy for attribute *A*, and *h* means the local height of attribute *A*.

$$prec(RT) = 1 - \frac{\sum_{i=1}^{N_A}\sum_{j=1}^{N}\frac{h}{|DGH_{Ai}|}}{|PT| * |N_A|} \tag{1}$$

3 The Proposed Scheme

The proposed data protection scheme contains two major phases: data classification and selective data protection (shown in Fig.1) to meet the requirements of high efficiency data protection in hybrid clouds. The values at risk (VaR) of enterprise data are identified in phase one. And the VaR results are used to partition data objects into

two sections for adopting different selective protection methods in phase two. Finally, high-risk and low-risk data are stored in hybrid clouds for authorized or unauthorized users, respectively. The details of this scheme are explained as following sections.

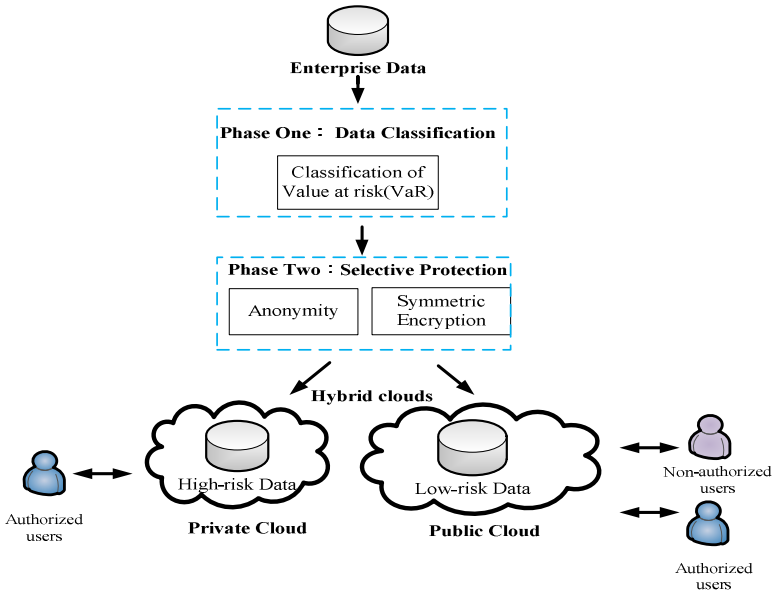


Fig. 1. The architecture of the proposed scheme

3.1 Data Classification

The improvement of data protection cost and complexity starts with the data classification procedure in this paper.

Step 1: Building attribute-value-conversion functions
The value conversion functions (as Eq. 2 shown) are used to assess the value of attributes.

$$V_{ij} = f(x_{ij}) \text{ , } (\ i = 1,...,n, j = 1,...,m \ ) \tag{2}$$

For example, x_{nm} is an item variable of the *Nth* attribute and *Mth* tuple. We take item variable x_{nm} into value conversion functions to get quantitative value, then we can build different value conversion functions according to different attribute categories. For categorical attributes, our study adopts hierarchical categorization method to divide attributes into several categories and each defined relative sensitive value. This method has advantages of enhancing speed of calculation of sensitive values. For numerical attributes, we build linear or non-linear functions to calculate sensitive values.

Step 2: Defining attribute weight

Present scheme define attribute weight in the privacy table according to security level of attributes. There are four attribute categories including of ID, QID, SA, and others in a relational database table and the security levels of them are ID > QID >SA>others. Our scheme also use a multi-criteria decision analysis technique, so-called Rank Order Centroid (ROC) [16], to accelerate attribute weight defining jobs for domain experts. The weight values are determined by attribute ranking and attribute number. If there are n attribute ranking $R_1,R_2,\cdots,R_n$, those weight will be $w_1,w_2,\cdots,w_n$, respectively. And it meets $1 > w_1 \geq w_2 \geq ... \geq w_n > 0$, where w_n is a weight of attribute in the table. If there are n attributes, the *kth* attribute-weight can be calculated by Eq.3：

$$w_k = \left(\frac{1}{n}\right)\sum_{k=i}^{n}\frac{1}{k}, (\ i = 1,...,n\ ) \tag{3}$$

Step 3: Calculating value at risk

We can get relative sensitive value of each item of attributes through attribute value conversion functions. Thus, we summarize the product obtained by multiplying sensitive value by attribute weight to get each tuple value at risk (VaR) R_j with Eq. 4.

$$R_j = \sum_{k=i}^{n} w_k * V_{ij}, (j = 1,...,m, i = 1,...,n) \tag{4}$$

Step 4: Classifying data by VaR

The security levels of each tuple are determined by a threshold T_A compared with VaR results. If $R_j > T_A$, then tuple T_j will be high-risk level. If $R_j < T_A$, then tuple T_j will be low-risk level.

3.2 Selective Data Protection

The efficiency of data protection in hybrid clouds depends on the outputs of data classification in phase one and the selective data protection methods in phase two. The proposed selective data protection methods for data distribution in public clouds are based on anonymity and symmetric encryption techniques. The former is used to add noise into the data while maintaining some overall statistical properties of the resulting table for unauthorized usage. And the latter is applied to rapidly encrypt and decrypt the required parts of low-risk data for authorized utilization. Therefore, each low-risk data in public clouds has two replicas for different usages.

The proposed selective data protection mechanism based on four attributes types. The main concept is that the important attributes (such as ID, QID and SA attributes) are processed in advanced to avoid information disclosure in public cloud. Considering the ID attribute has the characteristic of personal identification, it will cause privacy leakage problem if ID attribute is revealed. Meanwhile, it will lose data availability if all contents in ID attribute are deleted. The proposed scheme replaces original value of ID attribute with re-encode value. For example, the name or ID

number can be replaced with re-encode value, such as A001, A002, etc. The privacy preserving and information availability in public clouds are considered at the same time.

As to QID and SA attributes, the values of them are duplicated first and then anonymity and symmetric encryption techniques are simultaneously adopted. This method duplicates the field contents first. One copy is processed by anonymity; another copy is processed by symmetric encryption. Users can query and search the fields with anonymous values in public clouds after the precision level of data are adjusted. Non-authorized users are not allowed to access the fields with original data after symmetric encryption. Only authorized user can retrieve original value via a decryption key. On the contrary, the data of remaining other attributes are directly access in public clouds.

4 The Functional Analysis and Comparison

Tab. 1 lists the functional analysis and comparison results. Compared to other related protection methods, the proposed scheme provides not only functions of identity, authentication, access control and encryption, but also additional functions for data analysis in hybrid clouds (such as, data classification storage, keyword search, selective protection, data mining support, and hybrid cloud support). The proposed scheme is therefore more suitable for efficiency data protection in hybrid clouds than traditional methods.

Table 1. Functional analysis and comparison results

Functional Items	Gentry [8]	Sood[9]	Proposed scheme
Identity & Authentication	Yes	Yes	Yes
Access Control	Yes	Yes	Yes
Encryption	Yes	Yes	Yes
Data Classification Storage	No	Yes	Yes
Keyword Search	No	Yes	Yes
Selective Protection	No	No	Yes
Data Analysis Support	No	No	Yes
Hybrid Cloud Support	No	No	Yes

5 Conclusion

Traditional full data encryption in clouds can ensure privacy preserving but cause the low analysis efficiency problem. An efficient data protection scheme in hybrid clouds is proposed to improve data protection efficiency, rise up data availability and decrease the maintenance cost of cloud services. Data classification and selective data protection mechanisms are two key phases in this paper. Firstly, the enterprise data is evaluated in data classification process to produce VaR results for separating data objects into two sections for adopting different data protection methods in selective

data protection phase. Finally, high-risk and low-risk data are stored in hybrid clouds for authorized or unauthorized users, respectively. According to the functional analysis and comparison results, the proposed scheme is therefore more suitable for efficiency data protection in hybrid clouds than traditional methods.

References

1. IDC Inc. IDC IVIEW, http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf
2. Gartner Inc. Gartner Identifies the Top 10 Strategic Technology Trends for 2013, http://www.gartner.com/it/page.jsp?id=2209615
3. Li, J., Zhao, G., Chen, X., Xie, D., Rong, C., Li, W., Tang, L., Tang, Y.: Fine-grained data access control systems with user accountability in cloud computing. In: Proceedings of the 2010 IEEE Second International Conference on Cloud Computing Technology and Science (CloudCom), pp. 89–96. IEEE Press, Indianapolis (2010)
4. Kulkarni, G., Gambhir, J., Patil, T., Dongare, A.: A security aspects in cloud computing. In: Proceedings of the 2012 IEEE 3rd International Conference on Software Engineering and Service Science (ICSESS), pp. 547–550. IEEE Press (June 2012)
5. Mohamed, E.M., Abdelkader, H.S., El-Etriby, S.: Enhanced data security model for cloud computing. In: Proceedings of the 2012 8th International Conference on Informatics and Systems (INFOS), pp. 12–17 (May 2012)
6. Chuang, I.H., Li, S.H., Huang, K.C., Kuo, Y.H.: An effective privacy protection scheme for cloud computing. In: Proceedings of the 2011 13th International Conference on Advanced Communication Technology (ICACT), pp. 260–265 (2011)
7. Chaudhuri, S.: How different is big data? In: Proceedings of the 2012 IEEE 28th International Conference on Data Engineering (ICDE), p. 5. IEEE Press, Washington, DC (2012)
8. Gentry, C.: Fully homomorphic encryption using ideal lattices. In: Proceedings of the Symposium on the Theory of Computing (STOC), pp. 169–178 (2009)
9. Sood, S.K.: A combined approach to ensure data security in cloud computing. Journal of Network and Computer Applications 35(6), 1831–1838 (2012)
10. Vaidya, J., Clifton, C.: Privacy-preserving data mining: why, how, and when. IEEE Security & Privacy 2(6), 19–27 (2004)
11. Verykios, V.S., Bertino, E., Fovino, I.N., Provenza, L.P., Saygin, Y., Theodoridis, Y.: State-of-the-art in privacy preserving data mining. ACM SIGMOD Record 3(1), 50–57 (2004)
12. Sweeney, L.: K-anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems 10(5), 557–570 (2002)
13. Sweeney, L.: Achieving k-anonymity privacy protection using generalization and suppression. International Journal on Uncertainty Fuzziness and Knowledge-based Systems 10(5), 571–588 (2002)
14. Ciriani, V., Capitani di Vimercati, S., Foresti, S., Samarati, P.: k-Anonymity. Springer, New York (2007)
15. Xiao, X., Tao, Y.: Personalized Privacy Preservation. In: Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data, pp. 229–240 (2006)
16. Edwards, W., Barron, F.H.: SMARTS and SMARTER: improved simple methods for multi attribute utility measurement. Organizational Behavior and Human Decision Processes 60(3), 306–325 (1994)

Predicted Cost Model for Integrated Healthcare Systems Using Markov Process

Sangjin Jeong[1,2], Chan-Hyun Youn[3], and Yong-Woon Kim[2]

[1] Dept. of Information and Communications Engineering, KAIST, Daejeon, Korea
[2] Protocol Engineering Center, ETRI, Daejeon, Korea
{sjjeong,qkim}@etri.re.kr
[3] Dept. of Electrical Engineering, KAIST, Daejeon, Korea
chyoun@kaist.ac.kr

Abstract. Predicting cost in the health care environment is a challenging issue for healthcare professionals. Chronic disease is a long term disease that requires life time care. Physicians need to keep tracking patients' status over time including routine medical examinations. In this paper, we investigate the healthcare service scenario for chronic disease care. Then, we propose a predicted cost model for the investigated healthcare service. The cost model is based on the prediction of utilization and attendant costs through the development of a stochastic model, specifically a first-order Markov chain, can be adapted to specific diseases and/or events. The proposed model is at the initial stage and may require testing using a large administrative database of patients and hospital operational costs.

Keywords: healthcare, chronic disease, metabolic syndrome, service level agreement.

1 Introduction

Predicting cost in the health care environment is a challenging dilemma for medical professionals. The importance of a viable cost model incorporating outcomes measurement and payment schemes is of interest. Healthcare administrators want to assure that the delivery of services is appropriate as identified by government guidelines, rules and regulations. A critical starting point is to provide the framework necessary to provide a cost model that considers the general factors of healthcare encounters, patient diagnosis, treatment and the related costs that can be used to describe this complex problem. The stochastic nature of disease treatment can lead to substantial variation in experience between and among classes of enrollees, their diseases, and treatment utilization patterns. The most common approach to analyzing cost of healthcare service is the traditional method of summing the number of events occurring in the system over a period of time and calculating the mean and a standard deviation of cost. However, due to the emergence of new healthcare service models such as integrated healthcare system, there is an increasing demand for more sophisticated models to predict cost in the healthcare environments. A wide variety of

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*, Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_24,

conceptual and statistical models exist, both deterministic and stochastic, to measure utilization in health research. The typical types of deterministic models are traditional summing of events-based model, and decision tree-based decision analysis model. However, it is known that the limitation of the traditional models is its inability to account for non-symmetric aspect of cost data and the lack of consideration of the utilization patterns of the population. Decision tree-based models can be effective models in economic and policy analyses, because they can provide information to patients and practitioners about risk and cost. The difficulties with this model arise when timing becomes a concern. This problem becomes apparent when the time interval is several years or there are repeated events in a shorter time interval [7].

To resolve the limitations above, the Markov process has been widely used for modeling of epidemic progression of various diseases such as influenza, tuberculosis and HIV, and the pathways for subjects to predict utilization or possible pathways through the healthcare system [1-5]. Among them, to predict utilization in various healthcare plans and healthcare systems, Kapadia et al. studied 305 patients at a rehabilitation hospital over a six months period [5]. The authors developed a hospital service charging model by measuring the utilization patterns of the patients. Also, Beland studied ambulatory care in Canada. The author used the physician claims from clinic visits, hospitalization and emergency room visits and adopted a Markov chain to predict utilization to show the differences between population demographics such as age and gender. The author showed that the corresponding changes in the traditional model of counting visits to physicians can be modeled using Markov process. The results of the above literature review indicate that the Markov process is appropriate to estimate the utilization of a population of patients or enrollees. The Markov process can illustrate the difference in the treatment utilization patterns due to predictor variables such as gender and age [4]. Also, Leviton et al. examined the application of a generalized Markov process seems appropriate to predict utilization for patients with chronic or acute diseases [6]. In accordance with the previous literature, in this paper, we develop a Markov process-based cost prediction model for integrated healthcare system in home-hospital environment.

2 Markov Process Modeling for Integrated Healthcare Service

This section presents a predicted cost model for service scenario of home-hospotal environment integrated healthcare systems.Table 1 shows the states of integrated healthcare system in home-hospital environment used in this subchapter to develop the Markov process. The states represent typical disease management scenario of integrated healthcare system [7].

The Markov process is being employed as the first component of this model to predict utilization for this healthcare problem. The transition or change in utilization from state *i* to state *j* is influenced by the prior state is denoted as:

$$P_{ij} = \Pr[X_{n+1} = j \mid X_n = i], \tag{1}$$

where P_{ij} is the probability of going from state *i* to state *j* in one step or one increment in the time unit.

Table 1. States of integrated healthcare system in home-hospital environment for the Markov process

State 0 (m_0)	No identified chronic disease
State 1 (m_1)	Self-management at home
State 2 (m_2)	Consult an informant about abrupt change on at-home medical examination results
State 3 (m_3)	Visit hospital (periodic visit or abrupt disease state change)
State 4 (m_4)	Diagnosis by physician
State 5 (m_5)	Laboratory tests
State 6 (m_6)	Discharged from the hospital
State 7 (m_7)	Complete cure

Due to the patient's outpatient flow is managed by predetermined patient management care plan of hospital, it is not necessary to consider all the previous transitions, when determining next transition. All the transitions from the resident states require preset exit criteria and only the previous state influences the opportunity for transition to an alternate state. This means that being in a resident utilization state two transitions earlier is irrelevant to the state you are in presently. All the information that is needed is the previous transition state. This is using the memory less or Markovian property. The resident states of transition based on utilization for the Markov process will be defined as follows: The states defined in Table 1 lead to an eight states Markov chain with an absorbing barrier (State 7) and the resulting one step transition probability matrix is

$$\mathrm{P}_{7\times 7} = (P_{ij}), \text{ where } i = 0, 1, 2, \ldots, 7 \text{ and } j = 0, 1, 2, \ldots, 7 \tag{2}$$

We specify the time interval unit for the probability transition matrix for utilization to be one unit time T.

Consider the finite number of possible transitions for individuals in the model. Denote by the element

$$P_{00} = \Pr[X_n = 0 \mid X_{n-1} = 0], \text{ for any } n. \tag{3}$$

where the probability that an individual starting in m_0 stayed in m_0 after one time period.

Similarly, the probabilities for staying in the same state after one transition are denoted as follows:

$$P_{00}, P_{11}, P_{22}, P_{33}, P_{44}, P_{55}, P_{66}, P_{77} \tag{4}$$

$$\text{Let } f_{ii}^{n} = \Pr[X_n = i, X_v \neq i, v = 1, 2, \ldots, n-1 \mid X_0 = i] \tag{5}$$

be the probability that, starting from state i, the first return to state i occurs at the nth transition [8]. There are 64 possible transitions for an individual in this model. State m_7 (complete cure) is an absorbing barrier or state, defined as a state that once entered

cannot be exited [8]. This means that the probabilities of transition from m_7 to states m_0 through m_6 are zero and the probability of starting in state seven and staying in state seven is one. Hence in the probability transition matrix Eq. (2) becomes

$$P_{70} = P_{71} = P_{72} = P_{73} = P_{74} = P_{75} = P_{76} = 0 \;\&\; P_{77} = 1 \tag{6}$$

The matrix can be partitioned into 4 subsets. The set of transition probabilities, $\{P_{77}\}$, has the following two properties:

1. It has a period of one
2. Since $f_{77}^{n} = 1$, it is positive recurrent.

Combining the previous two properties leads to the conclusion that $\{P_{77}\}$ is an ergodic set. This set will be represented by the sub-matrix, $\mathrm{E}_{1\times1}$ and defined as follows:

$$\mathrm{E}_{1\times1} = \{P_{77}\} \tag{7}$$

The next submatrix of the partitioned matrix to be considered is the vector of zeroes,

$$\mathrm{O}_{1\times7} = \{P_{7\mathrm{i}} \mid i = 0,1,2,3,4,5,6\}. \tag{8}$$

Once in the absorbing state the individual cannot leave the state; hence all the cell entries axe zero. The third partitioned sub-matrix to be defined will be the transient states,

$$\mathrm{M}_{7\times7} = \{P_{\mathrm{ij}} \mid i = 0,1,2,3,4,5,6 \;\&\; j = 0,1,2,3,4,5,6\}. \tag{9}$$

The sub-matrix M includes all the transient states of the Markov chain. The probability of the first return, Eq. (5), for these states (m_0, m_1, m_2, m_3, m_4, m_5, m_6) is less than one. The final sub-matrix to consider is the transition from a transient state to the absorbing state. This will be defined as

$$\mathrm{L}_{7\times1} = \{P_{ij} \mid i = 0,1,2,3,4,5,6\} \tag{10}$$

An alternative form of the probability transition matrix can now be illustrated with dimensions of partitioned matrices:

$\mathrm{E}_{1\times1}, \mathrm{O}_{1\times7}, \mathrm{L}_{7\times1}, \mathrm{M}_{7\times7}$. It should be noted that the matrix $\mathrm{E}_{1\times1}$ is equivalent to the identity matrix, $\mathrm{I}_{1\times1}$. Replacing $\mathrm{E}_{1\times1}$ with $\mathrm{I}_{1\times1}$ in the matrix results in the following

$$\mathrm{P} = \begin{pmatrix} \mathrm{I}_{1\times1} & \mathrm{O}_{1\times7} \\ \mathrm{L}_{7\times1} & \mathrm{M}_{7\times7} \end{pmatrix} \tag{11}$$

3 Predicted Cost Model for Integrated Healthcare Systems Service Scenario

Determining the mean time or number cycles an individual occupies in a resident state requires some knowledge of linear algebra and the development of the fundamental matrix for Markov chain with an absorbing state. Kemeny and Snell [9] developed methodology for finding the mean time in each resident state before transition into the absorbing state. They proposed the following.

Let $\mathrm{B}_{n\times n}^{m}$ be a square matrix raised to the power m. If $B^m \to 0 \text{ as } m \to \infty$, then $(\mathrm{I}-\mathrm{B})$ has an inverse, and

$$(\mathrm{I}-\mathrm{B})^{-1} = \mathrm{I}+\mathrm{B}+\mathrm{B}^2+\ldots = \sum_{i=0}^{\infty}\mathrm{B}^i \tag{12}$$

For any Markov chain with an ergodic set, let the matrix M correspond to the set of transient states, as in Eq. (11). Then $(\mathrm{I}-\mathrm{M})$ has an inverse, and

$$(\mathrm{I}-\mathrm{M})^{-1} = \mathrm{I}+\mathrm{M}+\mathrm{M}^2+\ldots = \sum_{i=0}^{\infty}\mathrm{M}^i \tag{13}$$

Substituting the matrix M from Eq. (11) into Eq. (12) proves Eq. (13).

Let

$$\mathrm{N} = (\mathrm{I}-\mathrm{M})^{-1} \tag{14}$$

be the fundamental matrix for a Markov chain with an ergodic state [9]. The next consideration is the number of times for an individual that a transient state is occupied. Define η_{lj} to be the function assigning the total number of times that the process is in state m_j after starting from state m_l (restricting the choices to transient states, $\{m_j \mid j = 0,1,2,3,4,5,6\}$. This quantity will be will be expressed as the sum of indicator variables

$$\mu_{lj}^{k} = \begin{cases} 0, \text{ if the process is in state } m_j \text{ after } k \text{ steps} \\ 1, \text{ otherwise} \end{cases} \tag{15}$$

Determining the expectation of the number of cycles an individual stays in a resident transition state, conditional on having just entered the system, follows [9]. The mean number of days spent in m_j after starting in state m_l is $\mathrm{N}_{7\times 7} = E[\eta_{lj}]$ as can be seen from the following argument. It should be observed that $\eta_{lj} = \sum_{k=0}^{\infty}\mu_{lj}^{k}$. Hence,

$E[\eta_{lj}] = E\left[\sum_{k=0}^{\infty} \mu_{lj}^{k}\right]_{7\times 7}$. Note that the μ_{lj}^{k} the l, j element of M^{k}. Here η is the matrix whose l, j element is M^{k}. Then

$$E[\eta] = \sum_{k=0}^{\infty} E[\mu_{lj}^{k}] = \sum_{k=0}^{\infty} \mathrm{M}^{k} = \mathrm{N} \tag{16}$$

Denote the expected numbers of days in seven transient states by T', taken from the proper row of *N*.

The notation for the cost function is Eq. (17). Define the fixed cost to be a column vector, where each element of this 8×1 matrix is the averaged costs per utilization state of the system. It should be noted that the elements for the states "no use of services" and "Complete cure" have no allowable costs associated with them. The cost function can be represented as Eq. (17)

$$C_{8\times 1}^{'} = \{0, c_1, c_2, c_3, c_4, c_5, c_6, 0\} \tag{17}$$

The model is defined by multiplying T' and Eq. (17) with the result

$$\mathrm{F}(x_i) = \mathrm{T}_{8\times 1}^{'}(x_i)\mathrm{C}_{1\times 8}, \tag{18}$$

where x_i is the conditions of interest (gender, age, and diagnosis).

The value of the function F is the predicted cost for an individual. The vector of utilization, T', has a dimension of 1×8. The vector of cost, C, has a dimension of 8×1. Taking their product generates a scalar value, $\mathrm{F}_{1\times 1}$, which is the predicted cost given gender, age, and diagnosis.

4 Conclusion

Predicting cost in the health care environment is a challenging issue for healthcare system professionals. In this paper, we presented Markov process-based cost model to predict costs based upon healthcare service scenario of home-hospital integrated environments. The proposed cost model has following limitations. If the number of transitions is small in one or more resident states with the addition of one or more resident states becoming ergodic, then an unstable probability transition matrix is generated. The unstable matrix cannot provide appropriate estimates. The proposed model has not been tested on a large administrative database of claims. However, during the test phase of the proposed model, the following issues need to be considered. The first issue is that the restrictions due to the database with the lack of demographic identifiers. The absence of ethnicity and marital status may lead to questions about the changes in utilization patterns for these groups of enrollees. A second issue about evaluation is the choice of deleting the multiple events per day for an individual. This may cause the states remaining transient in the probability transition matrices. Therefore, the issues above need to be considered during further evaluation of the proposed model.

Acknowledgment. This research was supported by the ICT Standardization program of MKE (The Ministry of Knowledge Economy).

References

1. Aalen, O.O., Farewell, V.T., De Angelis, D., Day, N.E., Gill, O.N.: A Markov model for HIV disease progression including the effect of HIV diagnosis and treatment: Application to AIDS prediction in England and Wales. Statistics in Medicine 16, 2191–2210 (1997)
2. Auranen, K., Ranta, J., Takala, A.K., Arjas, E.: A statistical model of transmission of Hib bacteria in a family. Statistics in Medicine 15, 2235–2252 (1996)
3. Beck, J.R., Paulker, S.G.: The Markov process in medical prognosis. Medical Decision-Making 3(4), 419–458 (1983)
4. Beland, F.: Conceptualizing the utilization of ambulatory medical care as a process. Medical Care 26(2), 115–123 (1988)
5. Kapadia, A.S., Vineberg, S.E., Rossi, C.D.: Predicting course of treatment in a rehabilitation hospital: A Markovian model. Computers & Operation Research 12(5), 459–469 (1985)
6. Leviton, A., Schulman, J., Kammerman, L., Porter, D., Slack, W., Graham, J.R.: A probability model of headache recurrence. Journal of Chronic Disease 33, 407–412 (1980)
7. McGhee, C.R., Glasser, J.H., Chan, W., Pomeroy, N., Chan, W.: Forecasting Health Care Expenditures and Utilization Based on a Markov Process and a Deterministic Cost Function in Managed Care Settings. Lecture Notes-Monograph Series, vol. 43, pp. 229–238. Statistical Essays in Honor of Jack Hall, Crossing Boundaries (2003)
8. Ross, S.M.: Introduction to Probability Models, 3rd edn. Academic Press, San Diego (1997)
9. Kemeny, J.G., Snell, J.L.: Finite Markov Chains. D. Van Nostrand Company, Inc., Princeton (1960)

Gaussian Mixture Model Based on Hidden Markov Random Field for Color Image Segmentation

Khoa Anh Tran, Nhat Quang Vo, Tam Thi Nguyen, and Gueesang Lee*

Department of Electronics and Computer Science,
Chonnam National University, Gwangju, South Korea
{anhkhoayy,vqnhat,nguyentamdtmc}@gmail.com, gslee@jnu.ac.kr

Abstract. Gaussian Mixture Model (GMM) has been widely applied in image segmentation. However, the pixels themselves are considered independent of each other, making the segmentation result sensitive to noise. To overcome this problem for the segmentation process we propose a mixture model useing Markov Random Filed (MRF) that aims to incorporate spatial relationship among neighborhood pixels into the GMM. The proposed model has a simplified structure that allows the Expectation Maximization (EM) algorithm to be directly applied to the log-likelihood function to compute the optimum parameters of the mixture model. The experimental results show that our method has more advantage in image segmentation than other methods in terms of accuracy and quality of segmented image, and simple performance.

Keywords: Gaussian mixture models, Image segmentation, Hidden Markov random field, Expectation Maximization, Log-likelihood.

1 Introduction

Today, color image segmentation is useful in many applications in image processing and recognition systems. The results of image segmentation play important roles in providing more information for better imagery diagnosis. However, image quality is always impacted, even being corrupted, by high levels of noise. This is one of challenging works in accurate image segmentation. Several previous works have been carried out on image segmentation, in particular like mean shift-based method [1], graph-based method [2], clustering approach [3] etc. In recent years, Bayesian framework-based approaches [4] have gained popularity due to their simplicity and ease of implementation. Among these methods, standard Gaussian Mixture Model (GMM) [5], [6] is well known. One of the main drawback of this method is that the prior distribution π_j does not depend on the pixel index j and thus, not on the spatial relationship between the labels of neighboring pixels. Thus, the segmentation is extremely noise prone and illumination dependent. To overcome this disadvantage, mixture models with Markov Random Field (MRF) have been employed for pixel

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_25, © Springer-Verlag Berlin Heidelberg 2014

labeling [7], [8]. The distinct difference is that the prior distribution π_{ij} varies for every pixel x_i corresponding to each label L_j and depends on the neighboring pixels and the corresponding parameters. The disadvantages of the MRF methods lie in lacking robustness against high amount of noise and increase in computational cost. Several researchers have extended the models [9], [10] where an MRF models the joint distribution of the priors of each pixel, instead of the joint distribution of the pixel labels. A special type of MRF model noted as Hidden Markov Random Field (HMRF), is a stochastic process generated by a MRF. Its state sequence cannot be observed directly, but it can be observed through a field of observations [11]. HMRF theory provides a systematic approach for deriving optimality criteria such as those based on the maximum a posteriori (MAP) or the marginal posterior modes (MPM) [12]. Before applying the standard optimal criteria, model parameters have to be estimated. In recent years, several optimization algorithms have been proposed, and the Expectation Maximization (EM) algorithm is the most widely used by solving maximum likelihood (ML) estimation of the data. The spatially variant finite mixture model (SVFMM) was proposed in [13], where, a MAP estimation is used on Markov random field prior information of the pixel labels. Importantly, segmentation accuracy is quite sensitive to the initialization of segmentation algorithm by using the local optimization parameter estimation algorithms such as EM. Therefore, the initializations for these local algorithms have to be well selected and determined. In this work, a GMM based HMRF is proposed. The model is made based on following considerations – firstly, we use the HMRF-EM algorithm with GMM parameter for estimation that is capable of avoiding the initialization sensitivity problem. Secondly, the spatial information has been successfully incorporated in the model that allows the orders neighborhood system can being directly applied to compute the input parameters.

The content of this paper is organized as follows: In section 2, we present brief Hidden Markov Random field theory and our method. Section 3 includes the experiment results and finally the conclusion in section 4.

2 Methods

2.1 Hidden Markov Random Field Theory

Let x_i, $i = (1,2...,N)$, where each x_i is of dimension D, denote an observation at the ith pixel of an image. The neighborhood of the ith pixel is presented by δ_i. The target is to associate each x_i with a label in $(1,2...,K)$. For this classification, standard GMM assumes that each observation x_i is independent of the label L_j. The density function $f(x_i \mid \Pi, \Theta)$ at an observation x_i is given by:

$$f(x_i \mid \Pi, \Theta) = \sum_{j=1}^{K} \pi_{ij} \Phi(x_i \mid \Theta_j) \tag{1}$$

where, $\Pi = \{\pi_{ij}\}$, $j = (1,2,...,K)$ is the set of prior distributions of probabilities where π_{ij} denotes the probability that pixel x_i is in label L_j and satisfies the constraints:

$$0 \le \pi_{ij} \le 1 \quad \text{and} \quad \sum_{j=1}^{K} \pi_{ij} = 1 \tag{2}$$

Also, $\Phi(x_i \mid \Theta_j)$ is a component of the Gaussian mixture. Each component can be written in the form:

$$\Phi(x_i \mid \Theta_j) = \frac{|\Sigma_j|^{-1/2}}{(2\pi)^{D/2}} \exp\left\{\frac{\Delta^2}{2}\right\} \tag{3}$$

where $\Delta^2 = (x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j)$ is the squared Mahalanobis distance and $\Theta_j = \{\mu_j, \Sigma_j\}$, $j = (1,2,...K)$. The D-dimensional vector μ_j is the mean, the $D \times D$ matrix Σ_j is the covariance, and $|\Sigma_j|$ denotes the determinant of Σ_j. From Eq.(1), the joint conditional density of the data set $X = (x_1, x_2, ..., x_N)$ can be written as:

$$p(X \mid \Pi, \Theta) = \prod_{i=1}^{N} \left[\sum_{j=1}^{K} \pi_{ij} \Phi(x_i \mid \Theta_j) \right] \tag{4}$$

With the observation that x_i is considered to be independent given the pixel label, the spatial correlation between the neighboring pixels is not taken into account. In natural images, the neighboring pixels are highly correlated if they belong to the same object. If the correlation is not used, the segmentation can be very sensitive to noise, varying illumination and other environmental factors such as rain, wind or camera movements. MRF was introduced for segmentation in order to use this spatial information and has the following form:

$$p(\Pi) = Z^{-1} \exp\{U(\Pi)\} \tag{5}$$

$$U(\Pi) = \sum_{c \in C} V_c(\Pi) \tag{6}$$

where, Z is a normalizing constant, and $U(\Pi)$ is the smoothing prior (energy function), V_c is a clique potential function over all cliques $c \in C$. A clique c is a subset of sites in N that are all neighbors of each other, and C is a set of cliques.

For the image segmentation problem, the posterior probability density function given by Bayes rules can be written as:

$$p(\Pi,\Theta \mid X) \propto p(X \mid \Pi,\Theta)\, p(\Pi) \tag{7}$$

Usually in MRF, a multivariate Gaussian distribution is used to model $p(X \mid \Pi,\Theta)$, assuming the relationship between observations and labels follow the Gaussian distribution. Equation (7) can be rewritten as:

$$p(\Pi,\Theta \mid X) \propto \frac{1}{Z'} \exp\left\{ -\sum_{i \in N} \left[\frac{(x_i - \mu_{x_i})^2}{2\sigma_{x_i}^2} \right] + \log(\sigma_{x_i}) - \sum_{c \in C} V_c(\Pi) \right\} \tag{8}$$

With these assumptions, the HMRF-EM algorithm is given bellow:

1. Initial the parameter set $\Theta_j = \{\mu_j, \Sigma_j\}$.
2. Calculate the distribution $p(\Pi_i, \Theta_i \mid X_i)$.
3. Using current parameter set $\Theta^{(t)}$ to estimate the labels by MAP estimation:

$$X^{(t)} = \arg\max_{X \in \chi, i \in N} \left\{ p(X \mid \Pi,\Theta)^{(t)} P(\Pi) \right\} \tag{9}$$

$$= \arg\min_{X \in \chi, i \in N} \left\{ U(X \mid \Pi,\Theta)^{(t)} + U(\Pi) \right\}$$

where, χ is the set of all possible configurations of labels.

4. Calculate the posterior distribution for all $l \in L$ and all pixel x_i using the Bayesian rule:

$$P^{(t)}(l \mid x_i) = \frac{G(x_i;\theta_l) P(l \mid x_N^{(t)})}{\sum_{l \in L} G(x_i;\theta_l) P(l \mid x_N^{(t)})} \tag{10}$$

where, $x_{N_i}^{(t)}$ is the neighborhood configuration of $x_i^{(t)}$ and $G(x_i;\theta_l)$ is a Gaussian distribution function with parameters $\theta_j = \{\mu_j, \Sigma_j\}$.

Here we have

$$P(l \mid x_N^{(t)}) = \frac{1}{Z} \exp\left(-\sum_{j \in N} V_c(l, x_j^{(t)}) \right)$$

5. Use $P^{(t)}(l \mid x_i)$ to update the parameters:

$$\mu_l^{(t+1)} = \frac{\sum_i P^{(t)}(l \mid x_i) x_i}{\sum_i P^{(t)}(l \mid x_i)} \tag{11}$$

$$(\sigma_l^{(t+1)})^2 = \frac{\sum_i P^{(t)}(l \mid x_i)(x_i - \mu_l^{(t+1)})^2}{\sum_i P^{(t)}(l \mid x_i)} \tag{12}$$

2.2 Segmentation Algorithm

In the proposed algorithm, we use a second-order neighborhood system as a clique, which consists of the adjacent eight pixels [14]. The clique potential function is defined as:

$$U(\Pi) = -\beta \sum_{i \in N} \sum_{j \in N_i} \delta(x_i, x_j) \tag{13}$$

where, $N = \{N_i, i \in N\}$, N_i is the set of sites neighboring and $\delta(x_i, x_j) = 1$ if $x_i = x_j$ and $\delta(x_i, x_j) = 0$ otherwise.

We developed an iterative algorithm solve (9)

1. Initialize estimation by maximizing equation.
2. Using the iterated conditional models (ICM) algorithm initialization [15]: initialized by MAP solution based on considering the first-order neighborhood clique potential with Eq. 9.
3. HMRF-MAP segmentation by ICM iteration with equation Eq.9 and using the second-order neighborhood clique potential is adopted.

3 Experimental Results

To confirm any advanatges of our proposed method, our experiements have been set up with the same input and model parameters for our method and other popular methods applied in image segmentation, in particularly K-means, SMM [16], Standard GMM, and SVFMM [13] algorithms. Note that in the experiments, all the methods have been initialized with K-means. The results of our first experiment are presented in Fig. 1. In which Fig. 1a is the original color image having 4-class that we use as the input of segmentation. Figure 1b presents the result obtained by employing Gaussian noise (0 means, 0.001 variance). Figure 1c-1f present the results, in order, obtained by employing K-means, SMM, Standard GMM, and SVFMM algorithm methods. Finally, Figure 1g presents segmented image obtained from our method. The segmented images obtained by employing K-means, SMM, Standard GMM, and SVFMM algorithms (Fig. 1c-f) illustrate a lot remaining noise and a mix of differential clusters of objects on each segmented image. The mix of differential clusters is due to very sentitive of these models to noise. While the segmented image obtained from our method (Fig, 1g) present a better image. in term of high contrasted color and shape. The color of the components in segmented image obtained by our proposed method is smoother and more accuracy. Importantly, the noise is

dramatically reducing during the segmentation process. The PR indexes presented indicate that our method is more efficiency performance than the other four methods discusse.

(a) (b) (c) (d) (e) (f) (g)

Fig. 1. 97x177 testing data, (a) Original image, (b) Standard deviation of Gaussian noise (0 mean, 0.001 variance), (c) K-means (PR=0.63), (d) SMM (PR=0.69), (e) Standard GMM (PR=0.75), (f) SVFMM (PR=0.4), (g) Our method (PR=0.86).

In the second experiment, twelve real world color images (Fig. 2) extracted from from Berkeley dataset were used as the inputs of segmentation. In the second experiment, we apply the same approach of the first enperiement that our proposed segmentation method were compred to other 4 methods as mentioned above. The Probabilistic Rand Index (PR index) [17] has been used as the indication of comparison between our method and other 4 method. PR counts the fraction of pairs of pixels whose labeling are consistent between the computed segmentation and the ground truth, averaging across multiple ground truth segmentations to account for scale variation in human perception. The results of image segmentations obtained from our method are presented in Fig. 3. The 12 segmented images obtained from our method present clearly distinct segment regions/components. The PR results obtained from our method and other 4 methods are presented in Table 1. The PR results show that our proposed algorithm achieves higher PR index than other methods for the same input image, except the comparison between our method and SVFMM method for image 176035. These indicate that our proposed method provides a better segmentation process and result than other competing algorithms.

Fig. 2. Original images of Berkeley image segmentation dataset from left to right (a) 80099, (b) 135069, (c) 24063, (d) 118035, (e) 176035, (f) 55067, (g) 253036, (h) 86016, (i) 41004, (j) 183055, (k) 106020, (l) 108073.

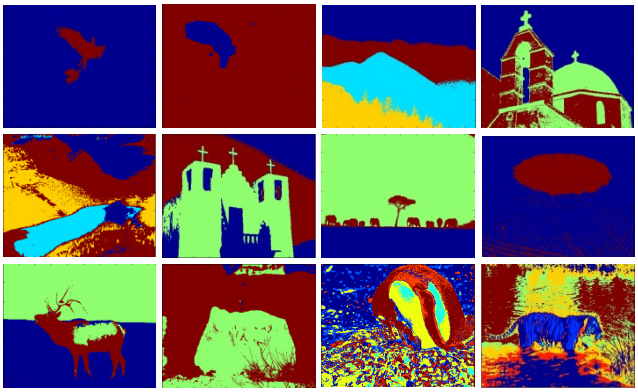


Fig. 3. Segmentation results of images from Berkeley by our proposed algorithm

Table 1. Comparison of performance of proposed method with other methods

PR indexes						
Image	k	K means	SMM	GMM	SVF MM	The proposed algorithm
80099	2	0.87	0.77	0.78	0.81	0.92
135069	2	0.94	0.84	0.82	0.80	0.98
24063	3	0.67	0.68	0.63	0.67	0.80
118035	3	0.69	0.71	0.73	0.51	0.89
176035	4	0.57	0.66	0.52	0.83	0.76
55067	4	0.66	0.63	0.51	0.66	0.97
253036	3	0.59	0.74	0.56	0.76	0.94
86016	2	0.47	0.59	0.66	0.73	0.81
41004	3	0.65	0.5	0.59	0.67	0.83
183055	3	0.69	0.59	0.61	0.56	0.85
106020	6	0.49	0.67	0.57	0.66	0.84
108073	7	0.71	0.66	0.61	0.71	0.83
Mean		**0.66**	**0.67**	**0.64**	**0.70**	**0.87**

4 Conclusion

In this work, a new mixture model has been presented for image segmentation. The model uses simple Gaussian Mixture by using Hidden Markov Random Field that aims to include spatial relationship among neighboring pixels. The model can account for outlier values and thus provides smoother segmentation than the other methods. Also, it has been kept fairly easy to manipulate the parameters of the technique and use the EM algorithm to compute the optimum values for the parameters of the mixture model. This algorithm is not much sensitive to initialize value that is an

ad advantage and good choice for several cases where good initialization is difficult to select. The future work would include identifying better methods to set the parameters of the automatically based on the image to be segmented.

Acknowledgement. This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST)(2013-006535). Also this research was supported by the MSIP(Ministry of Science, ICT&Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (NIPA-2013-H0301-13-3005) supervised by the NIPA(National IT Industry Promotion Agency).The authors would like to thank Dr. Nguyen Thi Hai Yen for her valuable comments and suggestions in this paper.

References

[1] Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(5), 603–619 (2002)
[2] Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(8), 888–905 (2000)
[3] Zhou, H., Schaefer, G., Celebi, M.E., Fei, M.: Baysein image segmentation with mean shift. In: Proc. of IEEE Conference on Image Processng, pp. 2405–2408 (2009)
[4] Li, S.Z.: Markov random field modeling in image analysis. Springer (2009)
[5] Nguyen, T.M., Jonathan Wu, Q.M., Ahuja, S.: An extension of the standard mixture model for image segmentation. IEEE Transaction on Neural Networks 21(8), 1326–1338 (2010)
[6] Titterington, D.M., Smith, A.F.M., Makov, U.E.: Statistical analysis of finite mixture distributions. Wiley, Hoboken (1985)
[7] Celeux, G., Forbes, F., Peyrard, N.: EM procedures using mean field like approximations for Markov model based image segmentation. Pattern Recognition 36(1), 131–144 (2003)
[8] Forbes, F., Peyrard, N.: Hidden Markov random field model selection criteria based on mean field like approximations. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(9), 10891101 (2003)
[9] Nikou, C., Galatsanos, N., Likas, A.: A class adaptive spatially variant mixture model for image segmentation. IEEE Transactions on Image Processing 16(4), 1121–1130 (2007)
[10] Robinson, M., Azimi-Sadjadi, M., Salazar, J.: A temporally adaptive classifier for multispectral magery. IEEE Transactions on Neural Networks 15(1), 159165 (2004)
[11] Li, S.Z.: Markov Random Field Modeling in Image Analysis, 3rd edn. Springer (2009)
[12] Nikou, C., Galatsanos, N., Likas, A.: A class-adaptive spatially variant mixture model for image segmentation. IEEE Trans. Image Process. 16(4), 1121–1130 (2007)
[13] Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions and the Bayesian restorationofimages. IEEE Trans. Pattern Anal. Mach. Intell., PAMI 6(6), 721–741 (1984)
[14] Besag, J.: On the statistical analysis of dirty pictures. J. Roy. Stat. Soc. B 48, 259–302 (1986)

[15] Sfikas, G., Nikou, C., Galatsanos, N.: Robust image segmentation with mixtures of students t-distributions. In: IEEE International Conference on Image Processing, vol. 1, p. 27327 (2007)

[16] Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proc. 8th IEEE Int. Conf. Comput. Vis., Vancouver, BC, Canada, vol. 2, pp. 416–423 (2001)

Performance Analysis of SPDY Protocol in Wired and Mobile Networks

HeeJung Kim[1], GyuSun Yi[1], HanNa Lim[2], JiCheol Lee[2], BeomSik Bae[2], and SungWon Lee[1,*]

[1] Department of Computer Engineering, KyungHee University, Korea
{heejung0310,lysters}@gmail.com, drsungwon@khu.ac.kr
[2] Samsung Electronics, Korea
{hanna.lim,jicheol.lee,bsbae}@samsung.com

Abstract. Google proposed the new application-layer protocol named SPDY with the purpose of complementing problems of HTTP/1.1 to improve web speed. In this paper we evaluate the SPDY protocol's performance in a variety of mobile environment to examine the characteristics of the SPDY protocol and compare the differences between the existing protocol and SPDY protocol. Also through this performance evaluation, we analyze the problem of SPDY and propose directions for improvement of this protocol.

Keywords: SPDY, 3G, WLAN, WiBro, Page load time.

1 Introduction

Explosive developments of mobile device and internet make us connect the internet anytime and anywhere. It also increases the size of the webpage size. However, HTTP/1.1, which we usually use, was presented in 1999 and it decreased web page speed despite high speed of the internet [1].

IETF(Internet Engineering Task Force) announced the development of HTTP/2.0, which is a newer version of HTTP/1.1 was presented after approximately 15 years when HTTP/1.1 first appeared, to adjust quickly changing web page environment and incorporate a SPDY Protocol [2]. SPDY is a new application-level protocol suggested by Google to reduce the web page load time [3,4].

The main purpose of this paper is to evaluate the performance of the SPDY protocol in a variety of mobile environment to examine the characteristics and problems of this protocol. Also, we cover the improvement directions of SPDY protocol.

The remainder of this paper is organized as follows. We briefly describe the SPDY protocol in Chapter 2, and discuss the related work in Chapter 3. We outline our experimental setup in Chapter4, and present our results in Chatper5, and lastly make a conclusion in Chapter 6.

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_26,

2 SPDY Overview

SPDY is suggested by Google to reduce the web page load time. SPDY adds a session layer atop of SSL over the TCP/IP model. Main functions of SPDY are followed.

First, SPDY maintains a single TCP connection between the server and the client. In single connection, it generates an independent stream that carries out multiple HTTP requests. It is unnecessary that transmission order between the stream is maintained. For this reason, each stream has unique stream ID.

Second, the client can request as many items as it wants from the server, and assigns a priority to each request. This prevents the network channel from being congested with non-critical resources when a high priority request is pending.

Third, SPDY reduces HTTP header size by compressing. By using header compressing, SPDY protocol can reduce more than 80% of header size.

Fourth, SPDY provides Server push that the server can push a resource to the client before the client has asked for it.

Above this, It provides features such as Server hint with aim to improve the speed of the web this other [3,4].

3 Related Works

3.1 Analysis of SPDY and TCP Initcwnd

CableLabs, not-for-profit research and development consortium, did performance evaluation about SPDY and TCP tune-up in variety of simulation environment. The result is that using both protocols results in a mean page load time improvement of 29%, compared to previous protocol [5].

3.2 SPDY Performance on Mobile Networks

Google evaluated the performance of SPDY, using a real phone, network emulation tool for emulated 3G network, SPDY-enabled Chrome browser, and a variety of pages from real websites (77 pages across 31 popular domains).

The result is that using SPDY results in an improvement of mean page load time by 23% across these sites, compared to HTTP. This is equivalent to a speedup of 1.3x for SPDY over HTTP [6].

Related works are similar with our research in the way that they measure page load time to evaluate the performance of SPDY. However, a handful of discrepancies in the outcome between the previous works and current study will be observed owing to the difference in the environment. This research was conducted based on the real mobile network.

Table 1. Information of Website in Variety Size

Site	Total Object	Domains	Page Size (KB)
Google Play	52	17	1592.43
2MB	88	4	2299.99
4MB	86	6	3985.02

Table 2. Information of website supporting SPDY Protocol

Site	Total Object	Domains	Page Size (KB)	Site	Total Object	Domains	Page Size (KB)
google plus	7	4	41.84	YouTube	60	13	735.65
gmail	11	4	51.46	Google news	53	6	740.46
Igvita.com	12	6	136.46	Google maps	53	5	836.92
webtide	18	2	177.54	Is the spdy protocol?	31	12	1002.91
Erianna	19	3	217	Google play	52	17	1592.43
google shopping	15	5	264.04	Humble bundle	99	24	1630.22
WordPress	17	11	268.25	Tech Crunch	174	55	1880.33
Twitter	12	3	288.23	Anizon	90	4	2353.93
CloudFlare	32	4	445.68	VIP word press	57	22	3802.77
Lilp Blog	25	4	648.77	gigaom	113	39	5499.83

4 Methodology

Performance evaluation was progressed with Ubuntu 11.04 built-in laptop with two web browsers Chromium and Firefox. We selected the page load time as performance evaluation value. It means the time of downloading whole web page resources.

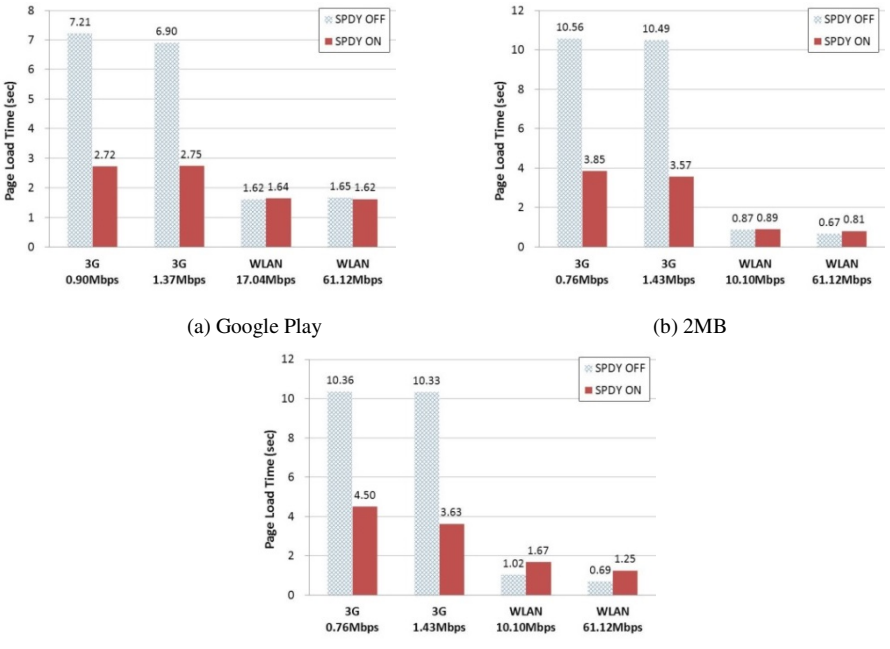
4.1 Evaluation over Web Page in Variety Size

First, we did performance evaluation in variety of web page size, kind and quality of wireless communication. Wireless communication is divided into two cases, either strong or weak signal, and we choose those points in our university. We also select three different sized websites, Google play and one site that supports SPDY protocol. More information of each is shown in Table 1.

To get more accurate results, page load time was measured repeated more than 100 times, and internet visited history was deleted before evaluation of new case was started.

4.2 Evaluation of Initial Connection Time

Next, we measured page load time of each site when we connected first time. For evaluation, we select twenty websites that support SPDY and the list is shown in Table 2. The 10 smallest sites are evaluated in the 3G network with poor signal, and the 10 largest sites are evaluated in the WLAN network with good signal. Each case of evaluation is repeated and when a new measurement is started, we delete the history of the Internet and also restart the laptop.



(a) Google Play (b) 2MB

(c) 4MB

Fig. 1. Page load time of each site in Chromium browser

4.3 Evaluation in WiBro Environment

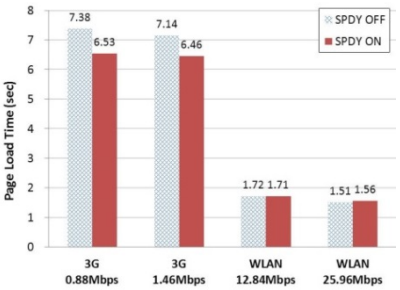
Finally, in order to compare the performance while moving, we measured page load time in WiBro network environment with moving by bus from the university to subway station. We measured SPDY enable case to move to the station from the school and measured SPDY disable case when go back to school. Google play site is selected for this experiment and history of internet is deleted before starting new measurement.

5 Performance Analysis

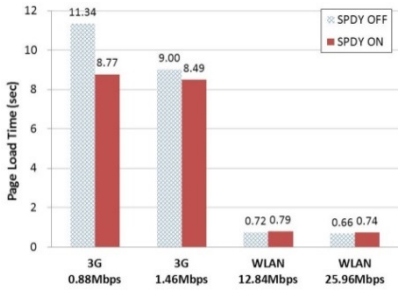
The evaluation is largely divided into three cases to collect more detailed results of SPDY Protocol. Results are as followed.

5.1 Evaluation over Web Page in Variety Size

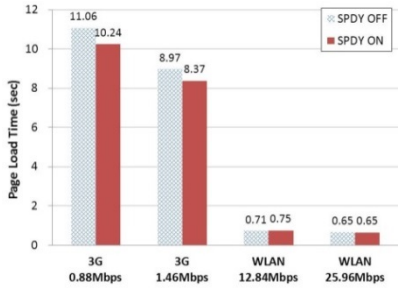
Figure 1 is a graph of measurement results in Chromium browser. (a) is a graph about Google Play site. In a 3G environment, the case of using the SPDY protocol showed a performance improvement about 60% compared to using the existing protocol. However, in high-speed wireless LAN environment, SPDY protocol did not display a



(a) Google Play



(b) 2MB



(c) 4MB

Fig. 2. Page load time of each site in Firefox browser

clear performance improvement, and there were some cases unveiled that using existing protocol is better. (B) is a graph of 2MB size webpage. In a 3G environment, it shows performance improvement more than 60%, but in a WLAN environment, conversely the performance was better when SPDY protocol was unused. (c) is a graph of 4MB size webpage. In a 3G environment, performance improvement of about 60% is appeared when using the SPDY protocol. However, the WLAN environment, the opposite is observed. When using the SPDY protocol, performance decreases about 64% ~ 80%.

Figure 2 is a graph of measurement results in Firefox browser. (A) is a graph of Google play site. In the 3G environment (relatively slow), it showed a performance improvement of about 10%, but in the WLAN environment, performance was decreased. (B) is a graph of 2MB size site. In a 3G environment, it showed a performance improvement of 5% ~ 23% when using the SPDY protocol. In WLAN environment, conversely performance was as bad as 12% when using SPDY protocol. Last, (c) is graph of the 4MB site. It showed a performance improvement of about 7% when using the SPDY protocol performance in a 3G environment. In a wireless LAN environment, however, performance of using SPDY protocol is the same or rather decreased compared to performance of not using SPDY protocol.

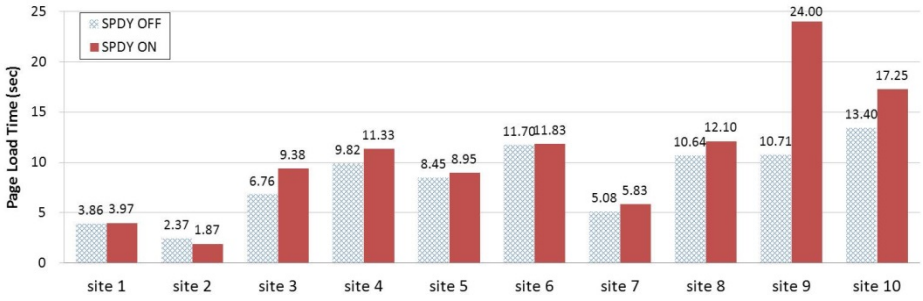


Fig. 3. Initial connection time of 10 sites in 3G network with Chromium browser

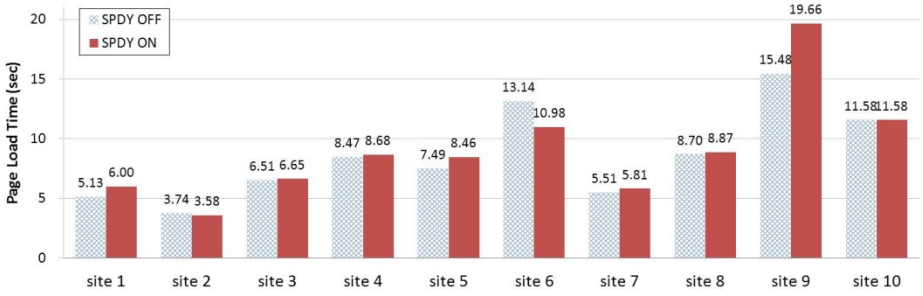


Fig. 4. Initial connection time of 10 sites in 3G network with Firefox browser

5.2 Evaluation of Initial Connection Time

Figure 3 is a result graph of measuring the 10 small size sites in a 3G environment with poor signal using Chromium browser. Besides the one exception of second site, the initial connection time when using SPDY protocol increased by two times SPDY protocol was not used.

Similar situation is showed in Figure 4, which is the result graph of using Firefox browser. It can be seen similar with Figure 3, Excepting the two sites that were measured second and sixth.

Figure 5 is a result graph of measuring the 10 large size site in WLAN environment with poor signal using Chromium browser. With the exception of three sites that are measured second, third and fifth, initial connection time is increased by 49% when using the SPDY protocol.

Figure 6 is a graph of the results in same environment as it can be seen above when using the Firefox browser. In comparison with previous results, the width of difference was not bigger. But with the exception of the three sites that were measured second, third and eighth, the initial connection time when using the SPDY protocol were all increased compared to the case that when SPDY protocol was unused.

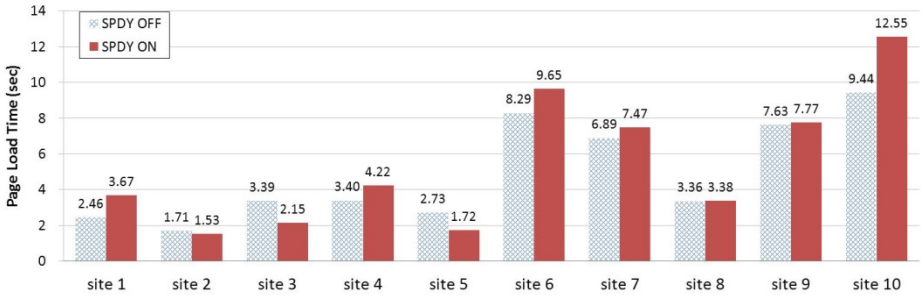


Fig. 5. Initial connection time of 10 sites in WLAN network with Chromium browser

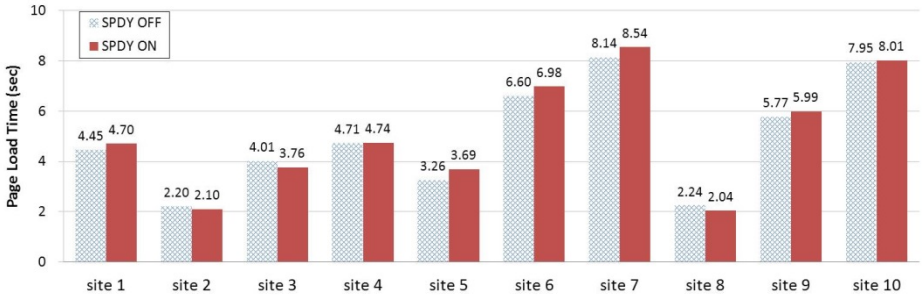


Fig. 6. Initial connection time of 10 sites in WLAN network with Firefox browser

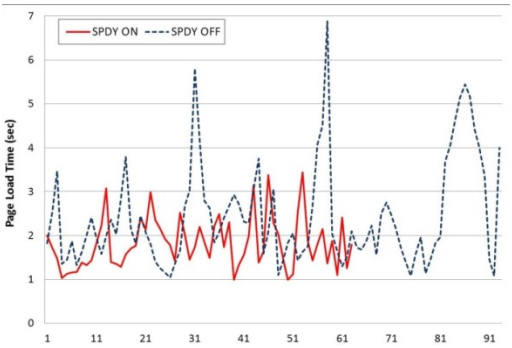


Fig. 7. Evaluation result in WiBro network environment

5.3 Evaluation in WiBro Environment

Finally, it is the result of performance measurement in the WiBro environment. Performance measurement was subjected at a certain time interval, but there was the distance difference on the bus route. Accordingly, we cannot measure with the same number, evaluation was occurred 63 times with enabling SPDY protocol and 93 times with disabling SPDY protocol. Figure 7 shows the results of this measurement.

As compared to the previous two experiments, network environment was not stable in this measurement and the result was also influenced by this environment. In Figure 7, excepting some large value caused by bad connection, it did not show a clear performance improvement as compared with the case of not using SPDY protocol when using SPDY protocol.

6 Conclusions

SPDY maintains a single TCP connection in which multiple streams performing HTTP requests are created. On top of that, with utilization of header compression, both the size of the request and the response header can be downsized by 80%. Accordingly, the performance of web sites is strengthened under SPDY protocol. Observing the result of performance evaluation, at 3G network environment stationary standpoint, we found that SPDY protocol effectively reduced web page load time around 10%-60%, compared with HTTP protocol. This successfully matches with the initial development goal of Google, "Let's make the web faster". However, complicated initial settings led to the lengthened initial connection time. Considering the rapidly shifting patterns of current high speed data communication, advanced measures should be prepared. Moreover, WLAN network environment in which SPDY protocol failed to see the improvements in performance. Last but not least is the SPDY protocol in which improvements in performance was unobserved in terms of mobility. This assigns a task for improvement in SPDY protocol.

This paper focused on a detailed evaluation and analysis of the issues related with SPDY. The expected implication is the creation of the room for developing advanced future SPDY protocol. Expected outcome of this research is we will be able to suggest directions for improvement of SPDY protocol.

Acknowledgement. This project is funded by Samsung Electronics.

References

1. Strangeloop, State of the Union: Ecommerce Page Speed&Web Performance, pp. 7–11 (2012)
2. Belshe, M., Thomson, M., Melnikov, A., Peon, R.: Hypertext Transfer Protocol version 2.0 (2013)
3. SPDY: An experimental protocol for a faster web, `http://www.chromium.org/spdy/spdy-whitepaper`
4. Belshe, M., Peon, R.: SPDY Protocol (2012)
5. White, G., Rice, D.: Analysis of SPDY and TCP Initcwnd. Analysis (2012)
6. SPDY Performance on Mobile Networks, `https://developers.google.com/speed/articles/spdy-for-mmobil?hl=ko`

An Efficient Data Aggregation Scheme for Protecting the Integrity of Sensitive Data in Wireless Sensor Networks

Hyunjo Lee, Tae-Hoon Kim, and Jae-Woo Chang*

Dept. of Computer Engineering, Chonbuk National University,
567 Baekje-daero, deokjin-gu, Jeonju-si, Jeollabuk-do, South Korea
{o2near,taehun3718,jwchang}@jbnu.ac.kr

Abstract. Since wireless sensor networks (WSNs) are resources-constrained, it is very essential to gather data efficiently. For this, data aggregation schemes are studied to minimize transmission cost in terms of the number of data packets. On the other hand, many applications want to preserve data privacy and integrity from the interception (or eavesdropping) of the data by an adversary. However, the existing schemes suffer from high communication cost. To resolve the problems, in this paper, we propose an efficient data aggregation scheme for protecting the integrity of sensitive data in WSNs. Our scheme makes use of the additive property of complex numbers to achieve sensitive data aggregation with protecting data integrity. With simulation results, we show that our scheme is much more efficient in terms of both communication overheads and integrity checking than the existing schemes for protecting integrity and privacy preserving data aggregation in WSNs.

Keywords: wireless sensor network, data aggregation, data integrity, data privacy, signature.

1 Introduction

In WSNs [1], data from sensor nodes are correlated in terms of time and space. For this, data aggregation which combines data coming from many sensor nodes has been actively researched in recent years. An extension of this approach is in-network aggregation which aggregates data progressively as data are passed through the network [4, 8]. In-network data aggregation can reduce the number of data transmissions and the number of nodes involved in gathering data from a WSN. Another issue of WSNs is the confidentiality of transported data. Since sensitive data are transported wirelessly among sensor nodes, they are typically prone to interception and eavesdropping. So, maintaining data privacy of a sensor node even from other trusted participating sensor nodes of the WSN is mandatory [2]. The other issue is data integrity [3, 9]. In communication, data integrity is simply defined as maintaining consistency and correctness of message (message without modification by adversaries). As data aggregation result is used for making critical decisions, the

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_27, © Springer-Verlag Berlin Heidelberg 2014

aggregation result must be verified before accepting it. For this reason, it is required to design a scheme for WSNs which can ensure the aggregated result has not been polluted (manipulation of data by an adversary) on the way to the query server and this issue is one of the main focuses of our research.

Since data privacy and integrity protection processes consume a significant amount of precious resource (i.e., limited power) of sensor nodes they shorten the lifetime of the WSN. Therefore, it is necessary to devise a light-weight scheme which can achieve data privacy and integrity protection efficiently. But, the existing work needs much resource of sensor nodes due to generation of unnecessary messages in the network. To solve this problem, in this paper, we propose an efficient data aggregation scheme that with protecting data integrity in WSNs. Our scheme utilizes complex numbers. The real unit of a complex number is used for concealing sampled data whereas the imaginary unit is used for providing data integrity checking.

The rest of the paper is organized as follows. In Section 2, we present some related work. Section 3 describes our integrity protecting sensitive (private) data aggregation scheme in detail. Simulation results are shown in Section 4. Along with some future research directions, we finally conclude our work in Section 5.

2 Related Work

Recently, He et al. proposed iPDA [3] and iCPDA [9] schemes for WSNs to support integrity checking. In the iPDA scheme, they protect data integrity by designing node disjoint two aggregation trees rooted at the query server where each node belongs to a single aggregation tree. In this technique, first, each sensor node randomly selects a set of L number of sensor nodes, within hop h, from each of the aggregation tree. Every sensor node slices its private data randomly into L pieces and L-1 pieces are encrypted and sent to the randomly selected sensor nodes of the aggregation tree keeping one piece at the same sensor node. In this process, a single sensor node receives multiple slices from multiple sensor nodes. The same process is independently done for each sensor node using another aggregation tree. Then, all the sensor nodes which received data slices from multiple sensor nodes decrypt the slices using their shared keys and sum the received data slices including its own. After that, each sensor node sends the sum value to its parent from the respective aggregation tree. In the same way, the sum data from another set of sensor nodes are transmitted to the query server through another aggregation tree. In the end, the aggregated data from two node-disjoint aggregation trees reach to the base station where the aggregated data from both aggregation trees are compared. If the difference of the aggregated data from the two aggregation trees doesn't deviate from the predefined threshold value the query server accepts the aggregation result, otherwise, it rejects the aggregated result by considering them as polluted data.

In the iCPDA, three rounds of interactions are required: Firstly, each node sends a seed to other cluster members. Next, each node hides its sensory data via the received seeds and sends the hidden sensory data to each cluster member. Then, each node adds its own hidden data to the received hidden data, and sends the calculated results

to its cluster head which calculates the aggregation results via inverse and multiplication of matrix. To enforce data integrity, cluster members check the transmitted aggregated data of the cluster head.

However, during protecting data privacy, both iPDA and iCPDA generate high traffics in the WSN. As a result, communication cost is significantly increased. Moreover, they support very weak data integrity checking because if any node modifies its sampled value 30 to 300 and uses the value 300 for aggregation process none of both methods can detects such misbehavior in the network.

3 An Efficient Data Aggregation Scheme for Protecting the Integrity of Sensitive Data

To overcome previously mentioned shortcomings of the iPDA, in this section, we propose an efficient data aggregation scheme for protecting both data privacy and integrity of the sensitive data in WSNs. Our scheme exploits complex numbers by using their additive property to aggregate sensor data in WSNs. Out of two parts of a complex number, the real part is used to hide the sampled data of a sensor node from its neighboring nodes and adversaries whereas the imaginary part is used for data integrity checking at both data aggregator and the sink node. For transmitting data, every sensor node transforms its sampled data into a complex number form by combining the sampled data with a unique private seed and appending an imaginary unit (a real number adjoined with i) with the modified sampled data. For this, first, the sampled value is mingle with a private seed and then the result appends another real number with i to give the value a complex number form (C = a + bi). The real number with i is the absolute difference value of the previous sample data and the current sample data of a node. Data can be aggregated in upper hierarchy levels during their transmissions to the query server by using algebraic properties of complex numbers. In particular, we apply the additive property of complex number for data aggregation. This is because, like in [2], we also focus on additive aggregation function (Sum). We know that other aggregation functions, such as Average, Count, Variance, Standard Deviation and any other Moment of the measured data, can be reduced to the additive aggregation function Sum [5].

3.1 SUM Data Aggregation with Protecting Data Privacy and Checking Data Integrity

1. Customizing the Sensed Data from the Source Nodes

For processing SUM aggregation function, the sampled sensitive data ds of each sensor node is, first, concealed in a by combining with a unique seed (sr) which is a private real number. The seeds can be selected from an integer range (i.e., space between lower bound - upper bound). By increasing the size of the range, we can further increase the level of the data privacy. To support data integrity, an integer value b – the difference of the previous sensed value and the current sensed value of

the sensor node– with i is appended to the a by using genCpxNum() function to form a complex number C = a + bi as shown in Table 1. Here, for achieving improved security, we use an efficient key establishment scheme proposed in [6]. In this scheme, every sensor node can generate a symmetric key Kx,y by using the information of pairwise key and node IDs. The Kx,y is shared by nodes x and y where the node x encrypts data by Kx,y and the node y decrypts the data by Kx,y. Moreover for encryption and decryption methods, we use the RC5 encryption algorithm [13] which is supported by TinyOS [12]. RC5 is a block-cipher with both a two-word input (plaintext) block and a two-word (ciphertext) output block, which are denoted by A and B, respectively.

Table 1. Customized data creation for each node

SN	Reading ds	Real Seed sr	Mask Value (a=ds+sr)	Difference Value bi	Complex Number (a+bi)
1	16	40	56	2i	56+2i
2	14	51	65	0i	65+0i
3	19	32	51	i	51+i
4	21	23	44	i	44+i
5	17	29	46	3i	46+3i
6	18	33	51	i	51+i
7	13	39	52	2i	52+2i
8	15	67	82	2i	82+2i

For instance, the reading 16 of node 1 is changed into 56+2i. For this, the reading 16 is added to 40 which is a private seed of the node 1 to get a result 56. In addition, assuming that 2i is the difference value of previous reading and current reading of the node 1, the 2i is appended to the result 56 to get 56+2i which is a complex number form of the 16 after data customization process. The node 1 includes its signature, i.e., 00000001, when it transmits the data as: <00000001, 56+2i>. Here, the signature means a special type of positive integer 2^n (where, n= 0 to $B_n \times 8$ -1, such that B_n is the number of free bytes available in the payload) which was presented in our previous work [10]. For the first round, the value of b is zero. We assumed that any sensor node cannot be compromised before sending first round data to the sink node. Every source sensor node keeps the original sensed value d of the current round to deduce b in the next round which is updated in each round of data transmission. Next, the source node encrypts the customized data R'1, i.e., R1 = a + bi, and the signature of the node by using a key Kx,y and transmits the cipher text Cj to its parent. In this way, our algorithm converts the sampled data into an encrypted complex number form. Hence, it not only protects the transmitting trend of private data but also doesn't let neighboring sensor nodes and adversaries to recover sensitive data even though they overheard and decrypted the sensitive data. This is the main principle of our scheme to preserve data privacy in WSNs.

2. Local Integrity Checking at the Intermediate Nodes

In this step, the parent sensor node (i.e., data aggregator) decrypts the received data by using respective pairwise symmetric keys of its child sensor nodes. For each child node, the parent node computes the difference value (b') of the two real units by using the stored previous data and received current data of the child node. For the first round, the value of b' is also zero. For this, the parent node always keeps the record of the previously received data from each of the child nodes and it updates the previous data by current one in every round. To support local integrity checking, the parent node first compares just computed difference value with the currently received difference value (imaginary unit) from the child node and then compares the difference value with local threshold δ. If the imaginary unit of the child's current data is equal to the computed difference value and the imaginary unit is not greater than δ then the parent node accepts the data of the child node. Otherwise, the parent node rejects the data of the child sensor node considering as polluted data. At the same time, it superimposes signatures of the contributed nodes by performing bitwise OR operation on the bit-streams of the node IDs and forwards the encrypted intermediate result 'Cr' towards the sink node.

3. Computing the Aggregation Result with Global Data Integrity Checking at the Sink Node

When the sink node receives all intermediate result sets Crs (partially aggregated encrypted customized data with superimposed signature) from the 1-hop child nodes, it decrypts them by using respective pairwise symmetric keys and computes the final aggregation SUM2 from Crs. After separating SUM2 into a real unit SUM2R and an imaginary unit SUM2IM, the sink computes the actual aggregated result SUM by subtracting SUM1R (a freshly computed sum value of the private seeds of the contributed source nodes) from SUM2R. For example, in Table 1, the total aggregated value collected at the sink node is 447+12i. The sink node separates 447+12i into 447 and 12i. The value 447 is computed by combining the readings of all the nodes with their private seeds. Since the sink node stores the information of the private seeds of all nodes, the sink node can get the actual result, i.e., 133, by subtracting the sum of private seeds of all nodes (314) from the total aggregated value (447). The final result SUM is always accurate and reliable because of the following two reasons. First, a complex number is an algebraic expression and hence the underlying algebra gives the accurate result of the aggregated sensor data. Second, since the private seeds are fixed integer values (i.e., seeds are not random numbers) after collecting data by the sink node it subtracts exactly the same values that have been added to the sensor data during data hiding process by every source node.

At the same time, before accepting the SUM, the sink node performs global integrity checking of SUM to assure whether the SUM2 has been polluted by an adversary in transit or not. For this, like parent nodes, the sink node also computes the difference value (B') of the two real units by using the stored previous data and received current data from the network. The sink node first compares just computed difference value B'i with the currently received difference value i.e., SUM2IM, from

the network and then compares the difference value (SUM2IM) with global threshold Δ (for every application, the maximum value for Δ = δ × N, where N is the total number of nodes in a network). If the imaginary unit SUM2IM of the current data from the network is equal to the just computed difference value B'i and the SUM2IM is not larger than Δ then the sink node accepts the data of the network and returned the actual SUM to the query issuer. Otherwise, the sink node rejects the SUM considering it as forged/polluted data by adversary or other nodes. For example, we assume that a local integrity threshold per node δ equals to 2i and the maximum value for a global threshold Δ can be computed as δ × N = 2i × 8 = 16i. If a sensor node N5 does not participate for data collection, the global integrity checking value Δ can be computed as δ × N = 2i × 7 = 14i. In this scenario, the received data is considered as a consistent one and is accepted by the sink node because the value computed at the sink node, i.e., 9i, is the same as the one received from the network and the value is less than the global integrity checking value, i.e., 9i < 14i.

4 Performance Analysis

In this section, we present simulation results of our scheme by comparing it with iPDA and iCPDA schemes in terms of communication overhead and integrity checking. For this, we use TOSSIM [7] simulator running over TinyOS [12] operating system and GCC compiler. We consider 100 sensor nodes distributed randomly in 100m × 100m area. We set the receiving power dissipation of 395 mW and the transmitting power dissipation of 660 mW which were used in directed diffusion [4].

Figure 1 shows communication overhead in terms of the number of messages generated in a WSN with respect to varying number of sensor nodes. Our scheme outperforms the iPDA and iCPDA schemes because the existing schemes generate unnecessary messages in the network. The reason is that in our scheme each sensor node can customize its data by itself and it doesn't need to generate extra messages in the network for data privacy and integrity checking. On the other hand, the iPDA and iCPDA schemes generate six messages and four messages, respectively, for privacy preservation and integrity checking. As a result, the iPDA and iCPDA schemes are very expensive in terms of communication overhead than our scheme.

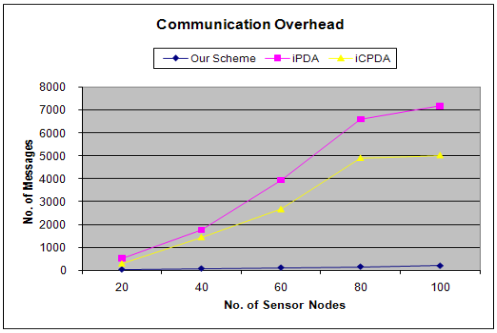


Fig. 1. Number of messages generated by the iPDA, iCPDA and our schemes

When adversaries manipulate messages in the network it is required to detect them. Figure 2 compares integrity checking feature of all the three schemes. It is shown that our scheme can detect every polluted message but the iPDA and iCPDA has very low rate of polluted message detection. The reason is that every node in our scheme performs local integrity checking of the coming data from the lower level nodes. But, only sink node checks the integrity in iPDA and so does the cluster heads in iCPDA.

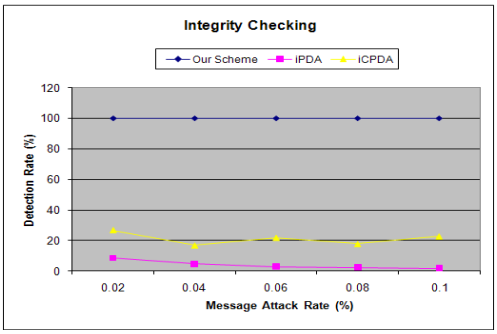


Fig. 2. Integrity checking by iPDA, iCPDA and our schemes when some messages are polluted

Table 2 shows computation overhead of data customization and data aggregation processes. The result shows that the iCPDA has the worst performance to aggregate data by preserving data privacy. The reason is that the iCPDA uses a complex computation method to achieve data privacy. On the other hand, the computation cost of our scheme is about two times faster than that of the iPDA. Our scheme and the iPDA are about 83 and 35 times faster than the iCPDA respectively. It means that both iPDA and our scheme reduce a significant amount of resource (CPU time) usage to achieve private data aggregation.

Table 2. Computational overhead for data customization and aggregation

Protocols	***Execution Time (in Secs.)***
iPDA	0.005924
iCPDA	0.219325
Our Scheme	0.002632

5 Conclusion and Future Work

In this paper, we proposed an efficient data aggregate scheme to protecting data privacy integrity for sensitive data. For maintaining data privacy, our scheme applies the additive property of complex numbers where sampled data are customized and given the form of complex number before transmitting towards the sink node. As a result, it protects the trend of private data of a sensor node from being known by its neighboring nodes including data aggregators in WSNs. Moreover, it is still difficult for an adversary to recover sensitive information even though data are overheard and

decrypted. Meanwhile, data integrity is protected by using the imaginary unit of complex-number-form customized data at the cost of just two extra bytes. Through simulation results, we have shown that our scheme is much more efficient in terms of communication overheads and integrity checking than the iPDA and iCPDA schemes. As future work, we will provide more simulation results by designing data integrity and sensitive data preserving scheme under collusive attacks.

Acknowledgement. This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(grant number 2013010099).

References

1. Yick, J., Mukherjee, B., Ghosal, D.: Wireless sensor network survey. Computer Networks 52(12), 2292–2330 (2008)
2. He, W., Liu, X., Nguyen, H., Nahrstedt, K., Abdelzaher, T.: PDA: Privacy-Preserving Data Aggregation in Wireless Sensor Networks. In: Proceedings of the 26th IEEE International Conference on Computer Communications, pp. 2045–2053 (May 2007)
3. He, W., Liu, X., Nguyen, H., Nahrstedt, K., Abdelzaher, T.: iPDA: An Integrity-Protecting Private Data Aggregation Scheme for Wireless Sensor Networks. In: IEEE MILCOM, pp. 1–7 (November 2008)
4. Itanagonwiwat, C., Govindan, R., Estrin, D.: Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks. In: MobiCom (2002)
5. Castelluccia, C., Mykletun, E., Tsudik, G.: Efficient aggregation of encrypted data in wireless sensor networks. In: The Second Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, pp. 109–117 (2005)
6. Blaß, E.-O., Zitterbart, M.: An efficient key establishment scheme for secure aggregating sensor networks. In: Proceedings of the 2006 ACM Symposium on Information, Computer and Communications Security, pp. 303–310 (March 2006)
7. Levis, P., Lee, N., Welsh, M., Cullar, D.: TOSSIM: Accurate and scalable simulation of entire TinyOS applications, `http://www.cs.berkely.edu/~pal/research/tossim.html`
8. Bista, R., Kim, Y.K., Chang, J.W.: A New Approach for Energy-Balanced Data Aggregation in Wireless Sensor Networks. In: CIT 2009, vol. 2, pp. 9–15 (2009)
9. He, W., Liu, X., Nguyen, H., Nahrstedt, K.: A Cluster-based Protocol To Enforce Integrity and Preserve Privacy in Data Aggregation. In: Proceedings of the 29th IEEE International Conference on Distributed Computing Systems Workshops, pp. 14–19 (2009)
10. Bista, R., Chang, J.W.: Energy Efficient Data Aggregation for Wireless Sensor Networks. Sustainable Wireless Sensor Networks (2010)
11. Zobel, J., Moffat, A., Ramamohanarao, K.: Inverted Files versus Signature File for Text Indexing. In: ACM TDS, vol. 23(4), pp. 453–490 (1998)
12. Hill, J., Szewczyk, R., Woo, A., Hollar, S., Culler, D.E., Pister, K.S.J.: System Architecture Directions for Networked Sensors. In: ASPLOS 2002, pp. 93–104 (2002)
13. Rivest, R.L.: The RC5 Encryption Algorithm. In: Preneel, B. (ed.) FSE 1994. LNCS, vol. 1008, pp. 86–96. Springer, Heidelberg (1995)

Design of a Multi-purpose Real-Time Tracking System for Electric Vehicles*

Junghoon Lee[1], Gyung-Leen Park[1,**], Hye-Jin Kim[1], Min-Jae Kang[2], Ho-Young Kwak[3], Sang Joon Lee[3], Eel-Hwan Kim[4], Jae-Do Song[5], and Hee Suk Kang[6]

[1] Dept. of Computer Science and Statistics
[2] Dept. of Electronic Engineering
[3] Dept. of Computer Engineering
[4] Dept. of Electric Engineering,
Jeju National University
[5] Daekyoung Engineering, Inc.
[6] Infomind, Inc.
Jeju-Do, Republic of Korea
{jhlee,glpark,hjkim82,minjk,kwak,sjlee,ehkim}@jejunu.ac.kr,
cmsong@familydk.co.kr, keystone@infomind.co.kr

Abstract. Aiming at building a city-wide smart transport system taking advantage of computational intelligence, this paper designs and develops a multi-purpose real-time tracking system for electric vehicles and charging infrastructures. Current battery remaining is monitored along with the current location for the EV side while the facility availability is mainly monitored for the charger side according to the standard data model and protocol. Its data model can efficiently cooperate with external information sources such as traffic condition updates. This framework implements a comprehensive user interface and abundant primitive application interfaces for potential electric vehicle services such as reservation, rent management, and scheduling. Upon this multi-purpose service platform, charging facility reservation services, sharing request processing services, and emergency rescue services are developed.

Keywords: electric vehicle, charging infrastructure, multi-purpose tracking system, telematics service, standard data model.

1 Introduction

According to the advent of the smart grid, modern power systems are getting smarter and smarter, taking advantage of sophisticated information and communication technologies, especially in optimization and artificial intelligence [1]. Just like other

* This research was financially supported by the Ministry of Trade, Industry and Energy (MOTIE), Korea Institute for Advancement of Technology (KIAT) through the Inter-ER Cooperation Projects.
** Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_28,

smart grid areas, smart transportation also pursues energy efficiency, while electric vehicles (EVs) are the key element, as they can reduce the consumption of fossil fuels. Here, for the sake of overcoming their drawbacks in short driving range and high price, computational intelligence can provide diverse useful services to EV drivers for their convenience. As those services are essentially belonging to location-based services due to mobility of EVs, it is necessary to track the real-time location of each EV. It can be achieved by vehicle-to-infrastructure communication such as WLAN and 3G [2].

Many EV-related services are expected to appear and will be commercialized, including rent-a-cars, sharing systems, public transportation, and logistics, not restricted to personal use. Hence, it is necessary for the EV telematics service framework to meet the diverse application-specific requirements. The current status of EV objects such as EV themselves, charging stations, and the like, are essential to develop a value-added service. It must be mentioned that battery charging is an essential part of all EV services [3]. We believe that upcoming EV businesses will be built on reservation, rent management, scheduling, and vehicle management. In this regard, this paper designs and develops a multipurpose real-time tracking system for EVs and chargers, integrating new business models on charging reservation, vehicle sharing, and emergency rescue.

2 Tracking System Design

As shown in Figure 1. Our data model consists of information on EVs, chargers, and external sources. For the EV side, the real-time location of each EV creates a spatio-temporal stream. The emergency report is dynamically issued and reported to the tracking server, which will forward it to the appropriate handler such as emergency rescue services according to the predefined logic. For the charging infrastructure side, each charger status is acquired and stored. In addition, our data model can cooperate with external information sources such as ITS, weather updates, disaster alarms, and real-time electricity price signals. Each one is managed in its own agency and retrieved by our database system on necessary basis. Such information blending is a common and essential feature in smart grid cities [4].

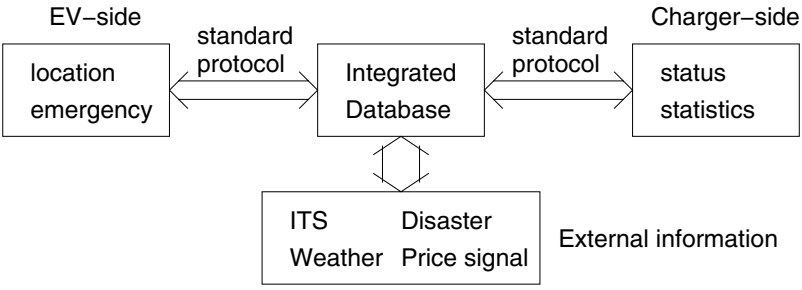


Fig. 1. Data model

To begin with, each EV reports its id, location, speed, SoC (State of Charge), and operation flag to indicate whether it can drive normally. This record is sent to the tracking system via the 3G communication channel under a flat-rate plan. Next, each charger follows the message exchange protocol and the common information model defined in IEC 61968, which is a semantic model for the interaction between different components in power systems [5]. Through this, our tracking system can monitor and control each charger. In addition, for the management purpose, the tracking system can trigger a series of control commands such as configuration file upload and download, firmware download, log file upload, and reset. Now, a value-added service can be implemented to support reservation, rent, and dispatch, taking advantage of add-on agent technologies.

Figure 2 depicts the implementation of charging infrastructure tracking system. It displays the locations of chargers installed in Jeju city on the city map. Jeju island has a coast line of about 200 *km* and a lot of tourist attractions. As a smart grid model city, the number of EVs keeps growing. Here, the type field indicates whether the charger is fast or slow. Each charger is assigned a unique management id. Currently, 4 chargers are directly connected to a prototype version of our tracking system for the validation of communication reliability and information accuracy before the full extension of the tracking system. By this connection, we can know whether a charger is currently connected to an EV, while the charger-issued alarm codes are monitored, stored, and retrieved. Additionally, this interface also shows weather information.

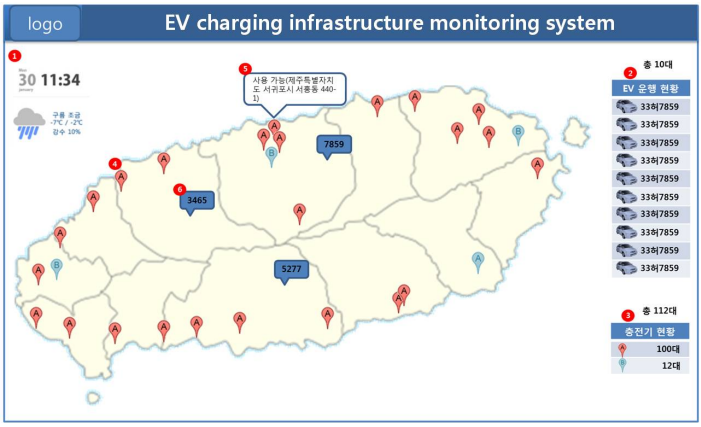


Fig. 2. Charging station tracking system

The tracking system also displays statistics on diverse EV charging activities for each station, as shown in Figure 3. Provided information includes daily amount of supplied electricity, hourly energy consumption, and real-time event logs sorted by the alarm level. Next, it also shows the number of charging transactions as well as the number of faulted chargers. Particularly, a temporal series is plotted in a graph form. Actually, not all chargers obey the interconnection standard and they are compatible only with a few EV batteries. At this stage, this information is very useful to EV

drivers. Moreover, the power supply statistics will be an important guideline for power provisioning and scheduling services on each charging station. A place can be considered as a charging station if it installs more than one charger, for example, shopping malls, airports, office buildings, and the like.

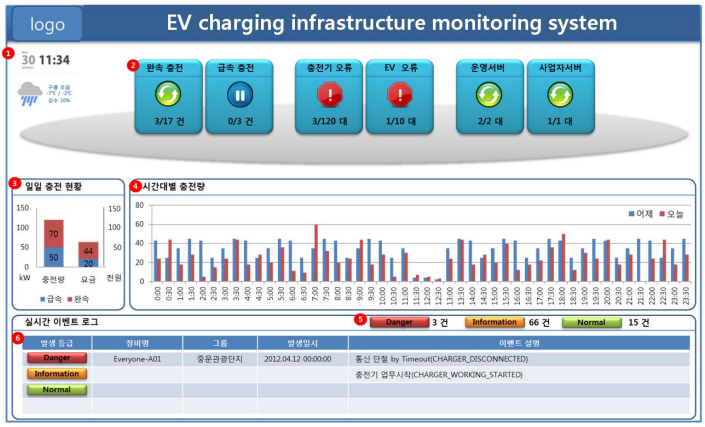


Fig. 3. Charging station statistics

3 Add-On Services

3.1 Charging Reservation Service

As the EV charging process takes too much time, an EV can wait if other vehicles are waiting for the charger. Hence, a reservation mechanism is essential for EV charging, as it can distribute charging load over more than one station, eliminating the waiting time [6]. Moreover, the charging schedule can find the best station where users can park and have their EVs charged while they are doing their work. For example, a tourist can take a tour while his or her EV is being charged. Figure 4 shows the interface design of our reservation system. It provides the charging station buttons by which a user can select and investigate current reservation status of each station. If a user is a shopper or tourist, he or she can decide when and where to go, considering remaining battery amount, stay time, and availability of chargers.

3.2 Emergency Rescue Service

In our emergency rescue service scenario, the PLM (Product Life-cycle Management) database manages all information regarding essential vehicles parts, mainly including batteries for EVs. In addition, it contains manufacture date, parts replacement schedule, service history, repair history, and the like. These records can be retrieved by the emergency rescue service system, ERSS from now on, by way of the central database. Basically, fault detection is reported to the EV tracking system, which continuously monitors the current status of each EV, as shown in Figure 5. The system forwards the report to the ERSS via the interface provided in the form of an

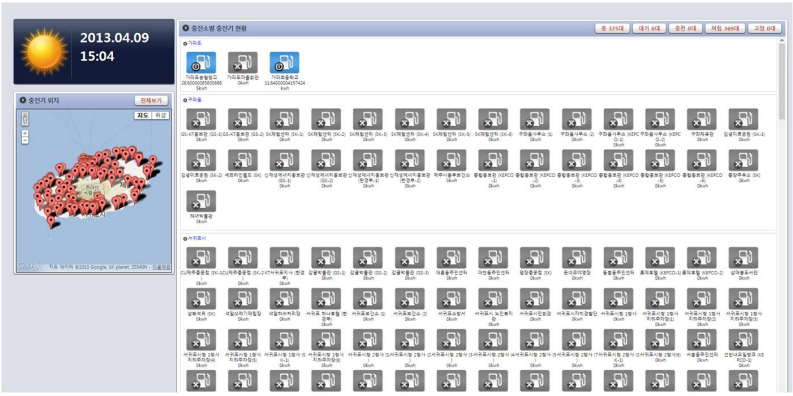


Fig. 4. Per-station monitoring and reservation interface

XML Web service. Considering the information from the PLM database and O&M (Operation and Management) supports, it decides how to cope with the fault and sometimes recommends a car service center capable of replacing necessary parts.

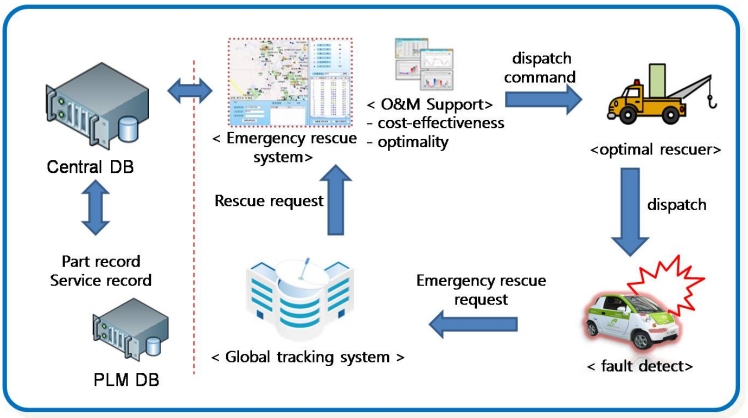


Fig. 5. Integration with emergency rescue systems

In case the EV cannot go to a service center, the emergency vehicle will be dispatched [8]. By continuously monitoring the real-time locations of emergency vehicles over the service area, ERSS can find the best vehicle capable of repairing the reported fault. Receiving the command from ERSS, an emergency vehicle will go to the rescue of the EV. To find the nearest emergency vehicle, ERSS implements the A* algorithm, calculating the point-to-point cost from the report location and each of emergency vehicles. However, if the number of emergency vehicles increases, the one-to-many Dijkstra algorithm can be applied, which sequentially scans adjacent nodes from the report location until an emergency vehicle is found. Anyway, ERSS can estimate the time interval for the selected vehicle to reach the report location and inform the EV of the waiting time.

4 Conclusions

The expectation of many EV-related businesses, makes it necessary to build a common and extensible service framework for diverse services. In this paper, we have designed and developed a multi-purpose tracking system for electric vehicles and charging infrastructures, aiming at providing an open environment for designing, developing, testing, and finally running a new EV services. Current battery remaining is monitored along with the current location for the EV side while the facility availability is mainly monitored for the charger side according to the standard data model and protocol. This framework implements a comprehensive user interface and abundant service interfaces for potential electric vehicle services such as reservation, rent management, and scheduling. Upon this multi-purpose service platform, charging facility reservation services, sharing request processing services, and emergency rescue services are developed.

References

1. Ramchrun, S., Vytelingum, R., Rogers, A., Jennings, N.: Putting the 'Smarts' into the Smart Grid: A Grand Challenge for Artificial Intelligence. Communications of The ACM 55(4), 86–97 (2012)
2. Ferreira, J., Pereira, P., Filipe, P., Afonso, J.: Recommender System for Drivers of Electric Vehicles. In: Proc. International Conference on Electronic Computer Technology, pp. 244–248 (2011)
3. Morrow, K., Karner, D., Francfort, J.: Plug-in Hybrid Electric Vehicle Charging Infrastructure Review. Battelle Energy Alliance (2008)
4. Karnouskos, S., Silva, P., Llic, D.: Energy Services for the Smart Grid City. In: IEEE Int'l Conference on Digital Ecosystem Technologies (2012)
5. McMorran, A.: An Introduction to IEC 61970-301 & 61968-11: The Common Information Model. University of Strathclyde (2007)
6. Lee, J., Park, G., Kim, H., Jeon, H.: Fast Scheduling Policy for Electric Vehicle Charging Stations in Smart Transportation. In: Research in Applied Computation Symposium, pp. 110–112 (2011)
7. Lee, J., Park, G.: Planning of Relocation Staff Operations in Electric Vehicle Sharing Systems. In: Selamat, A., Nguyen, N.T., Haron, H. (eds.) ACIIDS 2013, Part II. LNCS, vol. 7803, pp. 256–265. Springer, Heidelberg (2013)
8. Ibri, S., Nourelfath, M., Drias, H.: On the Integration of Dispatching and Covering for Emergency Vehicles Management System. In: International Conference on Machine and Web Intelligence, pp. 198–204 (2010)

Classification and Analysis of Time Synchronization Protocols for Wireless Sensor Networks in Terms of Power Consumption

Shi-Kyu Bae

Computer Eng. Dept. DongYang Univ., Korea
skbae@dyu.ac.kr

Abstract. Various Time Synchronization protocols (TSP) for a Wireless Sensor Network (WSN) have been developed, because a time relationship plays an important role in many WSN applications, as well. For TSP in WSN, constraints of energy as well as synchronization accuracy should be considered, especially. In this paper, we classify TSPs for WSN in terms of power consumption that is affected by synchronization style. And several TSPs are analyzed according to our classification. Simulation tests have been performed in order to evaluate the power consumption of those selected protocols.

Keywords: Wireless Sensor Network, Time Synchronization, Energy efficiency, Power Consumption, Performance, Analysis.

1 Introduction

Much research in various fields for WSNs has been performed so far. There is Time Synchronization issue among them. Time Synchronization is not a new topic, and a traditional method like NTP (Network Time Protocol) [1] was adopted for wired networks like Internet. Time Synchronization is important for a sensor node in WSN to get information accurately from other nodes in respect of a common clock criterion. Time critical applications anticipate an accurate or efficient synchronization among sensor nodes in WSN. There is a growing interest in the literature on proposals for time synchronization in WSNs. So, various time synchronization schemes have been developed until now. The importance and the way of analysis of the performance parameters are different depending on each synchronization scheme. So, it is not easy to evaluate and compare existing schemes. The main challenge for designing WSN is to minimize energy consumption. The time synchronization protocols for WSN have no exception, as well. Thus, synchronization needs to be energy efficient to prolong the lifetime of networks. The energy load should be evenly distributed among all nodes so that a certain node will not die too quickly.

In this paper, we classify TSPs for WSN and analyze the representative TSPs in terms of synchronization style that affects power consumption. The remainder of this paper is organized as follows. Section 2 presents works related to TSPs in WSN.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_29,

In section 3, energy model for WSN will be described. TSPs for WSN are classified in terms of synchronization style that affects power consumption, and several TSPs that are classified into categories above were analyzed in section 4. This paper ends with some concluding remarks in Section 5.

2 Related Works

TSPs were classified in several criteria [2][3].

- Master-slave versus peer-to-peer synchronization

While a master-slave protocol assigns one node as the master and the other nodes as slaves, any node can communicate directly with every other node in the network in a peer-to-peer type.

- Clock correction versus un-tethered clocks

Clock correction clock types perform synchronization by correcting the local clock in each node, but un-tethered ones maintain synchronization with time information table, without correcting clock.

- Internal synchronization versus external synchronization

In external synchronization, a standard source of time such as Universal Time (UTC) is provided. However, a global time base or real-time is not available from within the system in a internal synchronization type, which has a goal to minimize the difference between nodes.

- Probabilistic versus deterministic synchronization

The former provides a probabilistic guarantee on the maximum clock offset with a failure probability that can be bounded or determined. But, the latter guarantees an upper bound on the clock offset with certainty.

- Sender-to-receiver versus receiver-to-receiver synchronization

Receiver-to-receiver lets a reference node transmit once and other nodes exchange messages independently of the reference node. Reference Broadcast Synchronization (RBS) [4], which is in this category, lets a sender send a broadcast beacon for receivers' reference, and receivers except the sender participate in synchronization. RBS increases the accuracy by eliminating sender side's delay uncertainty even though it can't transmit exact reference time (or global clock). *Sender-to-reciever* indicates that one node send a message while the other receive it. Timing-sync Protocol for Sensor Networks (TPSN) [5] and Flooding Time Synchronization Protocol (FTSP)[6] are classified into this synchronization category which can transmit global reference clock through the network. It is a contrast to RBS. Receiver Only approach was added later than previous two methods in [3] and Pair-wise Broadcast Synchronization (PBS) [7], where a group of nodes are synchronized by listening to the message exchanges of a pair of nodes.

Besides, in [2],[3], the characteristics of the representative clock synchronization protocols for WSNs was reviewed and analyzed. We newly classify TSPs in WSN with other criteria and analyze the representative ones in this paper.

3 Power Evaluation Model

As power consumption for communication is known to be far larger than that for computing in cases of most WSN applications, power consumption of a sensor node is dominated by communication, rather than computing [8]. In the experiment with an example application (i.e. DSR routing protocol) [9], a transceiver module at each node had the larger energy consumption than processor or sensor module. At a transceiver module of each node, energy consumption in the transmit state is larger than that in the receiver state. In other words, for the same length of packet more energy is required for sending a packet than receiving one. However, in the same simulation mentioned above [9], the transceiver module showed larger energy consumption in the receiver state than in the transmit state. It is because that all packets within the communication area were received by a node, which is called as *Overhearing*. Overhearing occurs when a node receives packets that are not destined to it. Due to broadcast characteristics of wireless communication, all nodes overhear the packets transmitted from transmitter nodes. So, generally speaking, the number of packets received is much greater than that of packets transmitted. Overhearing, which is determined by some parameters like node density, spends significant amount of energy.

Assume that all sensor nodes are homogeneous and there are two basic energy consumption types, i.e. transmitting(E_{Tx}) and receiving(E_{Rx}). Static WSN that all sensor nodes remain fixed in their position once they have been deployed is also assumed. We use a simple energy dissipation model described in [10]. To transmit a b-bit message to a d-distant node, a transmitter node consumes

$$\begin{aligned} E_{Tx}(b,d) &= E_{Tx\text{-}elec}(b) + E_{Tx\text{-}amp}(b,d) \\ &= bE_{elec} + be_{fs}d^2,\ d< d_0 \end{aligned} \quad (1)$$

E_{elec}, the electronic energy, is determined by factors such as the digital coding, filtering, modulation, and spreading of the signal. And $e_{fs}d^2$, the amplifier energy, is determined by the distance from the transmitter to the receiver and the bit-error rate in free space (when the distance is less than a threshold distance, d_0).

To receive a b-bit message from a node which has the distance less than the threshold distance ($d_{0)}$, a receiver node consumes

$$E_{Rx}(b) = E_{Rx\text{-}elec}(b) = bE_{elec} \quad (2)$$

When a node sends a packet, the node not only consumes power as much as transmitting power, but also makes other neighbor nodes consume power as much as receiving power. In other words, due to broadcast nature of wireless communication, all the neighbor nodes which surround a transmitter node overhear the transmitted packets and consume the receiving energy.

Energy consumed at all nodes within a transmission range, R in general WSNs, when a b-bit message is sent by the transmitter node is

$$\begin{aligned} E_R(b,R) &= E_{Tx}(b,R) + (n(t)-1)\, E_{Rx}(b) \\ &= bE_{elec} + be_{fs}R^2 + (n(t)-1)*bE_{elec} \\ &= n(t)*\, bE_{elec} + be_{fs}R^2 \end{aligned} \quad (3)$$

Where E_R is the total energy consumed within a transmission range which is caused by the transmitter node, and n is the number of nodes within a transmission range including the transmitter which is centered. n(t) is the function of time depending on node density (the number of nodes per area) and mobility. As static WSNs without mobility are assumed, n depends on only node density and becomes constant.

E_R can be represented from (Eq. 3) as following

$$E_R(b,R) = n*bE_{elec} + be_{fs}R^2 \;\; = b * (n*E_{elec} + e_{fs*}R^2) \quad (4)$$

where n is the number of nodes within a transmission range including the transmitter node, depending on the node density deployed in a sensor field in static WSNs.

Eq. 4 shows that E_R is proportional to the size of messages for a static and uniformly deployed WSN. If a transmission range R is determined and fixed at the transmitter in a uniformly deployed static WSN, E_R is only proportional to the size of messages. Therefore, total energy consumption of a network for WSN protocols can be evaluated in terms of the total size of the transmission messages which were generated in the entire network.

$$\begin{aligned} E_{network}(b) &= \sum_{k=0}^{M} \quad E_R(b) \\ &= (n*E_{elec} + e_{fs} * R^2) * \sum_{k=0}^{M} \quad b \end{aligned} \quad (5)$$

where M represents the total number of messages occurred at all nodes in the network.

4 Classification and Analysis of TSPs

A network-wide synchronization operation is composed of more than a basic operation, which is performed among a set of nodes. The set of nodes participate in a basic synchronization operation which requires more than a packet transmission in either unicast or broadcast way, and the number of packet to be transmitted is different according to different synchronization schemes. Re-synchronization throughout a network will be repeated periodically or triggered by certain conditions (which is called as a round in this paper). Some synchronization schemes require a series of the synchronization step, basically. So, we define a *round* as one network-wide synchronization operation in this case for comparison with other schemes.

In previous section, it is shown that total energy consumption of a network for WSN protocols can be evaluated in terms of the total size of the transmission

messages which were generated in the entire network in Eq. 5. We classify the synchronization schemes into two categories by the style of synchronized progress in the light of the number of transmission messages; pair-wise synchronization (unicast-dominated transmission) and group-wise synchronization (broadcast-dominated transmission).

4.1 Pair-Wise Synchronization

A basic synchronization operation exchanges more than one message node by node sequentially or in parallel. A tree structure is used typically.

TPSN - Tree Topology : TPSN, the most prominent TSP, synchronizes each pair of nodes from a root node level by level with its own tree structure. Fig. 1 shows the operation of TPSN that is in pair-wise synchronization category.

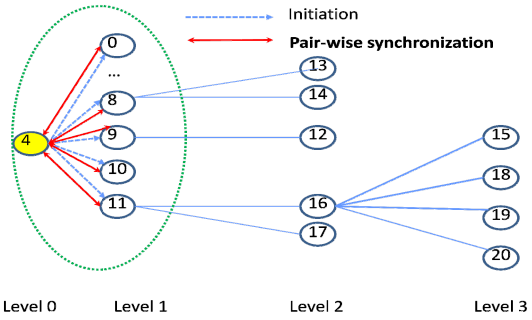


Fig. 1. Operation of TPSN : an example of pair-wise synchronization category

TPSN[5] is broken up into two phases; level discovery and synchronization phase. Level discovery phase creates the hierarchical topology of the network. In synchronization phase, all nodes synchronize with the lower level nodes with two-way communications, which synchronizes all nodes with the root node.

Suppose n be the number of nodes of an entire network. As the hierarchical tree of TPSN is constructed by simple flooding method, transmissions as many as n times are required. If h be the number of parent nodes at the tree, h ranges from 1 to (n-1). One initiation message per parent node and one round-trip message exchange per child (or link) are required in synchronization phase of TPSN. Thus, the total number of message transmissions per round can be calculated as $M_{TPSN} = h + 2(n-1) = 2n + h - 2$.

RBS - Mesh Topology. RBS[4] uses one broadcast and exchanges messages among all nodes except the sender node (mesh topology). So, n(n-1) times message transmissions are needed at maximum.

4.2 Group-Wise Synchronization

In this type, all or part of the neighbor nodes within a communication range are synchronized. And the entire network is synchronized consecutively or in parallel

with a unit of communication range (or broadcast domain). That is to say, the synchronized area having a communication range size may overlap. This category has typically a form of hierarchical cluster. Even though clustering needs additional overhead to construct clusters, cluster-based schemes have advantages with regards to performance, if re-clustering is not used too frequently. Fig. 2 shows the operation of the group-wise synchronization, which operates in a cluster-by-cluster way.

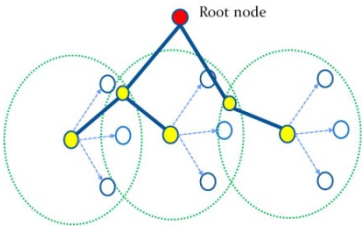


Fig. 2. An operation example of group-wise synchronization

The number of transmission messages per cluster is determined by a specific synchronization protocol.

FTSP. FTSP[6] synchronize in a diffusion way for every node. So, FTSP can be considered to be an implicit hierarchy structure and overlapped clusters. FTSP operates node-by-node within a communication range from a center node. The number of messages per round is equal to the number of entire nodes in a network. In FTSP, the order of synchronization operations is determined by the position of nodes (or distance from the previous node), and is different from TPSN which is determined by the tree constructed previously.

Reference Based, Tree Structured Time Synchronization [11]. A center node (called a reference node) synchronize all neighbor nodes within a communication region using broadcasting after exchanging two messages between the reference node and one of the neighbor nodes. Thus, three messages (two unicast and one broadcast) are required per cluster, so that the total number of messages per round in a network is three time the number of clusters.

Adaptive Time Synchronization for Homogeneous WSNs [12] . An approach of [12], which has cluster hierarchy, uses pair-wise communication between cluster heads including root node. After that, all cluster head nodes synchronize its cluster members by broadcasting delay information calculated from the message exchange between the cluster head and a specific node (one of the cluster members).

A classification of TSPs in terms of the number of transmission messages is shown in Fig. 3.

4.3 Discussion

RBS, of which the number of messages per round is $O(n^2)$, requires the most message exchanges. So, the most power will be consumed than any other TSPs. Tree-based pair-wise synchronization has the message complexity of O(n). However, cluster-based group-wise synchronization requires the message complexity of $O(h)$, where h is the number of cluster heads in a network. Therefore, cluster-based group-wise synchronization has an advantageous over pair-wise types in terms of power consumption, especially in dense WSN, even though it does not guarantee the reliability of transmission due to broadcast.

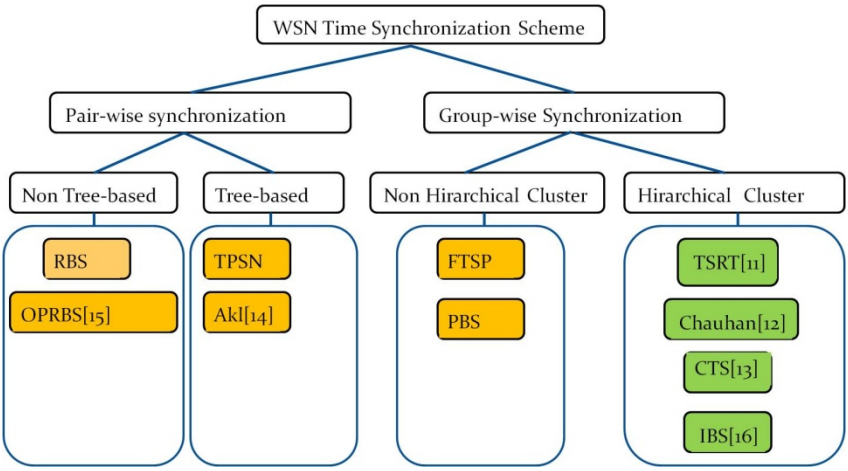


Fig. 3. classification of TSPs in terms of the number of transmission messages

5 Conclusion

Many time synchronization schemes for WSN have been developed until now. They, however, are designed and implemented for a different goal and operate in different ways. For TSP in WSN, constraints of energy as well as synchronization accuracy should be considered, especially. In this paper, we classify TSPs for WSN in terms of power consumption that is affected by synchronization style. And several TSPs are analyzed according to our classification. Group-wise with hierarchical cluster TSPs show the better performance of power consumption than pair-wise in our analysis. This paper will be useful in order to evaluate the power consumption of TSP in WSN.

References

1. Mill, D.: Computer Network Time Synchronization: the Network Time Protocol on Earth and in Space, 2nd edn. CRC Press (2011)
2. Sundararaman, B., et al.: Clock synchronization for wireless sensor networks: a survey. ELSEVIER Ad-hoc Networks 3, 281–323 (2005)
3. Rhee, I., et al.: Clock Synchronization in Wireless Sensor Networks_ An Overview: Sensors, 9 (2009), http://www.mdpi.com/journal/sensors
4. Elson, J., et al.: Fine-grained network time synchronization using reference broadcasts. In: ACM OSDI (2002)
5. Ganeriwal, S., et al.: Timing-Synch Protocol for Sensor Networks. In: ACM Sensys, USA (2003)
6. Maróti, M., et al.: The Flooding Time Synchronization Protocol. In: ACM SenSys 2004 (2004)
7. Noh, K., et al.: A New Approach for Time Synchronization in Wireless Sensor Networks: Pairwise Broadcast Synchronization. IEEE Tran. on Wireless Comm. 7(9) (September 2008)
8. Mahgoub, I., et al.: Sensor Network Protocols. CRC Press (2006)
9. Zhou, H., et al.: Modeling of Node Energy Consumption for Wireless Sensor Networks. Wireless Sensor Networks 3 (2011)
10. Heinzelman, W., et al.: An Application-Specific Protocol Architecture for Wireless Microsensor Networks, vol. 1(4) (October 2002)
11. Rahamatkar, S., Agarwal, A.: An Approach towards Lightweight, Reference Based, Tree Structured Time Synchronization in WSN. In: Meghanathan, N., Kaushik, B.K., Nagamalai, D., et al. (eds.) CCSIT 2011, Part I. CCIS, vol. 131, pp. 189–198. Springer, Heidelberg (2011)
12. Chauhan, S., et al.: Adaptive Time Synchronization for Homogeneous WSNs. International Journal of Radio Frequency Identification and Wireless Sensor Networks (2011)
13. Zurani, A., et al.: Clustered Time Synchronization Algorithm for Wireless Sensor Networks. International Journal of Recent Technology and Engineering (IJRTE) 1 (June 2012)
14. Akl, R., et al.: Hybrid Energy-Aware Synchronization Algorithm in Wireless Sensor Networks. In: The 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2007 (2007)
15. Jain, S., Sharma, Y.: Optimal Performance Reference Broadcast Synchronization (OPRBS) for Time Synchronization in Wireless Sensor Networks. In: International Conference on Computer, Communication and Electrical Technology, ICCCET (2011)
16. Bae, S.: Time Synchronization by Indirect-Broadcasting for Wireless Sensor Networks. In: Park, J.J(J.H.), Jeong, Y.-S., Park, S.O., Chen, H.-C. (eds.) Embedded and Multimedia Computing Technology and Service. LNEE, vol. 181, pp. 659–666. Springer, Heidelberg (2012)

The Vehicle Speed Detection Based on Three-Dimensional Information Using Two Cameras

Inwon Lee, Jong-pil Ahn, Eun-Ju Lee, Chi-Hak Lee, and Young-mo Kim

School of Electronics Engineering, Kyungpook National University,
Sankyuk-dong, Buk-gu, Daegu, Korea
ymkim@ee.knu.ac.kr

Abstract. This paper proposes a new vehicle speed detection method which is based on three-dimensional information of vehicle with two cameras. Two Cameras: one is a view of telescope, and the other is a view of wide. At first, the proposed method searches vehicle features such as edge, orientation or plate number, generates three-dimensional information from found features, and identifies an interesting features such as license plate. Finally, it is expected to get the height of the features and the speed of vehicle. Conclusion is that the proposed method is sufficient to be applied to the equipment which detects vehicle speed automatically.

1 Introduction

In intelligent traffic system, a speed detection system is an important system at the speed vehicle speed detection. In General, the over-speed detection system consists of one camera and two loop detectors which are installed in front of one of the cameras. The speed of the vehicle through the loop detector is detected, and if the speed of detection is determined the over-speed, the camera should be taken. In this way, to install the system, it is a road block, and the loop detector is buried under the road. And the camera should be installed with loop detector. So, enforcement section is limited, and the system is difficult to move, repair and so on.

Recently, there have been several works on the tracking of vehicles and classifying them into different types. Others, along with tracking, have estimated important traffic parameters by using image-processing techniques.

The several methods are the vehicle plate recognition[1,2], the vehicle model recognition[3], the vehicle velocity detection[4], and the tracking such as KLT[5], extended Kalman Filter[6], particle filter[7]. The basic method of vehicle tracking:

first, setting the region of interest,
second, recognizing vehicle entering the region of interest,
finally, performing tracking algorithm at vehicle entering the time.

In this paper, we propose a new vehicle speed detection method based on three-dimension information with two cameras. Two cameras estimate the height of the

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_30,

feature of vehicle. And the image of a wide angle camera estimates speed of vehicle using the transfer course of the feature. The feature is used to be license plate.

In section 2, the proposed speed detection method using two cameras is described, in which we describe a structure of it, detecting license plate for the feature, estimating height of the plate, and estimating speed of vehicle. We describe the experiments of the proposed method in section 3.

2 The Proposed Method

The visual scene of a real world is 3D, but image processing is performed on the 2D image extracted from 2D image sensor. To perform 2D image processing of the real 3D visual scene, 3D scene is projected onto 2D image plane. Figure 1 shows an example comparing distance of real world to that of image plane. Therefore, it is difficult to estimate speed of vehicle without the height of license plate.

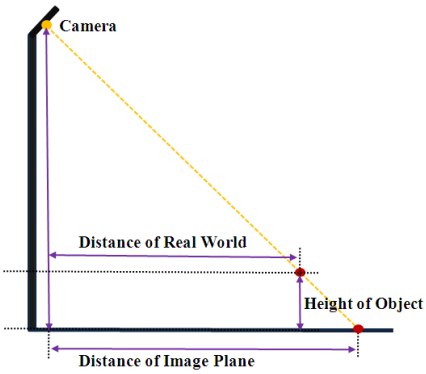


Fig. 1. Difference distance between real world and image plane

In this section, we explain that two cameras estimate the height of the feature of vehicle. And the image of a wide angle camera estimates speed of vehicle using the transfer course of the feature.

2.1 Structure of System

The proposed system using two cameras is installed on lane about six meters. And the proposed method estimates feature height using two cameras at difference viewing angle, and the speed of vehicle based on the feature on the wide angle camera. Figure 2 shows the system using two cameras on the top. One camera shows small region of interest for the 3D information of vehicle. Other camera shows large region of interest for the speed detection of vehicle. As shown in figure 3, lower camera is the narrow viewing angle, and upper camera is the wide viewing angle on the side.

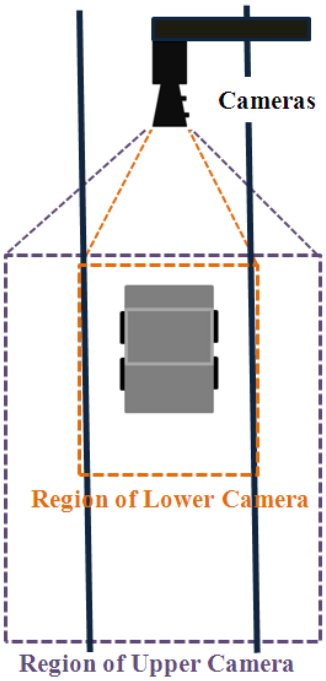


Fig. 2. Region of Cameras (Top View)

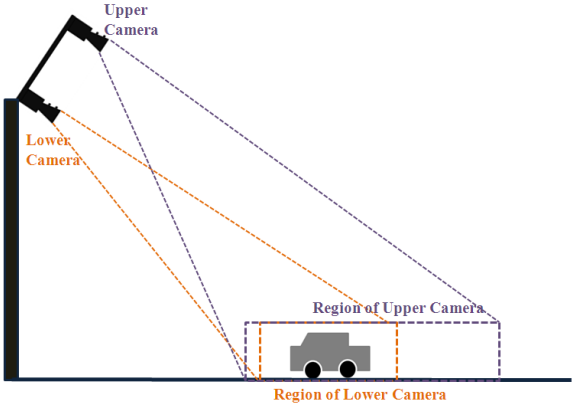


Fig. 3. Region(s) of (two) cameras (Side View)

2.2 Feature Extraction

The feature extraction method is recognition of license plate using input image.

The method:

first, extraction of license plate region using morphology and labeling algorithm,
second, correction of inclined license plate,
finally, recognition of license plate[1].

Figure 3 shows detection regions of two cameras. And the system is installed askew. The reason is that the rear vehicle can be hidden by the front vehicle.

2.3 Estimated Height of Feature

The estimated height of feature is derived with two cameras. If upper camera image and lower camera image find same feature, the height of feature can be obtained. Figure 4 shows the parameters needed to obtain the height of feature.

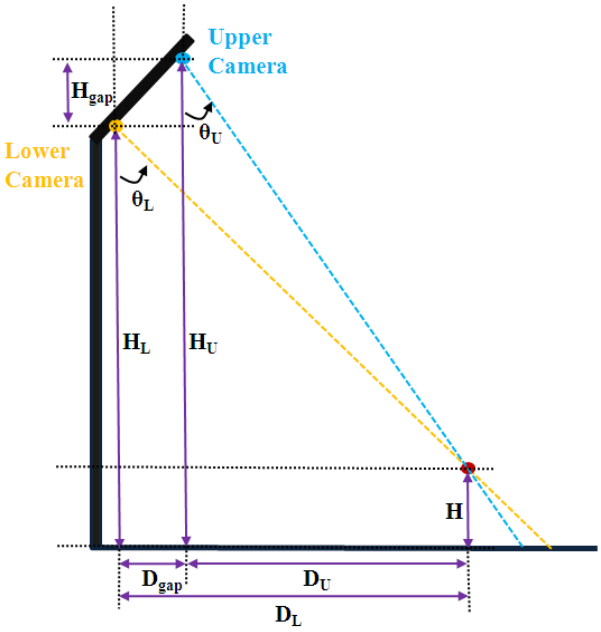


Fig. 4. Processing of the height

The function which derives feature height is defined in Eq. (1).

$$H = \frac{H_U \mathrm{X} \tan(\theta_U) - H_D \mathrm{X} \tan(\theta_D) - D_{gap}}{\tan(\theta_U) - \tan(\theta_D)} \tag{1}$$

The H means height of feature. The H_u and b_{ut} mean height of upper camera and angle between upper camera and feature, respectively. The H_D and θ_D mean height of lower camera and angle between lower camera and feature, respectively. The D_U, D_L and D_{gap} mean distance between upper camera and feature, distance between lower camera and feature, and distance between upper camera and lower camera, respectively.

2.4 Estimated Speed

The visual scene of a real world is 3D, but image processing of this information is performed on the 2D image extracted from 2D image sensor. So it is not possible to estimate speed of vehicle with a camera. If the height of feature is known, it is possible to estimate exact speed of vehicle. Figure 5 shows difference between distance on height of license plate and that on the road. The proposed method uses the height of license plate.

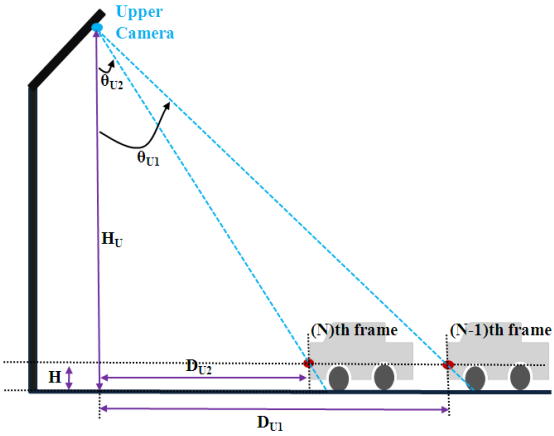


Fig. 5. Difference between distance on height of license plate and that on the road

2.5 Speed Detection of Vehicle Using Two Cameras

Figure 6 shows the block diagram of the proposed speed detection method using two cameras. The recognition model of license plate is extracted to the same feature for image of upper camera and lower camera in a sampling time T+dt. The height of license plate is obtained by using each image information in a sampling time T+dt. The proposed method detects the speed of vehicle using the height of license plate, and image of upper camera in sampling time T and T+dt.

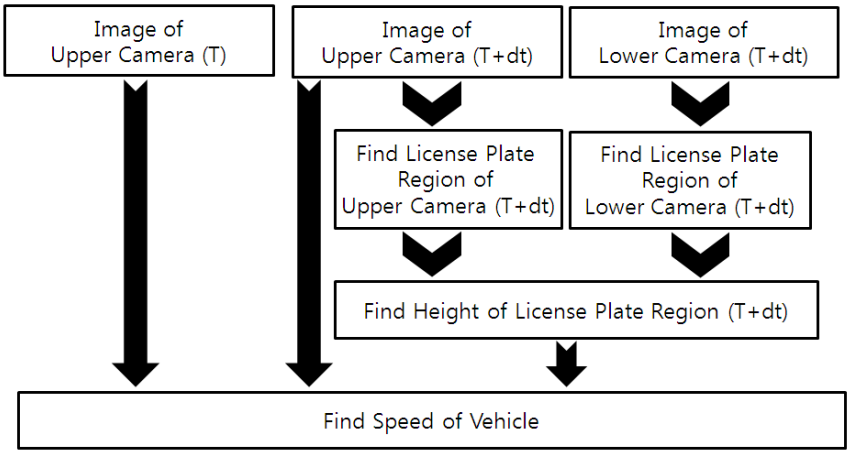


Fig. 6. Block diagram of the proposed model

3 Result

We demonstrated the crossroad at Ilsan district called "Lake Park" in Korea. Normally 100 vehicles one hour get through this crossroad during the green light signal on. For experiments, we prepare 'Intel Core i5 760 @2.80GHZ', '3.00GB RAM', and C++. And the gap of two cameras is 0.5 m.

Figure 7 shows the box rounding found license plate region on the image acquired from camera. The license number and region of license plate can be extracted from these frames using morphology operator and labeling algorithm. Figure 8 left shows the license plate region on the image from lower camera, and figure 8 right shows license plate region on the image from upper camera. After two cameras read the license number, the system finds the same license plate of the same vehicle and

Fig. 7. License plate region on the image from camera

extracts the features on the frames from the stereo camera. As shown in figure 9, the vehicle proposed method detects the speed between frames using transferred license plate and the time between frames. We can see how many speed detection of vehicle on Table 1.

Fig. 8. License plate region on the image from lower camera(lefr), license plate region on the image from upper camera(right), two cars are the same

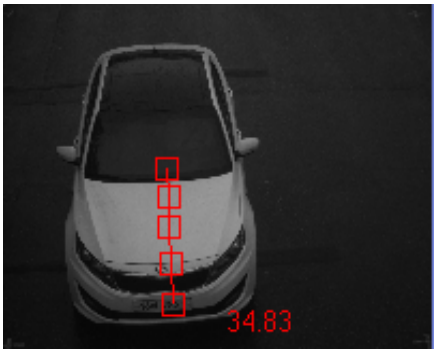


Fig. 9. Speed detection from upper camera

Table 1. The result of vehicle speed detection

Speed of Vehicle	No. of Vehicle	Avg. of Error Rate(%)	Min. of Error Rate(%)	Max of Error Rate(%)
50km/h ~ 55km/h	231	2.27695	-5.4915	7.120855
55km/h ~ 60km/h	256	1.09688	-3.9623	5.955455
60km/h ~ 70km/h	388	0.342392	-4.83644	4.525765
70km/h ~	110	-0.77542	-6.45455	2.349078

4 Conclusion

985 vehicles are detected for 10 hours a day at the crossroad. Approximately 100 vehicles are detected during a hour. The proposed method can play an important role for speed detection in intelligent traffic system.

Acknowledgments. This work was supported by National Research Foundation of Korea Grant funded by the Korean Government 2010-0025545.

References

1. Ko, M., Kim, Y.: Improvements in Real-Time Feature-Based License Plate Character Recognition. Key Engineering Materials 277-279, 355–360 (2005)
2. Mousa, A.: Canny Edge-Detection Based Vehicle Plate Recognition. International Journal of Signal Processing, Image Processing and Pattern Recognition 5(3), 1–8 (2012)
3. Lee, S., Gwak, J., Jeon, M.: Vehicle Model Recognition in Video. International Journal of Signal Processing, Image Processing and Pattern Recognition 6(2), 175 (2013)
4. Kalyan, A.S., Divakar, T., Rao, K.N., Chakravarthi, A.M.: Vehicle Velocity Prediction & Estimation in 2d Video for Night Condition. International Journal of Signal Processing, Image Processing and Pattern Recognition 4(2), 19–30 (2011)
5. Lim, Y., Lee, C., Kwon, S., Lee, J.: A fusion method of data association and virtual detection for minimizing track loss and false track. In: 2010 IEEE Intelligent Vehicles Symposium (IV), San Diego, United States, June 21-24 (2010)
6. Barth, A., Franke, U.: Estimating the Driving State of Oncoming Vehicles From a Moving Platform Using Stereo Vision. IEEE Transactions on Intelligent Transportation Systems 10(4), 560–570 (2009)
7. Yang, T., Kang, S.: Tracking for moving object using invariant moment and particle filter. In: Control Conference 2008, Kunming, Chinese, July 16-18 (2008)

Enhancing Spatial Information for Relief Work during Nuclear Accidents

Ming-Kuan Tsai* and Nie-Jia Yau

Research Center for Hazard Mitigation and Prevention, National Central University,
No.300, Jhongda Rd., Jhongli City, Taoyuan County 32001, Taiwan
twmktsai@ms95.url.com.tw, yau@ncu.edu.tw

Abstract. Several severe nuclear accidents have occurred in recent decades, necessitating immediate action to repair damage or prevent further harm. Since Taiwan has had four nuclear power plants, Taiwanese government agencies have devoted time to preparing emergency response for nuclear accidents. In light of these accidents, this study assists those Taiwanese government agencies to deal with spatial context-aware acquisition and representation. With Augmented-Reality (AR) and mobile Three-Dimensional (3D) graphics, this study proposes a mobile relief work system that helps relief workers to comprehend the relationship among their localities, the targeted structures, and the anticipated shelters. Based on the testing results regarding escaping victims in a simulated case study, relief workers quickly arrived at the targeted structures and successfully found the anticipated shelters. Over-all, this study is useful for those providing emergency response during nuclear accidents.

Keywords: Augmented-Reality, Disaster Management, Nuclear Accidents, Spatial Information, Three-Dimensional graphics.

1 Introduction

In the past, several nuclear accidents have became unfortunate disasters. For example, an explosion caused the Chernobyl nuclear power plant accident in Ukraine in 1986, and an earthquake triggered the Fukushima Daiichi nuclear power plant accident in Japan in 2011. The International Nuclear and Radiological Event Scale [1] indicates the level of these two nuclear accidents as the worst conditions. Many countries and civilians were exposed to widespread radiological materials. The ecological environment was contaminated, and the surrounding areas suffered economic losses of tens of billions of dollars.

Since Taiwan has four nuclear power plants, Taiwanese government agencies have established various emergency strategies to address nuclear accidents, especially for the use of information techniques; i.e. Global Positioning System (GPS), Geographical Information System (GIS), computer simulation, and Three-Dimensional (3D) graphics [2]. However, if nuclear accidents occur, because of the lack of spatial context-aware between accident sites and the neighboring areas,

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_31,

cooperation between disaster managers and relief workers is insufficient to provide effective emergency response. Therefore, this study focuses on improving the recognized problems (i.e., insufficient information service in GPS and GIS and inconvenient information operation in computer simulation and 3D graphics).

2 Literature Review

For relief work during nuclear accidents, with the rapid development of information technology, the following information techniques are useful:

- GPS and GIS: In many disasters, relief work is most effective in the first 48 hours following the incident [3]. When disaster managers completely understand the correlated information regarding nuclear accident sites and the affected areas (e.g., population, shelters, hospitals, spatial locations), they are able to rapidly draft response strategies (e.g., allocating first-aid resources and establishing temporary medical centers). The integration of GPS and GIS fulfills such a requirement, since GPS identifies the accurate localities of nuclear accidents and GIS provides a wide range of information acquisition [4, 5].
- Computer simulation and 3D graphics: Depending on distance and time, the distributions of radiological doses surrounding accident sites are different [6]. Disaster managers could assess the on-site radiological situation via monitoring data and help make emergency health protection decisions. When the monitoring data is sparse or the measured quantity is inappropriate, computer simulation (e.g., model predictions and statistical analysis) is an alternative tool to improve the understanding of the limited monitoring results [7]. Also, to assist disaster managers to comprehend the obtained accident information, visualization of analysis results through 3D graphics has frequently been used, incorporating elements such as hue and saturation of color, alterations in focus and resolution, broken lines, sharpness of pattern, and overlaying geometries [8].

3 Problems

Although the aforementioned information techniques help relief workers to cope with emergency scenarios during nuclear accidents, this study still recognizes two main problems. There are insufficient information services related to GPS and GIS and inconvenient information operation in computer simulation and 3D graphics, respectively.

In order to successfully arrive at accident sites, relief workers benefit from the integration of geographical information, electronic maps, and mobile devices [2]. For example, when relief workers attempt to understand the pathways from their locations to accident sites, they could immediately obtain assistance through route planning in Google Maps. However, due to several factors (e.g., bad visibility at night, ambiguous geographical in-formation in developing countries), relief workers still spend much time looking for the correct pathways [3]. Meanwhile, some GIS-based applications

merely show 2D/3D objects instead of the details (e.g., floor-plan and structural designing drawings), and some GPS-based applications work well only in outdoor environments [9]. Therefore, when relief workers evacuate victims from some structures (e.g., multistory buildings and large-area structures), they have difficulty identifying the indoor shelters. In other words, although relief workers successfully arrive at the targeted structure at an accident site, they still waste time searching for shelters and victims. Obviously, insufficient information service in GPS and GIS for relief workers presents a need for improvement.

Moreover, during nuclear accidents, radiological materials can be absorbed into structure components (e.g., roofs, windows, walls). The semi-collapsed structure components also provide crevices for the permeation of radionuclides [10]. Disaster managers may ask relief workers to validate the assessment of radiological dose, particularly in densely-populated cities. Recently, various monitoring sensors (e.g., seismometers and surveillance facilities) were installed in many new constructed edifices. Through the integration of the monitoring sensors and applications based on Computer-Aided Design (CAD), disaster managers can comprehend the extent of damage for structures, since the detected damage is displayed on structural design drawings. However, relief workers cannot move rapidly to specified spots to inspect the radiological doses present. Since many CAD-based applications are designed for computers and laptops [11], relief workers either have no devices to access structural design drawings or may be unfamiliar with the CAD-based applications. As a result, relief workers need to wait for inspection guidance from disaster managers. In other words, relief workers face inconvenient information operation in computer simulation and 3D graphics at accident sites.

4 Approach

To address these recognized problems, this study proposes a mobile relief work system consisting of two information techniques. Fig. 1 shows that this system offers transparent spatial context-aware acquisition and representation. The two adopted information techniques include Augmented-Reality (AR) and mobile 3D graphics. To complete the necessary components in the mobile relief work system, this study uses the Google Android development toolkits and Microsoft ASP.NET toolkits.

The mobile relief work system presents the specified geographical information on a Google Maps-based user interface and offers two system functions (i.e., AR-based out-door illustrator and mobile 3D-graphics indoor illustrator). During relief work at accident sites, Fig. 2 shows that the AR-based outdoor illustrator assists relief workers to arrive at the targeted structures. For instance, relief workers seek to find a six-floor building. After identifying the current global position, the AR-based outdoor illustrator displays the distance, pathway, and direction from their location to the targeted site. Based on this guidance, relief workers can enter the building. Since the information representation is combined with the periphery images, the relief workers are less likely to be confused by the surrounding environment.

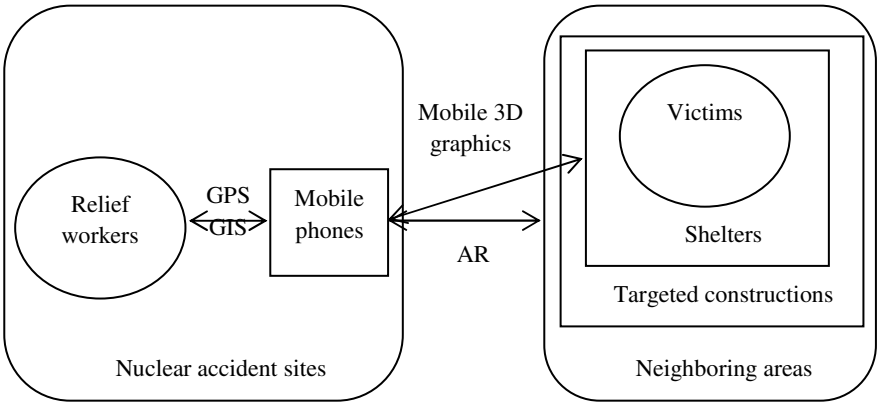


Fig. 1. Spatial context-aware acquisition and representation in this study

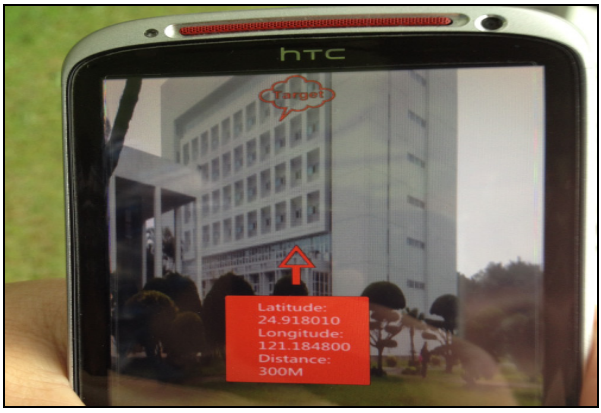


Fig. 2. Screenshot for AR-based outdoor illustrator

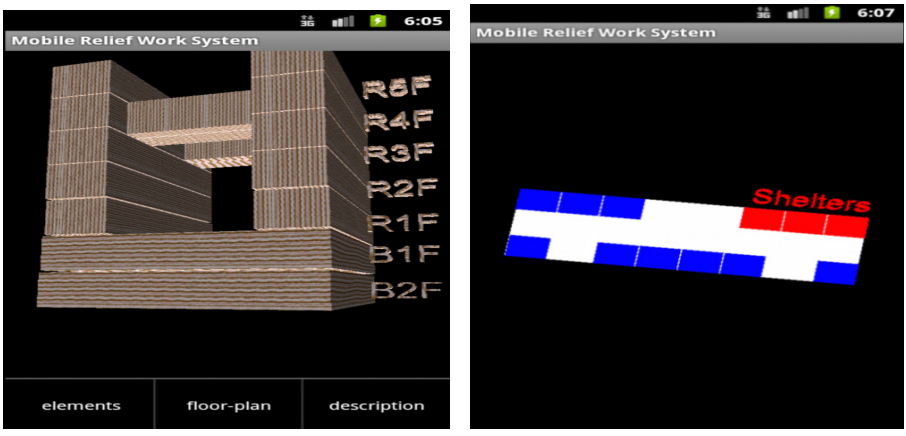


Fig. 3. Screenshot for mobile 3D-graphics indoor illustrator

When relief workers arrive at the targeted structure, Fig. 3 shows that the mobile 3D-graphics indoor illustrator helps workers find the shelters where victims are located or the inspection spots where radiological doses reach a state of alert. For example, in a hospital, the mobile 3D-graphics indoor illustrator presents the structural design drawings and illustrates various floors. Upon selecting a floor, relief workers can confirm the pathway from their location to the anticipated area. Since GPS may not work well at indoor environments, based on accelerometer sensors embedded in mobile phones, the mobile relief work system can automatically rotate 3D graphics objects while relief workers are moving. Therefore, relief workers effectively access and operate 3D graphics objects through the gesture and touch-enabled screens embedded in mobile phones.

5 Tests

During a simulated scenario, three relief workers used identical mobile phones to perform testes with electronic maps and the mobile relief work system. The relief workers needed to evacuate victims from three structures, including a mall, a university, and a 15-floor building. The shelters were in different locations of the three structures. The shelter of the mall existed on the north side of the three underground floors, that of the university was on the south side of the two underground floors, and that of the 15-floor building was located on the east side of the eighth floor. This study measured the usage time based on two stages (Table 1). In one stage (Outdoor), the relief workers moved from a specified location to the targeted structures; the other (Indoor) indicated when the relief workers arrived at the structures until they successfully found the shelters.

Table 1 shows that the testers completed the tests. Regarding the electronic maps, the relief workers took 26, 31, and 27 minutes to arrive at the structures and required 9, 19, and 17 minutes, respectively, to find the shelters. With the mobile relief work system, the relief workers needed less time (23, 28, and 22 minutes) to move to the structures and spent less time (9, 17, and 16 minutes) looking for the shelters. Clearly, in contrast to the electronic maps, the mobile relief work system offered better performance for the testers. For these tests, this study recognized that two variables affected the testing results: testing environment and experience. For the testing environment, the relief workers spent more time in the outdoor environment regardless of the electronic maps and mobile relief work system, since the distances from the outdoor locations to the structures were longer than those from the structures to the shelters. When the relief workers sought the shelters, the floor areas of the structures affected the use time. For example, the area of the university was wider than that of the mall and the 15-floor building, so the relief workers spent more time completing the test. Also, the experience decreased the usage time for relief work in the indoor environment. Since the tests with the mobile relief work system took place in the second testing, the relief workers might finish the tests based on their previous memories.

Table 1. Test results

	Using time (Minutes)			
Period	Electronic maps		Mobile relief work system	
Constructions (Shelters)	Outdoor	Indoor	Outdoor	Indoor
A mall (3rd underground floor)	26	9	23	9
An university (2nd underground floor)	31	19	28	17
A 15-floor building (8th floor)	27	17	22	16

6 Conclusion

During nuclear accidents, effective relief work abridges the losses. To offer spatial context-aware acquisition and representation for relief workers at nuclear accident sites, this study applies AR and mobile 3D graphics techniques to construct a mobile relief work system. This system helps relief workers to comprehend on-site geographic relationship among their localities, the targeted structures, and the anticipated shelters. The test results revealed that relief workers rapidly finished relief work through the mobile relief work system. For future research, application of the mobile relief work system in various disaster training courses would assist relief workers to enrich their experiences regarding emergency response.

Acknowledgement. The authors would like to thank all members in the Research Center for Hazard Mitigation and Prevention, National Central University. During the period of this study, the financial support came from NCU101G901-11, NCU102G901-11, and NSC102-2218-E-008-007.

References

1. International Atomic Energy Agency, http://www.iaea.org/Publications/Factsheets/English/ines.pdf
2. Montoya, L.: Geo-data acquisition through mobile GIS and digital video: an urban disaster management perspective. Environmental Modelling & Software 18, 869–876 (2003)
3. Farthing, D.W., Ware, M.: When it comes to mapping developing countries, disaster preparedness is better than disaster response. In: AGI GeoCommunity 2010: Opportunities in a Changing World 'Innovate - Connect - Succeed' (2010)
4. Haywood, S.H.: A method for displaying imprecision in early radiological emergency assessments. Journal of Radiological Protection 30, 673–685 (2010)
5. Battista, C.: Chernobyl: GIS model aids nuclear disaster relief. GIS World 7, 32–35 (1994)
6. Mettler Jr., F.A., Voelz, G.L.: Major radiation exposure - what to expect and how to respond. New England Journal of Medicine 346, 1554–1561 (2002)

7. El Harbawi, M., Mustapha, S., Choong, T.S.Y., Abdul Rashid, S., Kadir, S.A.S.A., Abdul Rashid, Z.: Rapid analysis of risk assessment using developed simulation of chemical industrial accidents software package. International Journal of Environmental Science and Technology 5, 53–64 (2008)
8. Bero, M.A., Gilboy, W.B., Glover, P.M.: An optical method for three-dimensional dosimetry. Journal of Radiological Protection 20, 287–294 (2000)
9. Bill, R., Cap, C., Kofahl, M., Mundt, T.: Indoor and Outdoor Positioning in Mobile Environments - a Review and some Investigations on WLAN-Positioning. Annals of GIS 10, 91–98 (2004)
10. Hong, K.J., Lazaro, M.A.: Radiological assessments for the national ignition facility. Fusion Technology 30, 1511–1515 (1996)
11. Tsai, M.K., Yau, N.J.: Improving information access for emergency response in disasters. Natural Hazards 66(2), 343–354 (2013)

Resource Analysis for Mobile P2P Live Video Streaming

Jongmyoung Kim and Seungchul Park

School of Computer Science and Engineering, Korea University of Technology and Education, 1800 Chungjeol-ro, Byeongcheon-Myun, Dongnam-gu, Cheonan, Chungnam, 330-708, Korea
scpark@kut.ac.kr

Abstract. Mobile resources such as smartphones and mobile links are quite different from the fixed resources from the viewpoint of processing power, battery power, and bandwidth. Since the mobile resources are more limited and dynamic than the fixed resources, realtime characteristics of the corresponding resources need to be timely reflected so as to support the quality of service of bandwidth-intensive and high power-consuming P2P(peer-to-peer) live video streaming applications. This paper will concentrate on analyzing the characteristics of mobile resources, mainly focusing on the processing power of smartphone, mobile WiFi bandwidth, and battery power which need to be importantly taken into consideration for mobile P2P live video streaming. Based on the analysis of those mobile resources, how to support quality of service for mobile P2P live video streaming will be also briefly discussed in this paper.

Keywords: Peer-to-Peer Communication, Live Video Streaming, Mobile P2P, Quality of Service.

1 Introduction

P2P(Peer-to-Peer) technology has been widely deployed for the video streaming applications as well as the conventional file sharing applications because of its high potential in scalability and availability[1]. Though most of the existing P2P video streaming applications have been developed on the basis of personal computers and fixed residential networks, currently mobile terminals including smartphones are heavily involved in the P2P applications, but as just client-only and free-riding peers[2,3]. That was because mobile resources such as smartphone and wireless link have been regarded to be more limited and dynamic than the personal computers and fixed residential networks which were widely deployed by the legacy P2P applications. In order for the P2P technology to spread out more widely, it is imperative to leverage the rapidly growing mobile smartphones as contributing intermediate peers, not just free-riders. Recently though there have been some research efforts for mobile P2P video streaming technology, they are just focused on checking its feasibility. Since the mobile environments have been rapidly changed so as to support higher processing capability and bandwidth, it is believed that this is right time to concretely analyze the mobile resources to check if they are affordable to support the QoS(Quality of Service) of bandwidth-intensive and high power-consuming P2P live video streaming applications.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_32,

In this paper, we concentrate on analyzing the processing power of smartphone, mobile WiFi bandwidth, and battery power of smartphone which need to be importantly taken into consideration to support the mobile P2P live video streaming applications. Regarding to the processing power of smartphones, we check if current smartphones have sufficient processing capability to support the P2P live video streaming applications as intermediate peers. We investigate two aspects of the mobile WiFi link - the bandwidth capability and the dynamicity which need to be very crucially considered in the peer selection and QoS control. Since the existing client-only mobile video streaming is already known as a high power-consuming application, the intermediate P2P live video streaming smartphones are believed to consume much more power. The battery power consumption of the intermediate P2P video streaming smartphone is precisely analyzed in this paper. Based on the analysis of those mobile resources, how to support quality of service for mobile P2P live video streaming is also briefly discussed in this paper.

2 Related Works

[2,3,4] proposed to deploy mobile devices just as client-only free riding peers in P2P streaming because mobile resources are too limited and dynamic. In those schemes, mobile terminals do not contribute their resources to the P2P streaming but only use the resources of other fixed nodes. Since, however, currently high-quality smartphones are so rapidly spreading and more popularly used as streaming devices than PCs, the well-known advantages of availability and scalability of P2P streaming cannot be expected any longer without leveraging the resources of mobile devices such as smartphones. [5,6] showed that a flexible smartphone-based P2P streaming could be implemented in dynamic mobile networks by using the standard RTSP(Real-Time Streaming Protocol) and the substream concept for a video stream. [5,6] considers only RTT(Round Trip Time)s to select a peer. In addition to the RTT, other dynamic characteristics of mobile resources such as processing power, battery remains, and link bandwidth need to be considered for the peer selection in order to support QoS more effectively. Since those mobile resources are more dynamic and limited, the QoS of a mobile P2P streaming need to be timely monitored and correspondingly supported. Though [7] suggested a policy-based dynamic QoS control scheme in which a policy server collects all QoS control related information and decides peers' QoS policies, it does not present what resources are needed to be monitored and how they are monitored in realtime. This paper is written with two objectives which differentiate this paper from the previous works. One is to analyze the characteristics of mobile resources to see if they are affordable to support P2P live video streaming. The other objective is to suggest how the characteristics of mobile resources should be reflected in the P2P overlay network construction so as to support QoS more effectively.

3 Characteristics of Mobile P2P Resources

Table 1 shows the test environment where characteristics of the mobile resources are measured. We used Samsung Galaxy smartphones, IEEE 802.11g WiFi network

connected with 100Base-TX and Gigabit Ethernet. H.264 video and 16-bit PCM audio were combined to generate 222Kbps stream. The characteristics of bandwidth were measured in three different link environments, signal strength of -47dbm, -55dbm, and -78dbm respectively.

Table 1. Test environment

Items	Specification
Smartphone	Samsung Galaxy S/S2/S3/S4
WiFi	IEEE 802.11n
Video&Audio Stream	228Kbps, H.264 Video(320x240 15fps), 16-bit linear PCM Audio
Environments(Signal Strength)	Case1(-47 dbm), Case2(-55 dbm), Case3(-78 dbm)

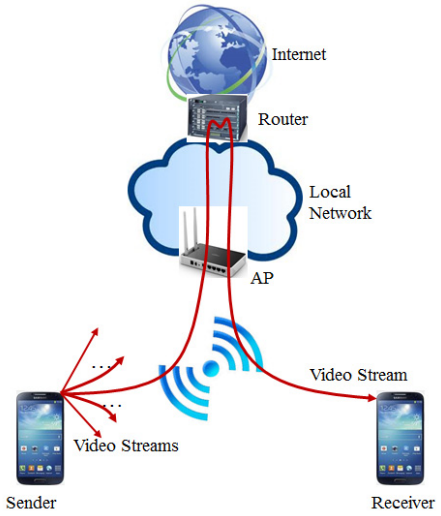


Fig. 1. Configuration of test network

The configuration of test network where the measurements were conducted is presented in Fig. 1. The sender smartphone captures video and audio in realtime, displays and sends the combined stream via an AP to one or more terminals. One of the streams is destined to the receiver, and the remained streams are destined to other terminals connected to a different AP. All streams sent by the sender are returned back at the default router of the local network which is composed of 100Base-TX and Gigabit Ethernet switches.

3.1 Processing Power of Smartphones

In order to check if the processing power of current smartphones are affordable to support P2P live video streaming, we monitored how the sender's FPS(Frames Per Second) were changing as the number of streams sent were increasing. After maximum(15) FPS of the corresponding camera and the related audio were captured first and displayed, and then the combined stream was relayed to other terminals. The number of peers to which streams were sent was gradually increased and how the frames sent per second were changing was measured in different types of Samsung Galaxy S, S2, S3, and S4. The result was presented in Fig 2. In the case of lowest model Galaxy S, the frames sent per second was kept until the number of streams sent reached around 40, S2 was 60, S3 was 85, and in the case of highest model S4, the frames sent per second did not drop until the number of streams sent reached over 200.

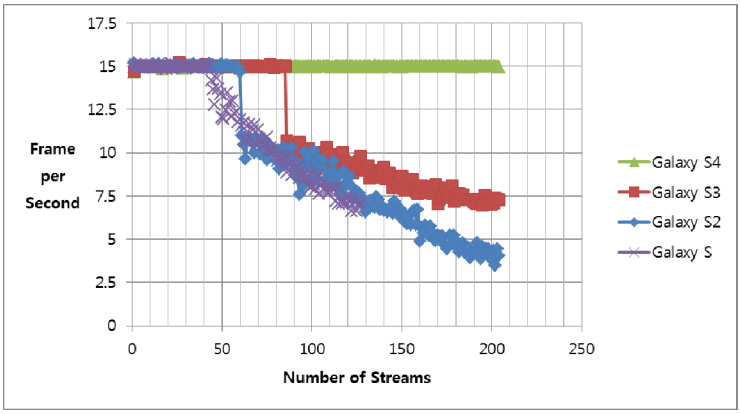


Fig. 2. Processing power of smartphones

From the above experimentation, we can see that current smartphones, even the lowest model, are capable enough to simultaneously support the live video and audio capturing, displaying, and relaying the video streams to more than several tens of peers. Since the processing power of smartphones are rapidly upgrading, their processing power is believed to be affordable to support P2P live video streaming. This means that the processing power is not a limited resource any longer for the P2P live video streaming. From the viewpoint of processing power, the smartphones can perform not only the role of intermediate peers but also the role of source peer in the P2P live video streaming.

3.2 Battery Power of Smartphones

Battery power of smartphones was checked with two objectives. First objective was to see if the battery power is powerful enough to support the P2P live video streaming. And second objective was to check the amount of battery consumption for the operations of capturing and displaying, and relaying the video and audio streams.

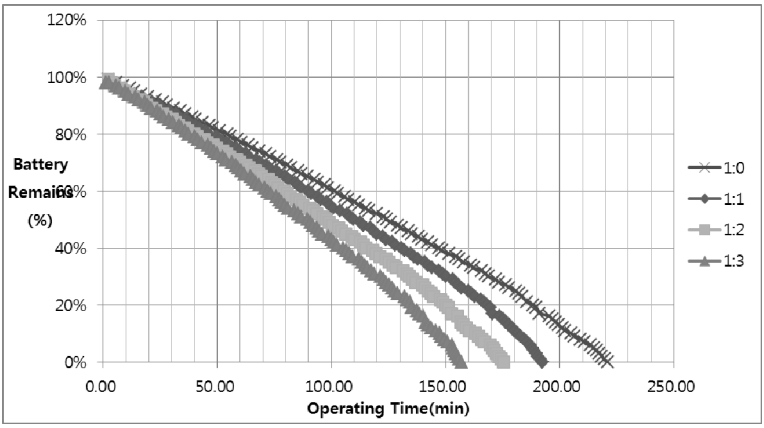


Fig. 3. Battery consumption of smartphones

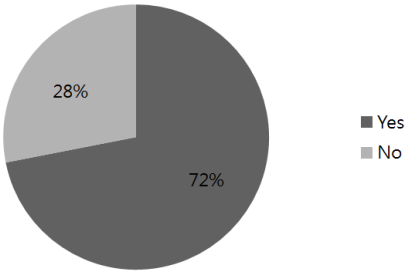


Fig. 4. Consciousness of battery remains of video streaming users

Fig. 3 shows how the battery of Galaxy S3 smartphone is consumed to support four different applications, capturing & displaying only(1:0), capturing and displaying : relaying to 1 peer(1:1), capturing and displaying : relaying to 2 peer(1:2), and capturing and displaying : relaying to 3 peer(1:3) respectively. From this experimentation, we learned that current smartphones, after fully charged, can endure more than 2 hours to simultaneously capture and display the video and audio, and relay the stream to several peers. That is understood to be affordable to support the P2P video streaming. However, we also learned that the video stream relaying, necessarily required by intermediate peers of the P2P video streaming, is a high power-consuming job. We surveyed 100 video streaming users to see how they are conscious of battery remains while streaming video in current client-only mode. As shown in Fig. 4, 76% of users answered that they are conscious of battery remains and most of them refrain from using the streaming service when the battery remains is going down below 10 to 30% so as to keep more crucial services such as phone alive longer. That means that users will get more conscious of battery remains when they are involved in the higher power-consuming P2P live video streaming as intermediate peers, and this will incur those users to churn more often. Thus, the status of battery

power of smartphones needs to be carefully taken into consideration in the peer selection in order to reduce the churning rate.

3.3 Bandwidth of P2P WiFi Link

The bandwidth of WiFi link was measured in three different environments where corresponding signal strengths are significantly different. Fig.5, Fig. 6, and Fig. 7 present the result of measurements.

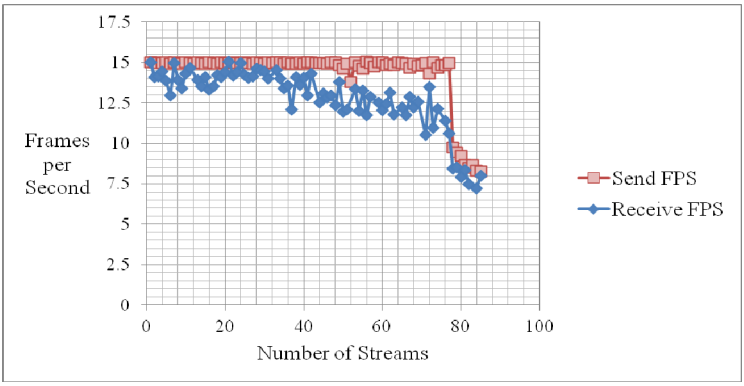


Fig. 5. Bandwidth of WiFi link : case1(-47dbm)

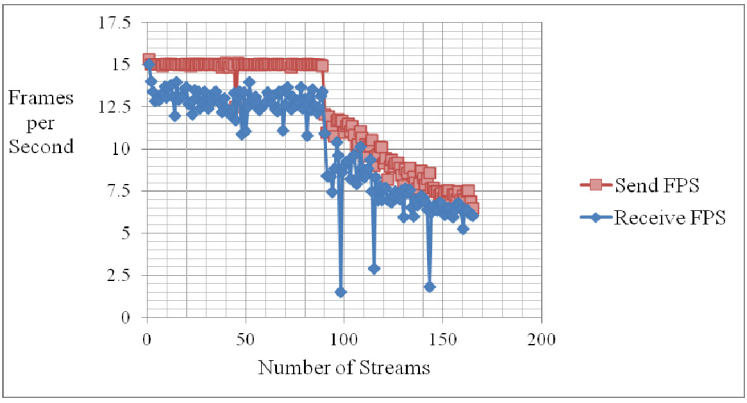


Fig. 6. Bandwidth of WiFi link : case2(-55dbm)

From the measurement results, as already known well, we can actually see that the mobile links are very dynamic from the viewpoint of bandwidth according to the environment. In the case 1 environment of signal length -47dbm, 14-15 FPS was kept in the receiver until the number of streams sent by the sender reached over 30+. In the environment 2 of -55dbm, approximately over 13 FPS was kept in the receiver until reaching nearly 80 streams sent. These show that current WiFi links have sufficient

bandwidth so as to be able to relay several tens of live video streams to peers if the condition of corresponding links is good enough. Whereas, in the case of poor link condition presented in Fig. 7, the FPS of receiver side downed often below 10 so that the QoS of the corresponding stream is not likely to be well maintained.

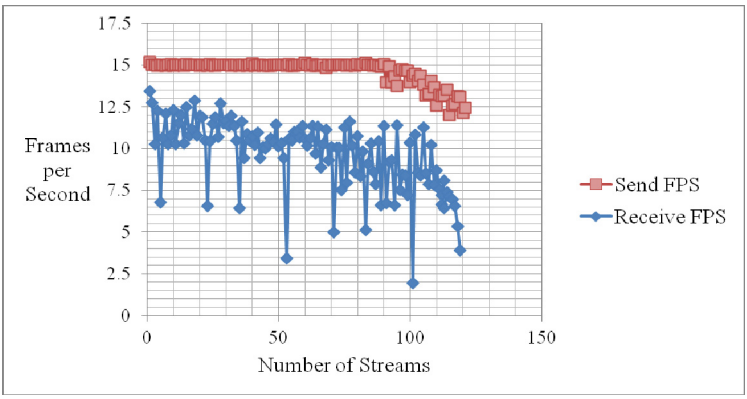


Fig. 7. Bandwidth of WiFi link : case3(-78dbm)

From the above experimentations, we can understand that widely deployed WiFi links are basically good enough to support the mobile P2P live video streaming. Since, however, the WiFi links are so dynamic, according to their conditions, as to result in high fluctuations of bandwidth, realtime monitoring and reflecting link status in the peer selection as well as the QoS control is compulsorily required in order to effectively support the mobile P2P live video streaming.

4 Concluding Remarks

Until several years ago, placing mobile terminals as client-only free-riding peers in the P2P streaming, because of their limited and dynamic resource characteristics, was naturally accepted. But, the situation has been changed. The number of smartphones was extremely increased, and their processing power and battery power have been rapidly upgraded. Moreover, mobile links, especially WiFi, have been widely spreading and their bandwidth has been getting higher and higher. From the measurements conducted in this paper, we learned that we do not have to place the mobile smartphone terminals as free-riders any longer. The result of measurements shows that current smartphones have already sufficient processing power to support the live P2P video streaming. Even though the battery power of mobile terminals has been regarded as a very limited resource, we showed that the battery power of current smartphones is affordable to support the P2P video streaming. However, we also learned that the status of battery power of smartphones needs to be carefully taken into consideration in the peer selection in order to reduce the churning rate because the video stream relaying is a high power-consuming task. Regarding to the mobile

WiFi link, we showed that current WiFi networks can provide sufficient bandwidth for P2P live video streaming. However, in order to overcome the dynamic characteristic of mobile links in P2P live video streaming, we learned that realtime monitoring of link status and its corresponding bandwidth need to be carefully considered in the peer selection as well as the QoS control.

References

1. Peltotalo, J., Harju, J., Jantunen, A., Vaatamoinen, L.: Peer-to-Peer Streaming Technology Survey. In: 7th International Conference on Networking, pp. 342–450 (April 2008)
2. Noh, J., Maker, M., Girod, B.: Streaming to Mobile Users in A Peer-to-Peer Network. In: Proceedings of the 5th International ICST Mobile Multimedia Communications Conference (September 2009)
3. Ting, W.-C., Liu, P.-C., Chang, S.-H., Chen, J.-W.: On Mobile Clients as Free Riders in Peer-to-Peer Live Streaming. In: IEEE PERCOM Workshops, pp. 471–476 (March 2011)
4. Sheu, S.-T., Huang, C.-H.: Mixed P2P-CDN System for Media Streaming in Mobile Environment. In: 7th International Wireless Communications and Mobile Computing Conference, pp. 657–660 (July 2011)
5. Peltotalo, J., Harju, J., Saukko, M., Vaatamoinen, L.: A Real-Time Peer-to-Peer Streaming System for Mobile Networking Environment. In: Proceedings of The INFOCOM and Workshop on Mobile Video Delivery (April 2009)
6. Peltotalo, J., Harju, J., Vaatamoinen, L., Bouazizi, I., Curcio, I.: D., D.: RTSP-based Mobile Peer-to-Peer Streaming System. International Journal of Digital Multimedia Broadcasting (2010)
7. Zhang, Y., Huang, X., Xia, H., Wang, N.: A Policy-based QoS Control Architecture for Mobile Peer-to-Peer Streaming Media. In: 2010 International Conference on E-Product E-Service and E-Entertainment, pp. 1–5 (November 2010)

A Study on Searchable Encryption System against Chosen-Ciphertext Attack

Sun-Ho Lee[1], Jae-Cheol Ryou[2], and Im-Yeong Lee[1,*]

[1] Department of Computer Software Engineering, Soonchunhyang University, Asan, Korea
{sunho431,imylee}@sch.ac.kr
[2] Department of Computer Engineering, Chungnam National University, Daejeon, Korea
jcryou@cnu.ac.kr

Abstract. The cloud computing service with which they may outsource the data storage and process them with various terminals at any time any where is getting vitalized as the network and computing technology develops. Especially, DaaS(Database as a Service) is widely used out of cloud computing services. However, it needs to encrypt the database stored in outsourcing storage because there is a security problem that the database stored in the server could be read by hackers and unethical server administrator without prior consent of the owner if the individual's sensitive information such as its physical information is stored without being encrypted in the database. However, entrusting the data storage makes no sense if the owner has to search the data by decrypting them after downloading all encrypted data into his terminals in order to search his data in safe with the existing encryption algorithm. A Searchable Encryption System has currently been appeared in order to solve such a problem. But, the existing searchable encryption system has the same typed trapdoors which are created to search the same keywords because numerous search queries are transmitted to the entrusted storage center and the administrator of the center shall be able to analogize the keywords and learn which data the user stores and search through such queries. Thus, this paper proposes a Searchable Encryption System using the one time trapdoor which prevents the unethical server administrator from analogizing the contents of search and data through search queries by creating the different trapdoor whenever the same user searches the same keywords.

Keywords: Searchable encryption, Data outsourcing, Outsourcing storage, Cloud storage, Distributed file system, Anti Replay Attack.

1 Introduction

The cloud computing service with which the enterprises and individuals entrust to store the data to reduce the responsibility to and save the operational cost for managing the server and are able to process the data with various terminals whenever and wherever is getting vitalized as the network and computing technology develops.

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*, Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_33,

Especially, DaaS out of cloud computing services is widely due to the strength that only the capacity actually used is charged without a limit to capacity extension. Actually, according to the survey of a research institute, 69% of people in USA was appeared to use the cloud computing. However, it needs to encrypt the database entrusted to be stored because there is a security problem that the database stored in the server could be read by hackers and unethical server administrator without prior consent of the owner if the individual's sensitive information such as its physical information is stored without being encrypted in the database. However, outsourcing the data storage makes no sense if the owner has to search the data by decrypting them after downloading all encrypted data into his terminals in order to search his data in safe with the existing encryption algorithm. A Searchable Encryption System has currently been appeared in order to solve such a problem[1-8]. It is an encryption method to search the data without decrypting the encrypted data using the characteristics of semi-homogeneous code and is currently studied by many researchers.

The process to use the searchable encryption system is as follows: First, the owner of the data shall encrypt the data which he desires to entrust to store and stores them into the 3rd place such as the Cloud. Then, he shall create the trapdoor by encrypt the keywords in order to search the data which have specific keywords out of data he stored. The trapdoor created shall be transmitted to the server and the server shall search the encrypted data with the trapdoor. However, the server shall not be able to know the information of keywords or data through the search process but only know whether or not the data fall into the keywords. It shall be problematic to assume that this searchable encryption system has a complete security with a reason that it searches the encrypted data with trapdoor. The existing searchable encryption system has the same typed trapdoors which are created to search the same keywords because numerous search queries are transmitted to the entrusted storage center and the administrator of the center shall be able to analogize the keywords and learn which data the user stores and search through such queries.

Thus, this paper proposes a Searchable Encryption System using the one time trapdoor which prevents the unethical server administrator from analogizing the contents of search and data through search queries by creating the different trapdoor whenever the same user searches the same keywords.

2 Preliminaries

In this section, we provide the necessary preliminary details.

2.1 Bilinear Maps

The bilinear map was proposed originally as a tool for attacking elliptical curve encryption by reducing the problem of discrete algebra on an elliptical curve to the problem of discrete algebra in a finite field, thereby reducing its complexity. However, this method has been used recently as an encryption tool for information

protection, instead of an attacking tool. Bilinear pairing is equivalent to a bilinear map. These terms are defined and the theory is described below.

Definition 1. Characteristics that satisfy an admissible bilinear map are as follows.

- Bilinear: Define a map $e = G \times G \rightarrow G_T$ as bilinear if $e(aP, bP) = e(P, Q)^{ab}$ where all $P, Q \in G$, and all $a, b \in Z$.
- Non-degenerate: The map does not relate all pairs in $G \times G$ to the identity in G_T. Note that G and G_T are groups of prime order, which implies that if P is a generator of G, e(P, P) is a generator of G_T.
- Computable: There is an efficient algorithm to compute e(P, Q) for any $P, Q \in G$. The following definition was constructed based on the bilinear map $e(aP, bQ) = e(P, bQ)^a = e(aP, Q)^b = e(P, Q)^{ab} = e(abP, Q) = e(P, abQ)$. With this map, the D-H decision problem can be solved readily for ellipses using the following equation: $e(aP, bQ) = e(cP, P) \Rightarrow ab = c$. Therefore, the following is the basis for resolving the difficulties of the bilinear map, which is used as an encryption tool by many encryption protocols.

Definition 2. When the elements G, P, aP, bP, cP (BDHP, Bilinear Diffie–Hellman Problem) are given, this relates to the $e(P, P)^{abc}$ calculation problem. In this study, the admissible bilinear map was used as the basis for secret number production during the key construction process between heterogeneous devices. This problem can be solved if the ellipse curve discrete mathematics problem can be solved. For example, a can be calculated from aP, so $e(P, P)^{abc}$ can be calculated using $e(bP, cP)^a$.

2.2 Requirements

Following requirements shall be satisfied in order to store and search the data in safe in such entrusted storage environment.

- **Confidentiality:** The communication data between the remote storage center and the client's terminal could be identified only by the authorized entity. The trapdoor being created to search the data shall be created in different form whenever the user searches the same keyword.
- **Traffic efficiency:** The communication quantity shall not be large for efficiency of energy and network resources between the client and the server.
- **Calculation efficiency:** An efficiency of arithmetic operation to create and search the index shall be provided. In addition, an efficiency of arithmetic operation to share the data safely with other users shall be also provided.

3 Proposed Scheme

In this paper, an searchable encryption system using the one time trapdoor is proposed considering the structural characteristics of entrusted storage center. This paper shall describe what steps shall be progressed by data storage and search scenario.

3.1 System Parameters

- ***p*:** prime number
- ***G*:** cyclic additive group of order *p*
- ***G*$_T$:** cyclic multiplicative group of order *p*
- ***g*:** generator of *G*
- ***e*:** bilinear map, $G \times G \rightarrow G_T$
- ***sk*$_*$:** *'s private key in Z_p
- ***pk*$_*$:** *'s public key in *G*
- ***k*:** data encryption key
- ***EC*$_k$():** symmetric key encryption by key *k*
- ***DC*$_k$():** symmetric key decryption by key *k*
- ***w*$_*$:** *th keyword of data
- ***W*:** set of w$_*$
- ***pd*:** plain data
- ***ed*:** encrypted data
- ***H*():** hash function, $G_T \rightarrow Z_p$
- ***H*$_1$():** hash function, $\{0,1\}^* \rightarrow G$
- ***H*$_2$():** hash function, $G_1 \rightarrow G_1$
- ***T*$_*$:** trapdoor searching keyword *

3.2 Definition

The detailed steps performed by the proposed method are as follows.

- **KeyGen:** The storage outsourcing users generate public key pairs, which are created prior to using the service. The storage outsourcing server should not know the user's private key. If the private key is leaked, the attacker can generate a trapdoor by acting as the owner of the private key. Thus, we generate a key pair based on the Discrete Logarithm Problem (DLP).
- **Enc(*sk, W, pd*)→*E, ed*:** The data owner creates the encrypted index, *E*, and encrypted data, *ed*, which only the owner can search by inputting their own private key, *sk*, and set of keywords, *W*, which are sent to the master server.
- **TGen(*sk, w*)→*T*$_w$:** To search the data safely, the user creates a trapdoor, T_w, which does not leak information related to the keyword *w*, which is being searched for using the private key *sk*. The trapdoor is sent to the master server. The storage outsourcing administrator should not be able to access information via a trapdoor.
- **Test(*E, T*$_w$)→"yes" or "no":** Using the trapdoor generated by the user's private key and the search keyword, the server performs a test to confirm whether the encrypted data contain the keywords. If the cipher text contains the keyword specified, the server sends a "yes" to the user and a "no" if not. Thus, the server cannot learn anything about the keywords or the data.
- **Dec(sk, E, ed)→pd:** The rightful owner of the encrypted data uses their private key to decrypt the encrypted data.

3.3 Storage Scenario

The proposed method considers the storage outsourcing structure so an encrypting index used for sharing and searching is stored on the master server. We assume that each user has received a key pair before using the storage outsourcing service (refer to Step 1). The user encrypts the necessary keywords during data searching so they can perform their own search later and send this to the master server (refer to Step 2). The master server sends chunk information to the user for data storage, who then divides the data into chunks and stores it on the designated chunk server.

Step 1. Key generation (KeyGen). Each storage outsourcing service user generates a key pair.
$x \in Z_p$ selection
$sk = x$ setting up
$pk = g^x$ setting up

Step 2. Index and data encryption (Enc). The data owner generates an encrypted index, which can be used for searching securely.
$A = pk_a^{hk}$ $(hk = H_1(k))$
$b_i = H_2(w_i)$
$B = \{b_1, b_2, \ldots, b_n\}$
$C = e(g, H_2(pk_a))^{hk} \cdot k$
$E_a = (A, B, C)$ output encrypted index for the master server
$ed = EC_k(pd)$ output encrypted data for the chunk server

3.4 Search and Reading Scenario

The user sends a trapdoor that can search data without exposing keyword information to the master server (refer to Step 1). The master server searches for the data with the keyword in the encrypted index using the trapdoor and then sends the chunk information that corresponds to the data to the user (refer to Step 2). The data retrieved is decrypted by the legitimate user (refer to Step 3). The user acquires the data by summing each chunk received from the chunk server that stores the data.

Step 1. Trapdoor generation (TGen). A user, *a,* who wants to search the data generates a trapdoor using the keywords and their secret key.
$r \in Z_p^*$
$X = pk_a^r$
$Y = H_2(w)^{sk_a \cdot r}$
$T_w = X \,||\, Y$

Step 2. Test. To confirm that the data contains the keywords sought by the user, the user performs the following tests with the public key, trapdoor, and crypt obtained from the server.

$e(pk_s,Y)=?e(X, b_i)^s$
$e(pk_s,H_2(w)^{ar}) =? e(pk_a^r, H2(w_i))^s$
$e(g^s, H_2(w)^{ar}) =? e(gar, H2(w_i))^s$
$e(g, H_2(w))^{ars} =? e(g, H2(w_i))^{ars}$

Step 3. Decryption (Dec). The user can perform the following decryption using their private key and the crypt obtained from the server.
$k = C/e(A, H_2(pk_a)^{-sk_a})$
$= C/e(pk_a^{hk}, H_2(pk_a)^{-sk_a})$
$= C/e(g^{ska \cdot hk}, H_2(pk_a)^{-sk_a})$
$= C/e(g, H_2(pk_a))^{hk}$
$= (e(g, H_2(pk_a))^{hk} \cdot k)/(e(g, H_2(pk_a))^{hk})$: output decryption key
$pd=DC_k(ed)$: output decrypted data

4 Analysis

The proposed method satisfies the following requirements.

- **Confidentiality:** With the proposed method, it is difficult for to analogize the contents of communication even though the vicious 3rd party taps the communication contents between the client and the server using the pairing. In addition, it is difficult to learn the keywords and contents of data which an unethical server administrator searches through queries because a different form of trapdoor is created whenever searched even though a trapdoor which searches the same keyword using an arbitrary value is created.
- **Traffic efficiency:** It provides an efficiency of communication quantity because only one round of communication process is needed for searching the keywords and super-encryption.
- **Calculation efficiency:** Unlike the existing method, the proposed method increased its capacity of arithmetic operation a little bit in order to create the trapdoor which changes every time.

5 Conclusion

A safe method to store and search the data which sets the security requirements and satisfies them considering the environment of entrusted storage center is proposed through this research. The proposed method provides a further safe search function with increased capacity of arithmetic operation comparing to previous researches. It is expected that a search using a multiple number of keywords shall become an important issue in order to search the data easily and smoothly from the entrusted storage center. Thus, a research on the super-encryption system which encrypts the indexes consisting of multiple keywords of variable lengths and search them smoothly shall be needed in future.

Acknowledgments. This research was supported by the MSIP(Ministry of Science, ICT&Future Planning), Korea, under the ITRC(Information Technology Research Center)) support program (NIPA-2013-H0301-13-1003) supervised by the NIPA(National IT Industry Promotion Agency).

References

1. Boneh, D., Crescenzo, G., Ostrovsky, R., Persiano, G.: Public Key Encryp-tion with Keyword Search. In: Proceedings of International Conference on the Theory and Applications of Cryptographic Techniques, Interlaken, Switzerland (May 2004)
2. Boneh, D., Waters, B.: Conjunctive, "Subset and Range Queries on Encrypted Data". In: Proceedings of the 4th Theory of Cryptography Conference, Amsterdam, Netherlands (Febraury 2007)
3. Hwang, Y.H., Lee, P.J.: Public key encryption with conjunctive key-word search and its extension to a multi-user system. In: Proceeding of First International Conference on Pairing-Based Cryptography, Tokyo, Japan (July 2007)
4. Bao, F., Deng, R.H., Ding, X., Yang, Y.: Private Query on Encrypted Data in Multi-User Settings. In: Proceeding of the 4th International Conference on Information Security Practice and Experience, Sydney, Australia (April 2008)
5. Kamara, S., Lauter, K.: Cryptographic cloud-storage. In: Proceedings of Workshops on Financial Cryptography and Data Security, Canary Islands, Spain (January 2010)
6. Ion, M., Russello, G., Crispo, B.: Enforcing Multi-user Access Policies to Encrypted Cloud Databases. In: International Symposium on Policies for Distributed Systems and Networks, Trento, Italy (June 2011)
7. Zhang, B., Zhang, F.: An efficient public key encryption with conjunctive-subset keywords search. Journal of Network and Computer Applications 34(1) (2011)
8. Yang, Y.: Towards Multi-user Private Keyword Search for Cloud Computing. In: Proceeding of International Conference on Cloud Computing, Singapore (July 2011)

A Compensation Cost-Aware Coordination for Distributed Long Running Transactions

Qing Lin and Jeongyong Byun

School of Computer Engineering, Dongguk University
qinglin9@gmail.com, byunjy@dongguk.ac.kr

Abstract. Web Service technologies make the automation of business activities that are distributed across multiple enterprises possible. Existing extended transaction protocols typically resort to compensation actions to regain atomicity and consistency. Firstly, a reservation-based transaction protocol is proposed to reduce compensation risk by preventing from concurrent conflicts. And time constraint is imposed on reservation to enhance the performance. Secondly, we theoretically illustrate a scheduling scheme that tries to satisfy various constraints as well as gain an optimized execution plan with minimum compensation cost. The specific solution is devised on a fundamental Serial Long Running Transaction (SLRT) Model and collected transactional properties of component services. Based on these conditions, a probabilistic cost evaluation method is defined. Finally, performance evaluation is conducted to prove the effectiveness of optimized schedule.

Keywords: Distributed Long Running Transaction, Compensation Cost, Reservation-based Protocol.

1 Introduction

The automation of business activities that are distributed across multiple enterprises becomes possible with the advent of the new generation of Internet-based technology, in particular, Web Services. Business activities typically involve related tasks that are loosely coupled and carried out over a long period of time [1].

However, such enterprise applications must operate with a high degree of availability and performance. WS-BusinessActivity, one of WS-Transaction components, which allows participants to commit unilaterally. Compensating actions may be invoked in the event of an error. However, technically compensating transactions can be difficult to design and program and even more difficult to test in a distributed Web Service environment. Due to various levels of service quality, poorly designed services induce inconsistent problems even unrecoverable [2]. What is more dangerous is cascading compensations caused by a large scale rolling back and recovery, especially long running transactions that are highly concurrently processed. Besides, considering business reality, some service providers may charge compensation cost, like ticket agents or retailer-stores, etc.

Therefore, in this paper, we proposed a reservation-based transaction protocol to coordinate business activities. This protocol is to solve high compensation risk by

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_34,

reducing the probability of database inconsistency. Then we dedicate to rescheduling of the life cycle of transaction to reduce compensation risk further. In order to do this, for one thing, a probabilistic method of quantitation and evaluation for schedule is devised as an important step before scheduling. For another thing, two scheduling algorithms are proposed to search for cost optimized schedule.

2 Related Work

The WS-Transaction specification describes coordination types that are used with the extensible coordination framework described in the WS-Coordination specification [4]. It defines two coordination types: Atomic Transaction (WS-AT) for individual operation and Business Activity (WS-BA) [3]. There are other protocols contrived to improve this protocol. One of them is written in paper [5], in which, the author criticizes the use of compensating transactions and suggests the use of a reservation to coordinate business activities. A great amount of analytical work is done in this paper. In contrast to their work, our work is mainly interest in a study of applying reservation-based protocol to WS-Transaction infrastructure at the middleware layer that integrates this protocol with workflow. Through simulation, we also discover that if time constraint is imposed to reservation, it obtains better performance.

In addition, the previous research [7] also explored scheduling, which maintains atomicity with minimum compensation cost and satisfies temporal constraints as well. In this paper, the probabilistic methodology for cost evaluation is inherited. However, different with those researches, our solution is raised specifically for the background of reservation-based protocol, in which every participant's job has two phases. These two phases bring more complexity than works [7]. What is more, the algorithm is designed to solve conflicts between resource holding time and minimum compensation cost. Meanwhile a multi-graph is proposed to improve algorithm's efficiency.

3 Motivating Scenario

We begin by using a simple example to show the motivation scenario. Fig.1 illustrates a data flow graph of a distributed purchase and supply planning system. This system needs to consume services and gain resources from multiple enterprises. The graph specifies the order in which component services are invoked, and the conditions under which a certain service may or may not be invoked. We use the term, serial long running transaction (SLRT) to denote orchestrated transaction as in Fig.1. In this scenario, Order Processing Service is performed earlier than others. According to its results, Inventory Service updates quantity in the database and Logistics Planning Service maps out a suitable delivery plan through the knowledge of goods type and destination location. The plan includes transportation cost and tools, e.g. truck. And then Finance Service is invoked to calculate the total fee that customer should pay for and payment request is sent to Bank Service immediately. This complex activity needs transactional support to guarantee its atomicity. Under the circumstance of reservation-based protocol, only when three positive reservation responses arrive

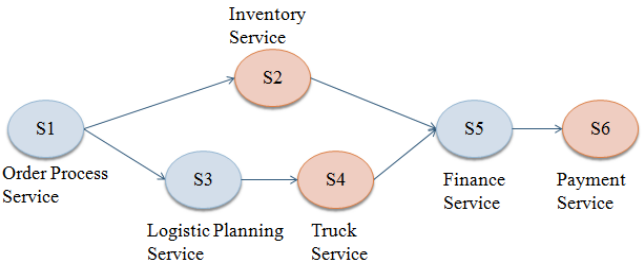


Fig. 1. SLRT of a distributed purchase and supply planning system

will the coordinator issue "complete" command to all the providers. Therefore it is necessary for every service provider imposes a time constraint for reservation due to better local resource allocation.

Let's see a natural schedule in Fig.2. The overall transaction experiences two fixed phases that we can see from Fig.2. The blue bar denotes reservation execution, and the orange one denotes completion execution. The white background represents deadline for reservation.

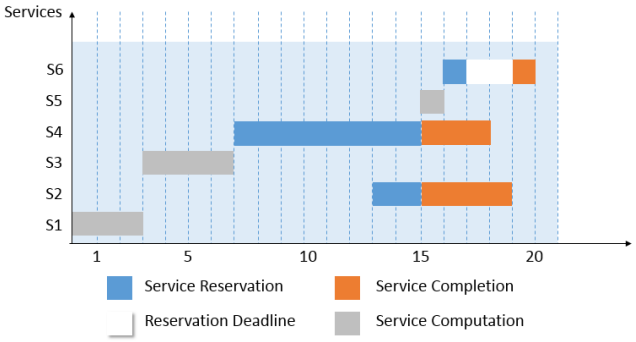


Fig. 2. A Natural Schedule

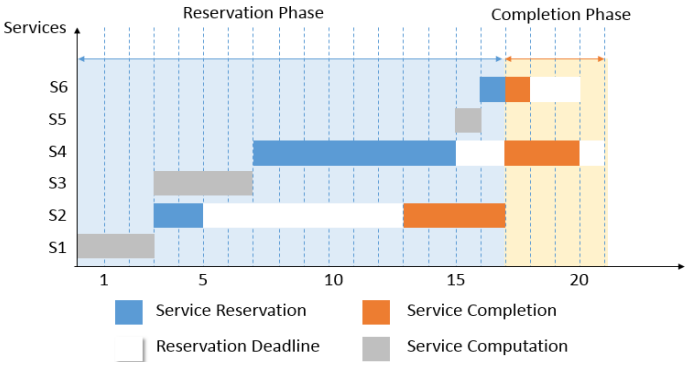


Fig. 3. A Cost Optimized Schedule

In case of aforementioned reservation time constraint problem, we assume transaction coordinator completes Inventory Service in advance, at time 13, with an aim to prevent the resource release. Fig.3 shows a cost aware optimized example. Apparently in a rescheduled transaction each individual service are pipelined to pursue cost optimization. Looking inside of the second schedule, Inventory Service is postponed behind Truck Service, which is prone to fail. Meanwhile, Bank Payment Service is scheduled at the rear of transaction since it is a high compensation cost service. Besides, mathematical evaluation of these two services is conducted as well. The result undoubtedly demonstrates the second schedule helps save cost more than 90% compared with nature scheduling according to formula that is defined in section 4.3.

4 Compensation Cost Aware Scheduling

4.1 Improved Reservation-Based Protocol

Reservation-based protocol is literally defined in Fig.4 [6]. Suppose there is a business activity involving *n* task from *n* participants (service providers). This model explains the conversation between coordinator and participant *j*. Here, *Dr(j)* is the time constraint for reservation that is declared by Participant *j*.

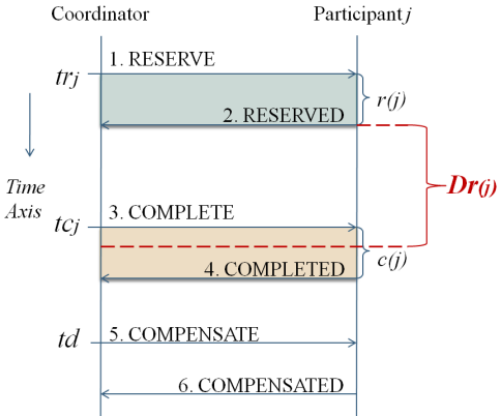


Fig. 4. Conversation Model of Reservation-based Protocol

4.2 Compensation Cost Model

We do provide a cost model based on the common knowledge about compensation penalty, by which we can develop our cost-aware optimization scheduling.

Compensation cost model is given for service *j* as follows:

$$\text{Cost}\ (S_j, td, tcj) = \begin{cases} 0 & td < tc_j \\ \alpha_j & tc_j \le td \le tc_j + c(j) \\ \beta_j & tc_j + c(j) < td \end{cases}$$

Here, α_j and β_j are non-negative penalty coefficients for Participant *j*'s completing and completed state respectively.

4.3 Compensation Cost Evaluation

The cost is produced when a failure occurs and failure occurs when negative responses from services' completion arrive at the coordinator.

Expected compensation cost, in case failure happens at tc_j +*c(j)*, can be computed by

$$Cost_{(tc_j+c(j))} = P_c \times P_f \times C$$

P_c is the success probability before time tc_j+*c(j)*
P_f is the failure probability at time tc_j+*c(j)*
C is the total compensation cost for all the involved services
The total cost is the integral of Cost from starting point to the termination of transaction is represented in the formula below. $S_{schedule}$ should decide *tc* and *tr* time for each service.

$$Cost(S_{schedule}) = \sum_{S_j \in S_{schedule}} Cost_{(tc_j+c(j))} \quad (1)$$

4.4 Multi-stage Graph Model Definition

An efficient solution of solving optimization problem, is to decompose the problem into different stages, and make intelligent decision for each stage. In the end, an optimized schedule could be figured out by composing all the decisions. For our compensation issue, the decision of completing services must increase compensation risk, for the reason that the failure risk as well as compensation requirements is increased.

A **multistage graph** is a graph *G = (V,E)* with *V* partitioned into *K>=2* disjoint subsets, which are called stage. All the Vertexes in *G* are classified into *K* stages. In other words, one stage must contain at least one vertex.

Stage: The vertexes are classified into *K* stages. Fig.5 is the illustration of stage segmentation for the purchase and supply system in our scenario. As we can see, a transaction's execution has two phases, reservation and completion. Reservations require sequential process which is decided by key path in data flow graph from Fig.1. In this example, the key path is sequence of *S1, S3, S4, S5, S6*. Meanwhile completions can run parallel. In other words, successful reservations on key path are

symbols of the end of stages, which are called as stage mark activities, e.g., *S1* reservation success indicates the end of *stage1*. Here, we define stage in a timed form, $stage = \{ ST_1, ST_2, ..., ST_k, ..., ST_K \}$, $ST_1, ST_2, ... ST_k, ... ST_K$ is the time that corresponding stages end. Since stage segmentation is acquired, our long running transaction can be scheduled based on this stage segments. Fig. 6 is one example according to stage in Fig. 5.

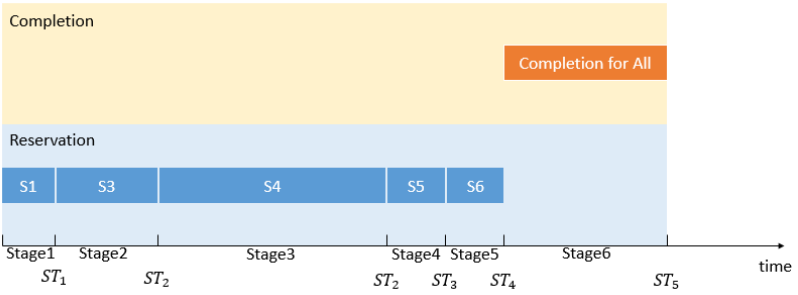


Fig. 5. Stage Segmentation Example

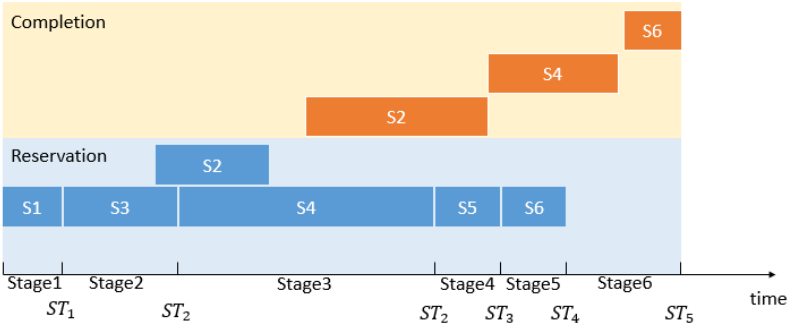


Fig. 6. An Example of Schedule based on Stage Segments

Vertex: The vertexes in multi-stage graph are the state of transaction. A vertex V_i is represented by two sets:

- $SET_{rev}(V_i)$, its members are reservation finished services.
- $SET_{comp}(V_i)$, its members are completion finished services.

For instance of Fig. 6, these two sets representing vertex *V* that belongs to *Stage 3,* are written as follows:

$$V = \begin{Bmatrix} SET_{rev}(V) = \{S_1, S_3, S_4, S_2\} \\ SET_{comp}(V) = \{\ \} \end{Bmatrix}$$

It means *V* is the temporary transaction status when services S_1, S_3, S_4, S_2 have been reserved, while none of them are completed yet.

Edge is a sub schedule for part of services, which leads transaction from V_{ki} *(stage k)* to $V_{(k+1)j}$ *(stage k+1)*. Here, i is the vertex number in *stage k*, and j is the vertex number in *stage k+1*, $i, j \geq 1$. This is illustrated in Fig.7.

Weight: Completions of services bring about compensation cost since more services need to be cancelled in case of transaction's abortion. The weigh on the edge denotes compensation cost that is raised by edge action. With an optimized schedule is found out for Edge $SE_{ki\rightarrow(k+1)j}$ in last step, the minimum cost, $Cost(SE_{ki\rightarrow(k+1)j})$, is already calculated by formula (1). Therefore, the value is assigned to the weight of the edge $SE_{ki\rightarrow(k+1)j}$.

Obviously, when transforming from one state to another, the transaction has to face increased compensation risk, for the sake of probability of failure and increased successfully completed participants. Based on the multistage graph, we assign the compensation cost to be the weight on the edge since that edge represents state transformation. Eventually many schedules are able to be found through multistage graph by selecting edge from one stage to the next. As shown in Fig.7, the red path is one of a number of schedules. To solve the problem, forward approach (also backward reasoning) can be applied.

In addition, we have to mention that each single sub schedule, also known as edge, can be optimized too. Traditional enumeration algorithm is efficient enough for sake of simplicity of inner stage scheduling problem.

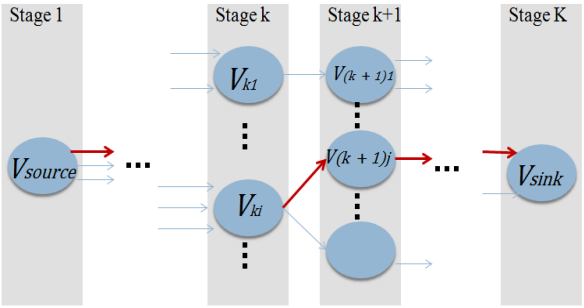


Fig. 7. Example of Multistage Graph and Path Selection

5 Simulation and Performance Evaluation

In this section, we study the performance of optimized schedule and test the correctness of cost evaluation formula. The simulations are carried out by CPN tool [9] and 3, 5, 7, 9 services are composed in one transaction, and costs are present in Fig.9 respectively.

Obviously optimized rescheduled transactions obtain least compensation cost, and almost two third of cost is saved compared with Reservation-based protocol, and 30

percent of the cost from original WS-BusinessActivity protocol. Therefore, our algorithm is able to guarantee atomicity of composite services under minimum cost, which greatly contributes to automation of business activities.

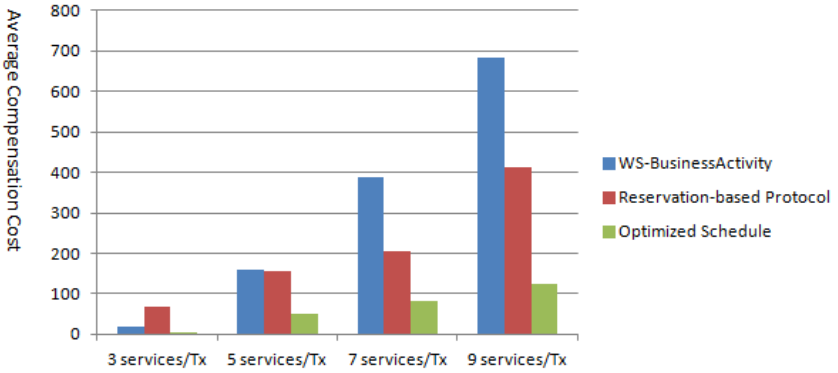


Fig. 8. Average Compensation Cost with Component Services Number Varied

6 Conclusion and Future Work

In this paper, we found that compensation can be considerably expensive and problematic. So our research contributes to a series of solutions for coordination distributed services in order to reduce compensation cost.

Firstly, a new protocol, defined as reservation-based protocol, is proposed to reduce the probability of compensation. Secondly, we theoretically illustrate a scheduling scheme that tries to satisfy various constraints as well as gain an optimized execution plan with minimum compensation cost.

In the future, we plan to apply our algorithms to Service Oriented Architecture platform, such as BPEL and WS-Coordinator framework.

References

1. Chatterjee, S., Webber, J.: Developing Enterprise Web Services: An Architect's Guide, p. 272. Prentice Hall PTR (2003)
2. Biswas, D.: Compensation in the world of Web Service Composition. In: Cardoso, J., Sheth, A.P. (eds.) SWSWPC 2004. LNCS, vol. 3387, pp. 69–80. Springer, Heidelberg (2005)
3. OASIS Web Service Business Activity (WS-BusinessActivity), found at: `http://docs.oasis-open.org/ws-tx/wstx-wsba-1.1-spec.pdf`
4. OASIS Web Service Coordination (WS-Coordination) Version1.2 (2009), `http://docs.oasis-open.org/ws-tx/wstx-wscoor-1.2-spec-os.pdf`
5. Zhao, W., Moser, L.E., Melliar-Smith, P.M.: A Reservation-Based Extended Transaction Protocol. IEEE Transactions on Parallel and Distributed Systems (2008)

6. Qing, L., Nasridinov, A., Hung, P.P., Byun, J.Y.: An Improved Reservation-based Protocol for Timely Web Service Transactions. Journal of KIISE: Computing Practices and Letters 17 (2011)
7. El Hadad, J., Manouvrier, M., Rukoz, M.: TQoS: Transactional and QoS-Aware Selection Algorithm for Automatic Web Service Composition. IEEE Transactions on Services Computing 3(1), 73–85 (2010)
8. An, L., Hai, L., Qing, L., Huang, L.-S., Xiao, M.-J.: Constraints-Aware Scheduling for Transactional Services Composition. Journal of Computer Science and Technology 24(4) (2009)
9. CPN tools download and documentations, found at, `http://cpntools.org/`

Ubiquitous Mobile Game Development Using Arduino on Android Platform

Andy S.Y. Lai and S.Y. Leung

Department of Information and Communications Technology,
Hong Kong Institute of Vocational Education, Hong Kong
andylai@vtc.edu.hk, AQA@live.hk

Abstract. This paper describes the use of Arduino on Android open-source platform in an experimental way to develop a remote control multi-player game in a ubiquitous computing environment. Android mobile is programmed to be a remote control device used to send sockets via Bluetooth network to Arduino microprocessor embedded in a toy car to control movements in all directions along its pathway. We applied the open-source microcontroller Arduino and the Java-Based technology Android in developing a multi-player mobile game in distributed ubiquitous computing, which strongly focuses on the emergence of technologies that embrace android mobile and Arduino open-sources. Our investigation focuses on an extended form of ubiquitous computing which game software developers utilize to develop remote control games for multi-players. We call this study an experimental ubiquitous computing application in which the Arduino embedded in the toy car can sense the color pattern changes with infrared along its pathway and instantaneously send the data via Bluetooth piconet to the connected Android mobile device. In turn, the Android mobile device sends the data to game server via web services on internet. Currently, mobile computing feeds data information into the game server. However, designing real-time ubiquitous mobile control game is still a daunting task and much theoretical and practical research remains to be done to reach the ubiquitous computing era. In this paper, we present the overall architecture and discuss, in detail, the implementation steps taken to create the Arduino and Android based remote control context-aware game. We prepare the client and server codes in ubiquitous computing, providing adaptive routines to handle connection information requests in telecommunication, logging and context formatting and delivery for speedy throughput and context-triggered actions.

Keywords: Distributed Ubiquitous Computing, Multi-player Game, Bluetooth, Context Aware System.

1 Introduction

One significant aspect of the merging mode of ubiquitous computing is the constantly changing execution environment. This work presents an experimental context-aware application that provides context-aware information to game server and game players in a mobile distributed computing environment. Our work is to apply the open-source microcontroller Arduino and the Java-Based technology Android in developing a

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_35,

multi-player mobile game in distributed ubiquitous computing, which strongly focuses on the emergence of technologies that embrace android mobile and Arduino open-sources. Our investigation focuses on an extended form of ubiquitous computing which game software developers utilize to develop remote control games for multi-players. We call this study an experimental ubiquitous computing application in which the Arduino embedded in the toy car can sense the color pattern changes with infrared along its pathway and instantaneously send the data via Bluetooth piconet to the connected Android mobile device. In turn, the Android mobile device sends the data to game server via web services on internet [1].

This experimental context-aware application can be viewed as a concrete example of automatic contextual reconfiguration and context-triggered actions in ubiquitous computing. Likewise, the implementation steps in this application can be viewed as micro-architectural elements of a ubiquitous computing framework that documents and motivates the semantics in ubiquitous computing in an effective way [2].

The rest of the paper is organized as follows: Section 2 presents the system architecture of remote control multi-player game using Arduino on an Android Platform. Section 3 describes the design and implementation of Android mobile control and explains the integration of Arduino microprocessor and Android device. Section 4 we conclude with a note on the current status of the project.

2 System Architecture

Android is the software activity designed to work with smartphones based on the Android open source operating system and Arduino is an open-source electronics prototyping platform based on flexible, easy-to-use hardware and software [3]. Yet Arduino can sense the environment by receiving input from a variety of sensors and can affect its surroundings by controlling lights, motors, and other actuators. The embedded microcontroller on the circuit board is programmed using the Arduino programming language (based on Wiring) and the Arduino development environment (based on Processing). Arduino can be stand-alone or they can communicate with software running on a computer. In addition, we adopt a software component called Amarino which consists of libraries help to interface with Android mobile application and Arduino microprocessor, and it brings in a new and complete dimension for mobile telecommunication [4]. We build our own interfaces for remote control application in Android application to access the smartphone sensors, like the accelerometer, light sensor, compass and touchscreen. Subsequently, we can freely control the remote Arduino embedded toy vehicle. This project provides three different game playing modes which comprise the toy vehicle physical motions, the game battle with infrared, and the toy vehicle time racing on track. Our study in this paper is to provide an embedded system in a remote toy vehicle in a distributed computing environment to support single-player and multi-player modes [5, 6] with context-awareness [7, 8].

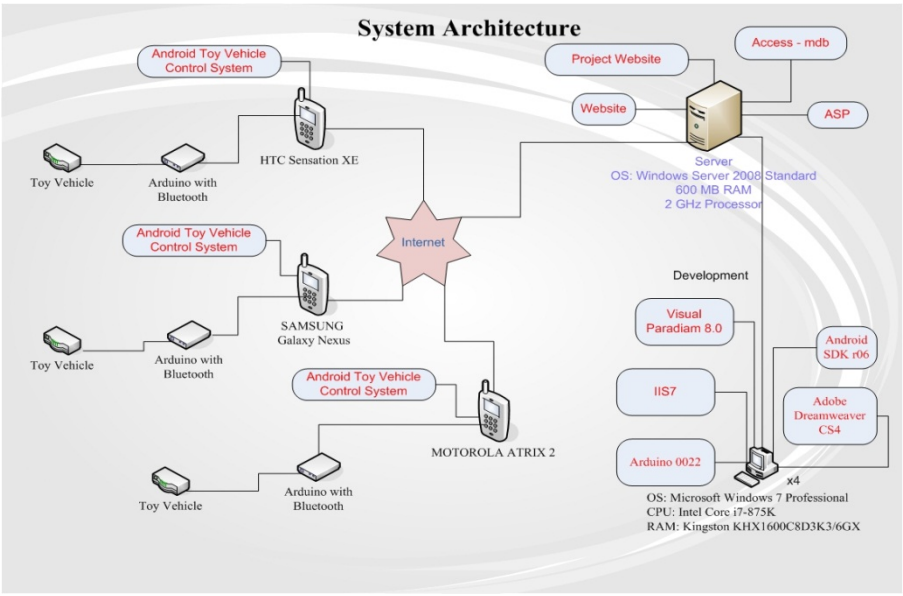


Fig. 1. Network Architecture for Arduino Multi-Player Game on Android Platform

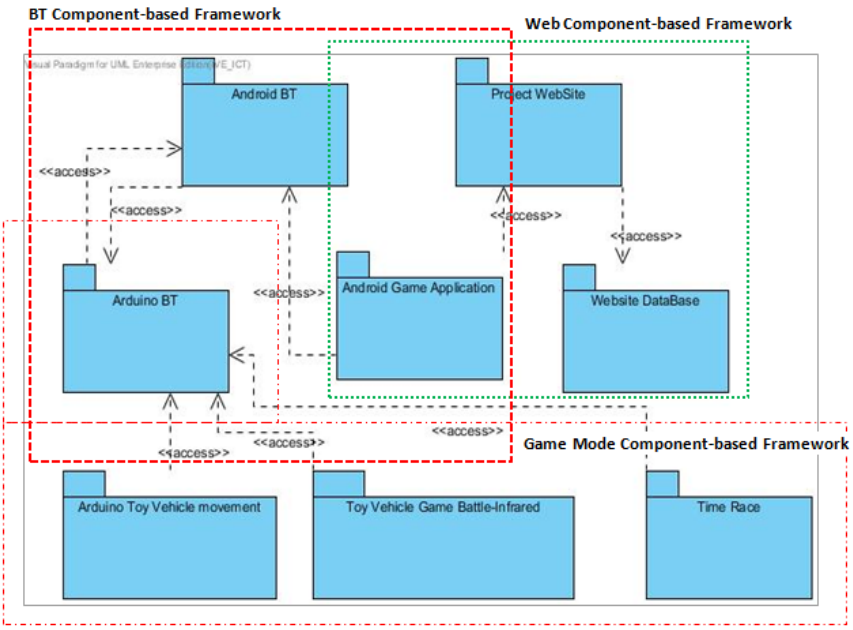


Fig. 2. Component-based Software Architecture for Arduino Multi-Player Game

Figure 1 shows a pictorial representation of the network architecture of Arduino multi-player game on Android platform. It consists of the distributed network components: Android Mobile Controllers, Arduino Embedded Toy Vehicle, Bluetooth-based Signal Transmission System, Web IIS Server, and Database Server. Figure 2 presents the component-based software architecture of a remote control multi-player game using Arduino on Android platform, which shows the communication between Android Application and toy vehicle via the bluetooth technology embedded in a smartphone and an Arduino-embedded toy vehicle.

The physical movement with remote control, the game battle fighting with infrared, and the car racing along destined tracks are the game playing components and they are correlated and dependents, which are illustrated and presented in the software architecture in the figure above. The network data communication takes places between different game playing components physically in Android mobile application and Arduino circuit microprocessor.

3 Android Motion Control Design and Implementation

The Android user interface design for Android mobile phone remotely controls a toy vehicle to move forward and backward at different speeds, turn right and left for different angles. The sequence diagram in Figure 3 presents how the Amarino, a toolkit software library, helps to establish a bluetooth connection on an Android mobile; and it provides a library called MeetAndroid used in Arduino working environment. The baud rate could set to 57600 or 115200 depending on the Arduino board. The Arduino BT board we used in this project works with 115200 baud.

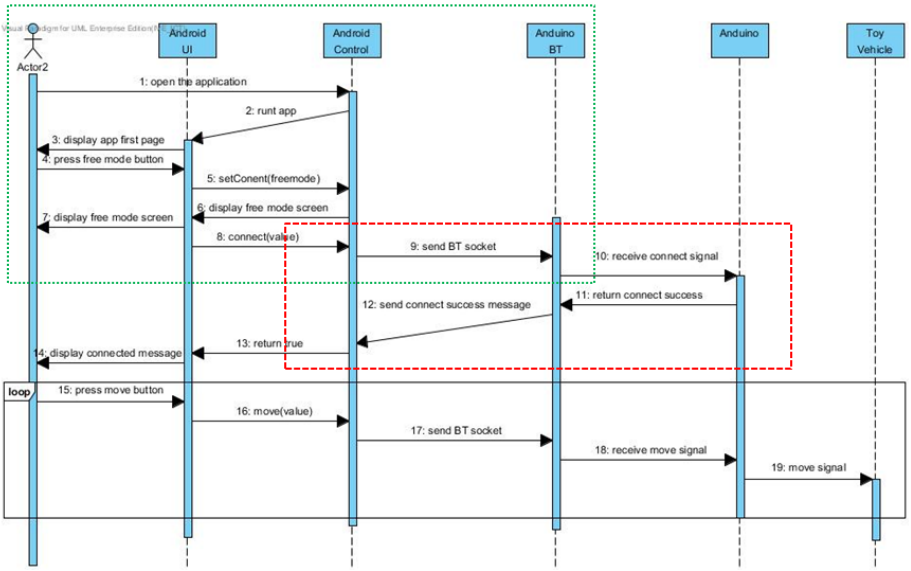
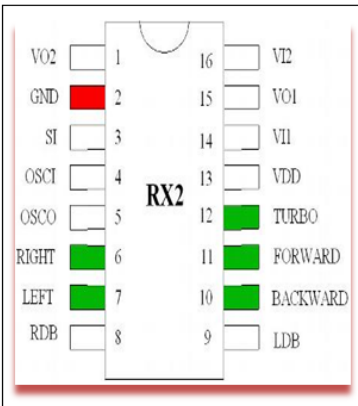


Fig. 3. Sequence Diagram of Android Motion Controls for Arduino Bluetooth Game

Amarino is embedded in Android for bluetooth network communication, which is perfect for controlling the limited movements in toy vehicle in our project. In order to control a toy vehicle the Android mobile needs to send signals remotely to the digital pins of the ADK board in Audrino as shown in Figure 4. Arduino circuit board has pin assignment for receiver to control physical motion, which has pin#12 for motion forward, 11 for backward, 10 for move left, 9 for move right, and 8 for reverse accordingly.

To define in which direction the toy vehicle should move, we have implemented a C++ program that utilizes the methods, MeetAndroid.receive() and MeetAndroid.send(), to receive and send data respectively [4]. The data, in turn, will be used to tilt the motor to move forward or backward in a certain direction on a horizontal floor ground. The Arduino program sketch shown in Figure 4 is responsible for receiving the remote control data and controlling the motor to move to the correct position and direction. The assumption of baud rate 9600 per second (bps) for the serial port, in facts, is acceptable to most the devices.



Pin Assignment in Receiver

```
#include <MeetAndroid.h>

MeetAndroid meetAndroid;
int photocellPin = 4;
int point = 0;
int forward = 12; // Pin 12 - Forward
int reverse = 11; // Pin 11 - Reverse
int left = 10;        // Pin 10 - Left
int right = 9;        // Pin 9  - Right
char val;     // Data from the serial port
```

```
void setup() {
  pinMode(forward, OUTPUT);
  pinMode(reverse, OUTPUT);
  Serial.begin(9600);        // at 9600bps
}

void loop() {
  meetAndroid.receive();
  digitalWrite(IR, HIGH);
  point = analogRead(photocellPin);
  meetAndroid.send(point);
  performCommand();
}

void performCommand() {
  if (Serial.available())  {
     val = Serial.read();
  }
  if (val == 'f') { // Forward
     go_forward();
   } else if (val == 'z') {     // Stop Forward
     stop_go_forward();
   } ...
}

void go_forward() {        // Forward
  digitalWrite(forward, LOW);
}

void stop_go_forward() { //StopForward
  digitalWrite(forward, HIGH);
}
```

Fig. 4. Game Motion Pin Assignment in Receiver using MeetAndroid Library in Amarino

Amarino in Android helps to send data to the Arduino board [4]. Android has an inner class ArduinoReceiver, a subclass of BoardcastReciever. An Android activity which has been developed to receive and handle motion data sent remotely by the Arduino circuit board. The way that ArduinoReceiver does to retrieve the data added to the intent with String data type. After assured the data type of the object in coding, the data will be retrieved and placed in a TextView field called mValueTV on main layout xml of Android activity.

4 Concluding Remark

In this paper, we presented an experimental Android Bluetooth-Based game development in mobile computing and network telecommunication. The Amarino has been employed as a support tool for bluetooth connection between Android mobile game players and the Arduino embedded toy vehicle. We demonstrated that bluetooth is a good candidate in wireless networking technology to support distributed applications in a mobile remote control game in a distributed system environment. Android mobile is programmed to be a remote controller providing services to send sockets via a bluetooth network to a toy vehicle embedded with Arduino microprocessor in order to control its motions in directions along its pathway. Our investigation focuses on an extended form of microprocessor computing which game software developers utilize to develop remote control games with multi-players. We call this study an experimental wireless network computing application in which the Arduino embedded in the toy car can sense the color pattern changes with infrared along its pathway and instantaneously send the data via bluetooth piconet to the wireless connected Android mobile device for processing. In this paper, we present the overall architecture and discuss, in detail, the implementation steps taken to create the Arduino and Android based remote control context-aware game. We prepare the client and server codes in ubiquitous computing, providing adaptive routines to handle connection information requests in telecommunication, logging and context formatting and delivery for throughput and context-triggered actions.

References

1. Lai, A.S.Y.: Mobile Bluetooth-Based Multi-Player Game Development in Ubiquitous Computing. Journal of Computational Information Systems (JCIS) 6(14) (December 2010)
2. Gu, T., Pung, H.K., Shang, D.Q.: A service-oriented middleware for building context-aware services. Journal of Network and Computer Applications 20, 1–18 (2005)
3. Bohmer, M.: Beginning Android ADK with Arduino. Apress (December 2012)
4. Arduino talk with Android using Amarino (2011), `http://www.amarino-toolkit.net`
5. Lai, A.S.Y.: Meta-based Distributed Computing Framework for Distributed Computing System. In: Embedded and Multimedia Computing, EMC 2011. LNEE, vol. 102, pp. 405–414. Springer, Heidelberg (2011)
6. Lai, A.S.Y.: Meta Level Component-Based Framework for Distributed Computing Application. PhD's Dissertation, Computer Science, Engineering and Applied Science, Aston University, Birmingham, UK (2008)
7. Cano, J., Manzoni, P., Toh, C.K.: UbiqMuseum: A Bluetooth and Java Based Context-Aware System for Ubiquitous Computing. Journal of Wireless Personal Communications (2006)
8. Schilit, B., Adams, N., Want, R.: Context-Aware Computing Applications. In: Proceedings of Mobile Computing Systems and Applications, Santa Cruz, CA, pp. 85–90. IEEE Computer Society (December 1994)

Performance Analysis for PUF Data Using Fuzzy Extractor

Hyunho Kang[1], Yohei Hori[2], Toshihiro Katashita[2], Manabu Hagiwara[3], and Keiichi Iwamura[1]

[1] Dept. of Electrical Engineering, Tokyo University of Science, 6-3-1 Niijuku, Katsushika-ku, Tokyo 125-8585, Japan
{kang,iwamura}@ee.kagu.tus.ac.jp
[2] Research Institute for Secure Systems (RISEC), National Institute of Advanced Industrial Science and Technology (AIST), Central 2, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan
{hori,t-katashita}@aist.go.jp
[3] Dept. of Mathematics and Informatics, Faculty of Science, Chiba University, 1-33 Yayoi-cho, Inage, Chiba 263-0022, Japan
hagiwara@math.s.chiba-u.ac.jp

Abstract. The extraction of a stable signal from noisy data is very useful in applications that aim to combine it with a cryptographic key. An approach based on an error correcting code was proposed by Dodis et al., which is known as a fuzzy extractor. Physical unclonable functions (PUFs) generate device-specific data streams, although PUFs are noisy functions. In this paper, we describe a method for preparing a PUF key during fuzzy extractor implementation. The experimental results showed that all possible combinations of input message length and the number of correctable errors were tested for a BCH code with codeword length N, which was the length of the PUF responses.

Keywords: Arbiter PUF, Fuzzy Extractor, Physical Unclonable Function (PUF).

1 Introduction

The fuzzy extractor scheme defined in [1] can derive reliable bit strings from noisy data and it is a very useful approach for applications that aim to combine it with a cryptographic key. Physical unclonable functions (PUFs) generate device-specific data streams by using manufacturing variations of each LSI. High security authentication is possible during secret key generation using PUFs, if a system requires the best extraction scheme.

Experimental studies of fuzzy extractors [2][3][4] have received considerable attention since this approach was proposed by Dodis et al., 2004. However, it appears to be difficult to implement initially. Thus, a test and an illustration of how to produce a key may facilitate a better understanding of a practical fuzzy extractor. We report the results of some implementation examples using

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*, Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_36,

PUF data and we present a detailed implementation diagram. We hope that this paper will help users to understand the implementation of this scheme.

The remainder of this paper is organized as follows. Section 2 presents the implementation approach and a detailed diagram. The experimental results and conclusions are provided in Sections 3 and 4, respectively.

2 Implementation Approach

The key result provided by Dodis et al. [1] demonstrated that fuzzy extractors can be built from secure sketches using strong randomness extractors, as shown in Fig. 1. During the generation procedure, **SS** is applied to noisy data **w** and a random input message **r** is used to obtain **s**, while a strong extractor **Ext** with randomness of **x** to **w** is used to obtain almost uniform randomness **R**. The pair **(x, s)** is stored as helper data, **P**. During the reproduction procedure, the helper data is used to regenerate the output **R** from new noisy data **w′** based on **Rec(w′,s)** and **Ext(w,x)**.

During the implementation of this scheme, there are two important considerations, i.e., a combination of information reconciliation and privacy amplification. The information reconciliation step guarantees the elimination of noise from the measured noisy data. Privacy amplification guarantees the uniform distribution of the derived key bits. A BCH code and SHA-256 hash function were used to address these two basic requirements.

In this paper, I examine the fuzzy extractor performance of our Arbiter PUF by presenting results for all possible combinations of the message length and the number of correctable errors for a BCH code with a fixed codeword length (i.e., 127, 255, and 511).

Figure 2 shows the implementation diagram for Fuzzy extractors using the BCH code and hash function (N = 255). This diagram helps us to understand how the system operates.

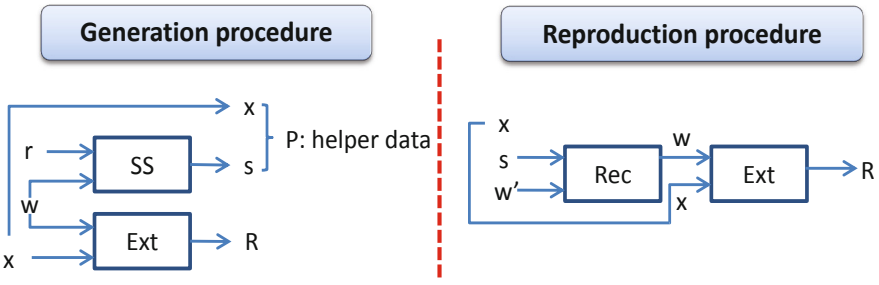


Fig. 1. Typical scheme for a fuzzy extractor [1]

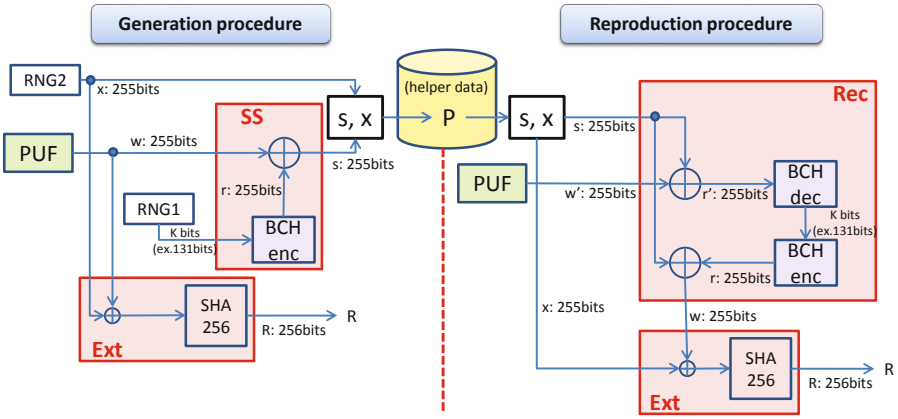


Fig. 2. Implementation diagram for our fuzzy extractor (N = 255)

3 Experimental Results

The FPGA used in this experiment was a Xilinx Virtex–5LX (xc5vlx30–ffg324), which operated on a SASEBO–GII evaluation board [5][6]. All of the performance results in this paper were generated using MATLAB. For example, the BCH code implementation is readily available in the Communications System Toolbox in MATLAB.

3.1 Performance of the Two PUFs Tested

In this section, we discuss the performance of the two Arbiter PUFs that we tested, before moving onto the fuzzy extractor performance. Reliability and uniqueness are commonly used to evaluate the performance of PUFs. In this study, we selected challenge response pair data for 100 test iterations using 500

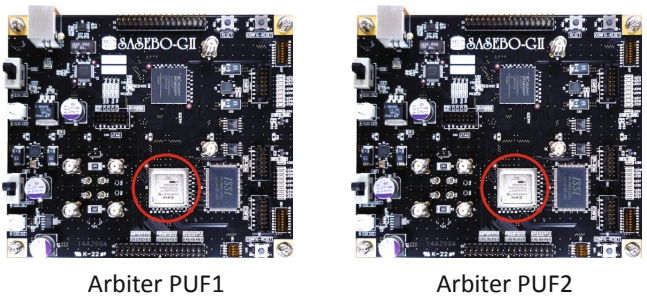


Fig. 3. The two PUFs tested on the SASEBO-GII

IDs from each tested Arbiter PUF (two different FPGAs were implemented using the same circuit structure, as shown in Fig. 3).

Reliability. A comparison of the *SC(Same Challenge) Intra* and *DC(Different Challenge) Intra* of PUF1 in Figure 4 shows that the HD of the PUF was divided into two distinct classes, depending on the properties of the challenge. The peaks of the two sets of histograms were clearly separated, which indicated that there were no errors in terms of the false acceptance rate and false rejection rate.

Uniqueness. We tested the uniqueness of the two Arbiter PUFs by finding all of the *SC Intra* and *SC Inter* HDs. As shown in Figure 5, there were no identification errors because there were no overlaps in the Intra and Inter SC distributions. To maintain stable security, it is desirable to separate the two distributions adequately. Thus, three types of response length were used to test the performance variation.

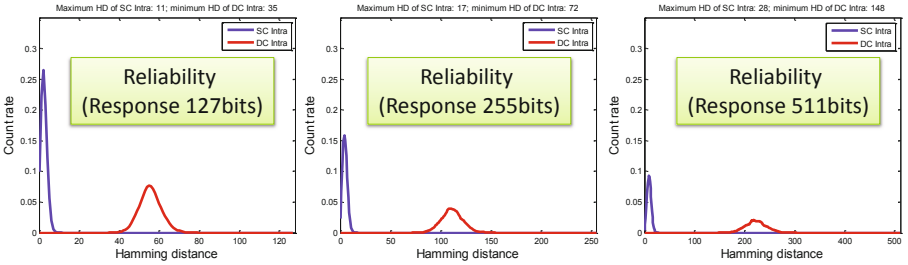


Fig. 4. Reliability

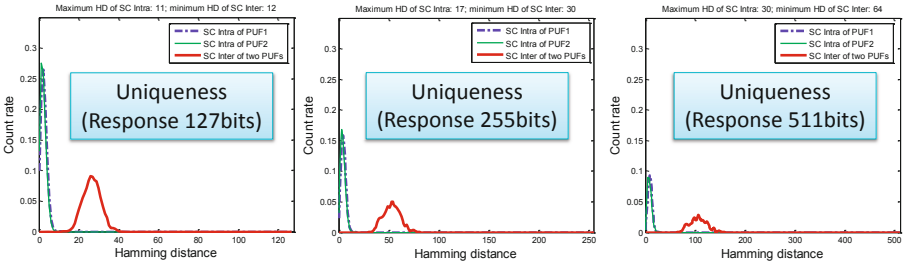


Fig. 5. Uniqueness

As you must realize, PUFs [7] do not provide exactly the same output each time. However, for identification purposes, errors may be no problem as long as the distance between the IDs is large enough (as we have seen in the test).

3.2 Fuzzy Extractor Performance of Our Arbiter PUFs

As mentioned earlier, all possible combinations (see Table 1∼3) were used as parameters of the BCH code in each response length to examine the fuzzy extractor performance of our Arbiter PUFs.

Figure 6∼7 show the Hamming distance between two extracted keys when the two tested PUFs were the same, which demonstrates the dependency of the number of correctable errors on the testing index. Figure 6 shows that the response errors in all tests were corrected from an index of 5. In Fig. 7, the errors were corrected from indices of 10 and 17, respectively.

(Note: the test index of the enrolled PUF response occurs first, as shown in Fig. 6∼7.)

Table 1. Number of correctable errors in the BCH code, for N = 127

index	N	K	t	index	N	K	t
1	127	120	1	10	127	57	11
2	127	113	2	11	127	50	13
3	127	106	3	12	127	43	14
4	127	99	4	13	127	36	15
5	127	92	5	14	127	29	21
6	127	85	6	15	127	22	23
7	127	78	7	16	127	15	27
8	127	71	9	17	127	8	31
9	127	64	10				

Table 2. Number of correctable errors in the BCH code, for N = 255

index	N	K	t	index	N	K	t
1	255	247	1	18	255	115	21
2	255	239	2	19	255	107	22
3	255	231	3	20	255	99	23
4	255	223	4	21	255	91	25
5	255	215	5	22	255	87	26
6	255	207	6	23	255	79	27
7	255	199	7	24	255	71	29
8	255	191	8	25	255	63	30
9	255	187	9	26	255	55	31
10	255	179	10	27	255	47	42
11	255	171	11	28	255	45	43
12	255	163	12	29	255	37	45
13	255	155	13	30	255	29	47
14	255	147	14	31	255	21	55
15	255	139	15	32	255	13	59
16	255	131	18	33	255	9	63
17	255	123	19				

Table 3. Number of correctable errors in the BCH code, for N = 511

index	N	K	t	index	N	K	t	index	N	K	t	index	N	K	t
1	511	502	1	16	511	367	17	31	511	238	37	46	511	103	61
2	511	493	2	17	511	358	18	32	511	229	38	47	511	94	62
3	511	484	3	18	511	349	19	33	511	220	39	48	511	85	63
4	511	475	4	19	511	340	20	34	511	211	41	49	511	76	85
5	511	466	5	20	511	331	21	35	511	202	42	50	511	67	87
6	511	457	6	21	511	322	22	36	511	193	43	51	511	58	91
7	511	448	7	22	511	313	23	37	511	184	45	52	511	49	93
8	511	439	8	23	511	304	25	38	511	175	46	53	511	40	95
9	511	430	9	24	511	295	26	39	511	166	47	54	511	31	109
10	511	421	10	25	511	286	27	40	511	157	51	55	511	28	111
11	511	412	11	26	511	277	28	41	511	148	53	56	511	19	119
12	511	403	12	27	511	268	29	42	511	139	54	57	511	10	127
13	511	394	13	28	511	259	30	43	511	130	55				
14	511	385	14	29	511	250	31	44	511	121	58				
15	511	376	15	30	511	241	36	45	511	112	59				

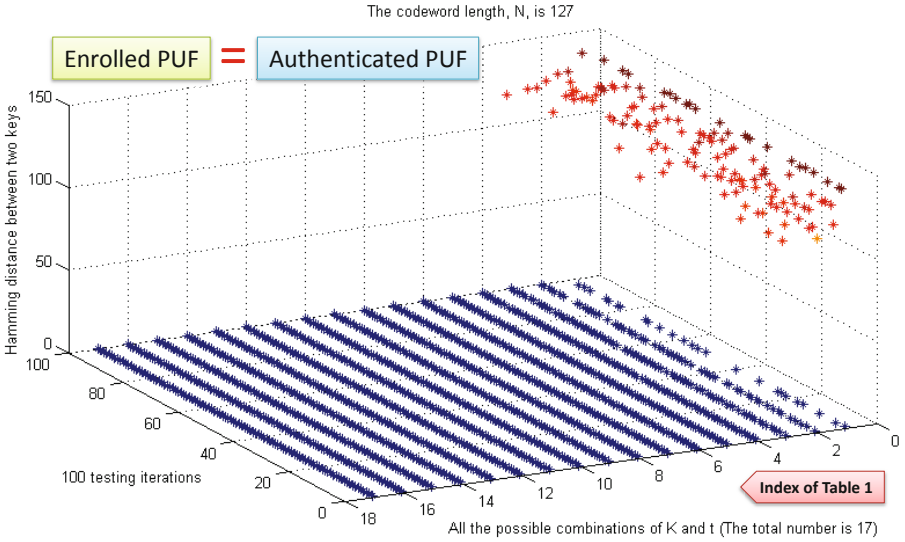


Fig. 6. Hamming distance between two extracted keys (N=127)

Figure 8∼9 show the Hamming distance between two extracted keys when two different PUFs were tested. Figure 8 shows that the response errors in all tests were corrected from an index of 17 because of the authentication of different Arbiter PUFs. In Fig. 9, the errors were corrected from indices of 33 and 56, respectively.

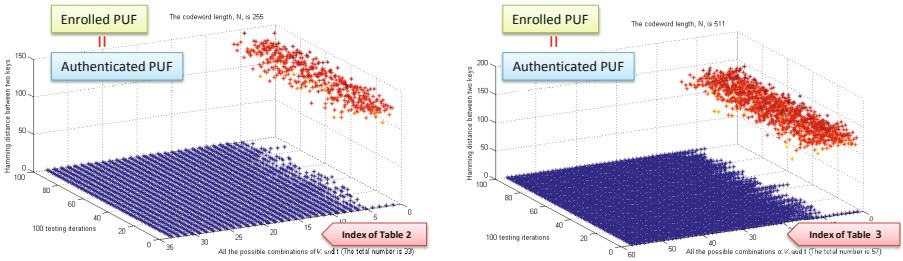


Fig. 7. Hamming distance between two extracted keys (left) N = 255; (right) N = 511

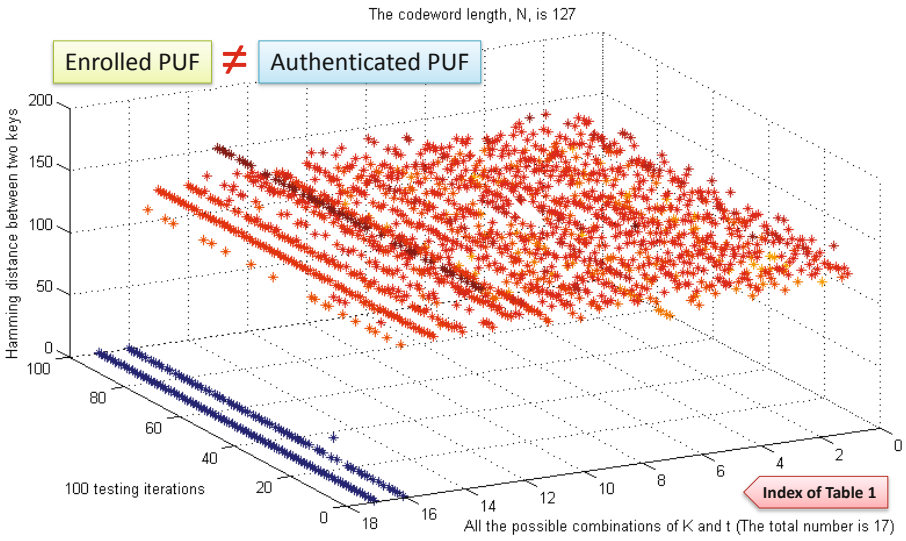


Fig. 8. Hamming distance between two extracted keys (N = 127)

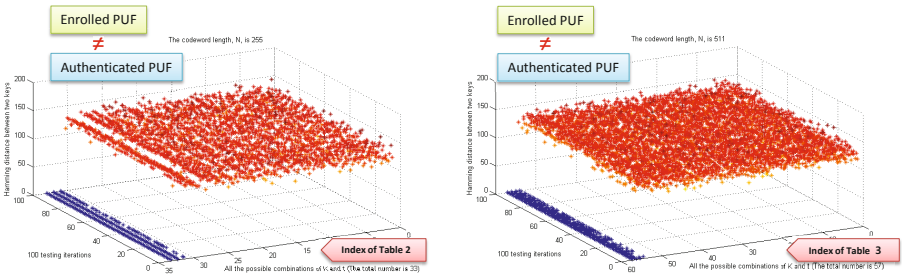


Fig. 9. Hamming distance between two extracted keys (left) N = 255; (right) N = 511

4 Conclusion

Since PUFs always produce bit errors at their output, error correction is one of the major topics concerning PUFs. In this paper, we showed the results of some implementation examples using our Arbiter PUF data and we present a detailed implementation diagram. This study must be helpful to facilitate the understanding of fuzzy extractor implementation.

References

1. Dodis, Y., Ostrovsky, R., Reyzin, L., Smith, A.: Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data (A preliminary version of this paper appeared in Eurocrypt 2004). SIAM J. Comput. 38(1), 97–139 (2008)
2. Bulens, P., Standaert, F.-X., Quisquater, J.-J.: How to strongly link data and its medium: the paper case. IET Inf. Secur. 4(3), 125–136 (2010)
3. Imamverdiev, Y.N., Sukhostat, L.V.: A Method for Cryptographic Key Generation from Fingerprints. Automatic Control and Computer Sciences 46(2), 66–75 (2012)
4. van der Leest, V., van der Sluis, E., Schrijen, G.-J., Tuyls, P., Handschuh, H.: Efficient Implementation of True Random Number Generator Based on SRAM PUFs. In: Naccache, D. (ed.) Quisquater Festschrift. LNCS, vol. 6805, pp. 300–318. Springer, Heidelberg (2012)
5. Satoh, A., Katashita, T., Sakane, H.: Secure implementation of cryptographic modules. Development of a standard evaluation environment for side channel attacks. Synthesiology-English Edition 3(1), 86–95 (2010)
6. The download site of Side-channel Attack Standard Evaluation BOard support file in AIST (National Institute of Advanced Industrial Science and Technology), `http://www.risec.aist.go.jp/project/sasebo/`
7. Bohm, C., Hofer, M.: Physical Unclonable Functions in Theory and Practice. Springer (2013)

High-Speed Block Cipher Algorithm Based on Hybrid Method

Bac Do Thi[1] and Minh Nguyen Hieu[2]

[1] University of Information and Communication Technology, Thai Nguyen, Viet Nam
dtbac@ictu.edu.vn
[2] Le Qui Don Technical University, Ha Noi, Viet Nam
hieuminhmta@ymail.com

Abstract. This paper proposes 3 different designs of the new 64-bit block cipher diagram. A new feature of the designs is the application of hybrid CSPN (Controlled Substitution Permutation network). Designs with particular advantages will make the selection more appropriate for each target of applications. However, design shall meet all the security requirements to protect applications against the well-known threats.

Keywords: hybrid CSPN, SDDO (Switchable Data Dependent Operation), Block cipher, key schedule.

1 Introduction

Information security is one of the essential requirements of the communication system to ensure the system safety; one of solutions to the system security shall be applied with ciphers. However, the selection of appropriate ciphers with targets of different services is a significant factor for success of the service. Especially, such factor is required and evaluated carefully to apply to the high speed communication applications factors. Proposed several algorithms have been orientated the high-speed communications networks such as CIKS-1 [3], DDP-64 [3], Cobra-H64 [3], Cobra-S128 [4], etc. The previous studies was applied the homogeneous CSPN-based design, i.e., based on only one type CE (Controlled Element). However, creating hybrids CSPN from two type combined CE for use in cryptographic applications is a solution to be considered flexible and suitable for many different applications and support to promote the strong points of CEs in each specific design. This paper proposed a new block cipher called BM123-64. It was developed on the basis of different SDDOs, in which these operators are built from heterogeneous or hybrid CSPNs.

This paper is structured as follows: section 2 presents the design model of heterogeneous and hybrid CSPNs; section 3 presents the new BM123-64 block cipher; section 4 presents the recommendations, forecasts and conclusions.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_37, © Springer-Verlag Berlin Heidelberg 2014

2 CSPN Design Used in Cryptographic

The general structure of CSPN was described in detail by Nikolay A. Moldovyan in [3] and is applied to construct the DDOs (Data Dependent Operation), SDDOs operators applied in many algorithms. Basically, both of CSPN types have the following common characteristics: only one CE in the diagram; multiple layers between which fixed permutations are interposed. According to [3], CSPN deems to be homogeneous. However, for the purpose of maintaining the generality of this design, a modification is proposed based on the combination of two separate CE called the hybrid CSPN model. Specifically, if the odd numbered layers have used one CE, the even numbered layers will be used with the other CE. An illustration of hybrid CSPN used in BM123-64 algorithm shall be shown below, $F'_{16/64}$ hybrid operators (Fig. 1). Design of hybrid $F'_{16/64}$ includes 4 layers, in which logical function of elements $F_{2/2}$ selected from layer 1 and 3 shall be represented by the equation (1, 2, 3), while of element $F'_{2/2}$ selected from 2, 4 layers shall be represented by equations (4, 5, 6) (see section 3). The permutation I_1, I_2 are described in Table 1.

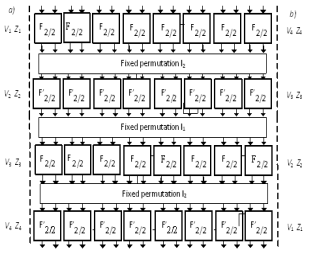


Fig. 1. Hybrid structure of $F'_{16/64}$, $F'^{-1}_{16/64}$

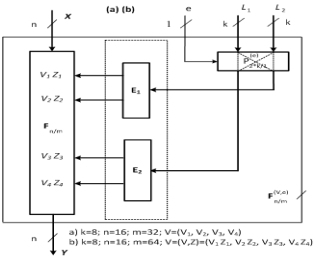


Fig. 2. Structure of $F^{(V,e)}_{16/32}$(a) and $F^{(V,e)}_{16/64}$ (b)

3 Design of Algorithm BM123-64

BM123-64 block cipher will be proposed with size of 64-bit block, key length of 256 bits. It includes 8 transformable rounds which use $\text{Crypt}^{(e)}$ transformations as shown in Fig.3. The difference of the proposed algorithm from others is that, the algorithm is

built on the basis of CSPNs, SDDOs combined with the S_{4x4}, $S^{-1}{}_{4x4}$ boxes and fixed permutations I, I_1 (Table 1). In the propose algorithms, $F_{n/m}^{(V,e)}$ SDDOs operators shall be applied. SDDOs are built on the basis of homogenous or hybrid CSPNs after embedding of SCO (Operator Controlled Switchable). The steps of BM123-64 are described as follows:

1. For $j = 1$ to 7 do: $\{(L, R) \leftarrow \mathbf{Crypt}^{(e)}(L, R, U_j, Q_j); (R, L) \leftarrow (L, R)\}$.
2. $(L, R) \leftarrow \mathbf{Crypt}^{(e)}(L, R, U_8, Q_8)$.
3. $\{(L, R) \leftarrow (L \oplus U_9, R \oplus Q_9); (L, R) \leftarrow (L, R)\}$.

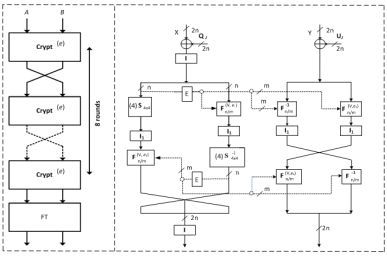


Fig. 3. Structure round transformation $\text{Crypt}^{(e)}$ of BM123-64

Table 1. Description of fix permutations and E expanded operator in the algorithm

I	(1)(2,18)(3)(4,20)(5)(6,22)(7)(8,24)(9)(10,26)(11)(12,28)(13)(14,30)(15)(16,32)(17)(18,2)(19)(20,4)(21)(22,6)(23)(24,8)(25)(26,10)(27)(28,12)(29)(30,14)31)(32,16)
I_1	(1)(2,5)(3,9)(4,13)(5,2)(6)(7,10)(8,14)(9,3)(10,7)(11)(12,15)(13,4)(14,8)(15,12) (16)
I_2	(1)(2,3)(3,2)(4)(5)(6,7)(7,6)(8)(9)(10,11)(11,10)(12)(13)(14,15)(15,14) (16)
E	$E(X)=(X, X^{<<<4}, X^{<<<8}, X^{<<<12})$.

In proposed algorithms the switchable bits e_i (i=1..4) depend on bit e ($e \in \{0,1\}$). Defining encryption (e=0) and decryption (e=1) mode and e_i is determined as follows: $e_1 = e \oplus e'_1$, $e_2 = e \oplus e'_2$, $e_3 = e \oplus e'_3$, $e_4 = e \oplus e'_4$ (see Table 2).

Table 2. The keys cheduling and lists the switch bits in BM123-64

No. rounds *j*	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**FT**
$Q_j =$	K_1	K_2	K_5	K_7	K_3	K_6	K_8	K_4	K_1
$U_j =$	K_3	K_4	K_8	K_6	K_2	K_7	K_5	K_2	K_3
$e'_1 =$	1	1	1	0	0	1	1	0	-
$e'_2 =$	0	1	1	0	1	1	1	1	-
$e'_3 =$	0	0	0	1	1	0	0	1	-
$e'_4 =$	1	0	0	1	0	0	0	0	-

In each round transformation we use only one 32-bit subkey combined with both the left and the right data subblocks. This algorithm enables both the encryption and decryption processes to be used the same algorithm. Besides, it is shown that both the encryption and decryption processes use the same key scheduling and transformation encryption and decryption mode only needing transformation of the controlling bit (Table 2). For the purpose of larger encryption space and algorithm and selection of the more appropriate algorithm with application's target, three different designs of $F_{n/m}^{(V,e)}$ operator shall be proposed the design diagram of the Crypt$^{(e)}$ shown in Fig. 2.

Case 1. In transformations Crypt$^{(e)}$ applied $F_{16/64}^{(V,e)}$ that is built from homogenous CSPNs. In 2208 [3], among invertible $F_{2/2}$ elements suitable for encrypted and highest non-linear application (NL= 4), selected $F_{2/2}$ shall be described by following logical functions:

$$y_1 = vzx_1 \oplus vz \oplus vx_1 \oplus vx_2 \oplus v \oplus z \oplus x_1 \oplus 1; \quad \text{NL}(y_1) = 4; \quad (1)$$

$$y_2 = vzx_2 \oplus vz \oplus vx_1 \oplus vx_2 \oplus zx_1 \oplus x_2 \oplus v \oplus z \oplus 1; \quad \text{NL}(y_2) = 4; \quad (2)$$

$$y_3 = vzx_1 \oplus vzx_2 \oplus zx_1 \oplus x_1 \oplus x_2; \quad \text{NL}(y_3) = 4, \quad (3)$$

Also according to [4], the corresponding differential values of the element $F_{2/2}$ are listed in Table 3. The biggest advantage shall be the highest linear logical functions, which give support to peak disturbance standard of algorithms. SDDO $F_{16/64}^{(V,e)}$ (Fig. 2) designed on the basis of $F_{2/2}$ will comply with earlier prototype proposed in [5]. This design is also applied with the operator $F_{16/64}$, built by operators of homogeneous CSPN. To control the distribution of the controlling vector in $F_{16/64}^{(V,e)}$ we use permutations $P_{2/1}$ which are connected parallel. The expansion of bits in E_1 and E_2 is performed as follows: 8 bits is an input of the extension block E_1 (or E_2) then the controlling vector $(V, Z) = (V1, Z1, V2, Z2, V3, Z3, V4, Z4)$ uses the switchable controlled block $F_{16/64}$ created as follows: $V_1 = L_1, V_2 = L_1^{<<<4}, V_3 = L_2^{<<<4}, V_4 = L_2$; $Z_1 = L_1^{<<<2}, Z_2 = L_1^{<<<6}, Z_3 = L_2^{<<<6}, Z_4 = L_2^{<<<2}$.

Table 3. Probabilities $\boldsymbol{Pr(ijk) = Pr(\Delta_i^Y/\Delta_j^X, \Delta_k^Y)}$ of DC of $F_{2/2}$ and $F'_{2/2}$ element

i j k	001	002	011	101	110	120	002	102	201	202	
$F_{2/2}$	0.25	0.125	0.1875	0.375	0.75	0.5	0.125	0.5	0.375	0.375	
i j k	110	210	001	011	021	222	112	221	220	101	121
$F'_{2/2}$	1	0	0	0.5	0.5	1	1	1	1	0.5	0.5

Case 2. In Crypt$^{(e)}$ transformations applied with $F_{16/64}^{(V,e)}$ from hybrid CSPN. In this design, two different CE elements, rather than one element $F_{2/2}$ shall be used for CSPN design. For the generality of the design, first CE shall be called $F_{2/2}$ and other shall be called $F'_{2/2}$, in which, the selected first CE sill is $F_{2/2}$ as case 1, and $F'_{2/2}$ shall be describe via logical functions as follows:

$$y_1 = vzx_1 \oplus vzx_2 \oplus vx_1 \oplus vx_2 \oplus zx_1 \oplus zx_2 \oplus z \oplus v \oplus x_2; \quad \mathrm{NL}(y_1) = 2; \tag{4}$$

$$y_2 = vzx_1 \oplus vzx_2 \oplus vz \oplus vx_1 \oplus vx_2 \oplus zx_1 \oplus zx_2 \oplus x_1; \quad \mathrm{NL}(y_2) = 2; \tag{5}$$

$$y_3 = vz \oplus v \oplus z \oplus x_1 \oplus x_2; \quad \mathrm{NL}(y_3) = 4. \tag{6}$$

For, respective differential value of element $F'_{2/2}$ shall be listed in Table 3. In comparison with case 1 above, logical functions of the selected $F'_{2/2}$ will have lower linearity but higher differential value (Table 3), or better avalanche effect than case 1, i.e., stronger the prevention of cryptanalysis threads. The design of hybrid $F'_{16/64}$ will be shown in Fig. 1, SDDO model in this case shown in Fig.2b.

Case 3. Transformations $\mathrm{Crypt}^{(e)}$ applied with $F_{16/32}^{(V,e)}$ and $F_{16/64}^{(V,e)}$ from homogenous CSPNs ($F_{16/32}^{(V,e)}$ applied with CE of $F_{2/1}$, $F_{16/64}^{(V,e)}$ applied CE of $F_{2/2}$). In this design, a hybrid feature different from design described in the case 2 is application of 2 CEs belonging to the same class. Unlike case 1, SDDOs left and right branch of the diagram are designed from a CE class $F_{2/2}$. Left branch of the diagram in Figure 3 shall be designed with CSPN from CE of class $F_{2/1}$, and right branch of the diagram shall be designed with CE of class $F_{2/2}$ as described in case 1. Thus, the design 3 has advantage of the reduction of the resource cost which at 32 block/round of $F_{2/2}$. Selected $F_{2/2}$ describe via logical functions as follows: $y_1 = vx_2 \oplus x_2 \oplus x_1 \oplus v \oplus 1$; $y_2 = vx_1 \oplus x_2$. Similar with designed SDDO $F_{16/64}^{(V,e)}$ as above, $F_{16/32}^{(V,e)}$ shall be built from $F_{16/32}$ basing on the embedded SCO by the parallel combination of 8 blocks $P_{2/1}$. The expansion of bits in E_1 and E_2 is performed as follows: 8 bits is an input of the extension block $E1$ (or $E2$) then the controlling vector $V = (V1, V2, V3, V4)$ uses the switchable controlled block $F_{16/64}$ created as follows: $V_1 = L_1, V_2 = L_1^{<<<4}$, $V_3 = L_2^{<<<4}, V_4 = L_2$.

4 Discussion and Conclusions

Using DDO to design the cipher shall be applied with simple above-mentioned key schedule in studies. At the same time, if any SCO is embedded for SDDO production from DDO, SDDO also eliminates the weak keys due to no complex key generation process, which has been proven in previous studies. In addition, DDO application in the encryption algorithm specified in [3] by several authors is not able to find randomly a known traces and types of attack. Moreover, the DDO-based algorithms and linear cryptanalysis are not more efficient than differential cryptanalysis [2,5]. On the basis of this argument, we present some discussion on the content related to the proposed design.

Evaluate According to NESSIE Statistical Standard. According to the results, after third testing rounds, (case 3 shall include 6 testing rounds), the algorithms meet the safety requirements of NESSIE.

Evaluate Linear Cryptanalysis (LC) and Differential Cryptanalysis (DC). The proposed designing in this paper had been developed based on the linearity of CSPNs analyzed in detail in [3]. In addition, according to the preliminary assessment of the other DDO-based encryption algorithms shown in [3, 4], the LC is not more efficient than DC. At the same time, according to the calculation results, the DC of the proposed algorithm using different SDDOs are better than differential value of several traditional algorithms (table 4), and best in case 2 due to the best differential of element $F_{2/2}$; the design diagram of the proposed algorithm was able to prevent the DC after round 4 (case 1, 2) or after round 5 (case 3).

Table 4. Security comparision of some cipher with BM123-64

Cipher	*Rmax*	DC		P(r)
		Min	P(Z)	
DDO-64	6	$(0, \Delta_1^R)$	$P(2) \approx 2^{-29}$	$\approx 2^{-87}$
COBRA-F64A	10	$(\Delta_1^R, 0)$	$P(2) < 2^{-20}$	$< 2^{-100}$
BM123-64 (case 1)	8	$(0, \Delta_1^R)$	$P(2) \approx 2^{-32,5}$	$\approx 2^{-137}$
BM123-64 (case 2)	8	$(0, \Delta_1^R)$	$P(2) \approx 2^{-34}$	$\approx 2^{-145}$
BM123-64 (case 3)	8	$(0, \Delta_1^R)$	$P(2) \approx 2^{-29}$	$\approx 2^{-121}$

Evaluate Performance During Algorithm Application to FPGA. According to result of the proposed algorithm application to FPGA and the comparison of other traditional algorithms (Table 5), the performance evaluation was reviewed basing on both formulas; first **IE = S/R** (S:speed, R:resource); second **IE = S/(R x F)** (F:Frequency). BM123-64 was implemented in Vitex-5. It is predicted that performance of proposed algorithm is much higher than other algorithm such as DDP-64, SPECTR-64, etc. Thus, in term of all mentioned above cases, the proposed algorithms meet safety requirement on prevention of known attacks.

Table 5. FPGA Synthesis Results of BM123-64 and Comparisons

Cipher	Block size	*Rmax*	*N*	R (#CLBs)	F (MHz)	S (Mbps)	Performance	
							S/R	S/R× F
BM123-64(1)	64	8	8	2066	166	10624	5.14	30.97
BM123-64 (2)	64	8	8	2066	165	10560	5.11	30.97
BM123-64 (3)	64	8	8	1810	162	10368	5.73	35.35
DDP-64	64	10	10	3440	95	6100	1.77	18.67
SPECTR-64	64	12	12	7021	83	5312	0.76	9.12

All main results include in this paper as follows:

1. Proposing hybrid CSPNs on the basis of development of homogenous CSPN. The proposal supports to produce CSPNs larger space for the development of cryptography applications of high-speed communication network.
4. Proposing BM123-64 algorithm with 3 different specific designs. The analytical results show that applications are capable of against the known attacks.
5. Estimating the costs of applications to FPGA, evaluating performance and comparing results with traditional efficient encryption algorithm.

References

1. Moldovyan, N.A., Moldovyan, A.A., Eremeev, M.A., Summerville, D.H.: Wireless networks security and cipher design based on data-dependent operations: Classification of the FPGA suitable controlled elements. In: Proceedings of the CCCT-2004, Austin Texas, USA, pp. 123–128 (2004)
2. Moldovyan, N.A., Moldovyan, A.A., Eremeev, M.A., Sklavos, N.: New class of Cryptographic Primitives and Cipher Design for Network Security. International Journal of Network Security 2, 114–125 (2006)
3. Moldovyan, N.A., Moldovyan, A.A.: Data-driven Ciphers for Fast Telecommunication Systems. Auerbach Publications Talor & Francis Group, New York (2008)
4. Goots, N.D., Moldovyan, A.A., Moldovyan, N.A.: Variable bit permutations: linear characteristics and pure VBP-based cipher. Comput. Sci. Moldova. 13, 84–109 (2005)
5. Minh, N.H., Bac, D.T., Duy, H.N.: New SDDO-Based Block Cipher for Wireless Sensor Network Security. International Journal of Computer Science and Network Security 10, 54–60 (2010)
6. Moldovyan, N.A.: On Cipher Design Based on Switchable Controlled operations. International Journal of Network Security 7, 404–415 (2008)

Decision Engine Based Access Router Discovery Scheme in IEEE 802.11

HwaMin Lee[1], DooSoon Park[1], DaeWon Lee[2], and SungJai Choi[3,*]

[1] Department of Computer Software Engineering, Soonchunhyang University, Asan, ChungNam, 336-745, Korea
{leehm,parkds}@sch.ac.kr
[2] Division of General Education, SeoKyeong University, Jeongneung 4-dong, Sungbuk-gu, Seoul 136-704, Korea
daelee@skuniv.ac.kr
[3] Department of Electronic Engineering, Gachon University, 1342 Seongnamdaero, Sujeong-gu, Seongnam-si, Gyenggi-do, Korea
csj0717@gachon.ac.kr

Abstract. In this paper, we focus on AR selection problem of mobile host (MH). A mobile host (MH) may have several available networks when entering a new area. The user of MH decides to use one of access router (AR) at AR list. However, the decision only with the subsystem identification (SSID) and signal strength could not provide appropriate connection to the MH. To provide appropriate connection, more information of AR is needed. In this paper, we extend prefix information option on router advertisement message to obtain AR's status information. Then, we proposed decision engine (DE) on the MH that analyzes ARs and decides the appropriate AR automatically by AR's status information. In experimental analysis, proposed AR discovery process has several advantages. For the MH, wireless connection period is increased, the power consumption is decreased, and the signaling overhead is reduced. For AR and router, the load balancing is provided and the network topology can also be more efficient.

Keywords: Discovery Process, Neighbor Discovery Protocol, Access Router, 802. 11, Decision Engine.

1 Introduction

With the development of mobile communications and Internet technology, there is a strong need to provide connectivity for roaming devices to communicate continuously with other devices on the 802.11 wireless networks. Depending on the point of attachment, the access to an 802.11 network can be achieved in two different modes. A MH can form spontaneous networks as a mobile router (MR) (ad-hoc mode) or it can get connected to an access router (AR) which is directly connected to a backbone

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*, Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_38,

(infrastructure mode) [1]. In both modes, mobility appears as the key benefit of 802.11, providing the users the possibility to move inside and between cells. When moving out of the range of its current point of attachment, an MH should quickly discover and attach to a new point of attachment to reconnect to the network. It is known as a handoff. The MH performs operations each modes. In infrastructure mode, it consists in finding a new AR. In ad-hoc mode, an MR may additionally need to discover new services, and eventually update routing states if multi-hops protocols are used.

The 802.11 is focused on layer 2 handoff [2,3,4,5,6]. When a MH starts up or enters into a new cell, it needs to discover its environment: radio frequencies, neighbor point of attachment, and available services. Generally, the MH may have several available AR. The user of MH decides to use one of AR only with the subsystem identification (SSID) and signal strength for user's internet connection [1].

In this paper, we focus on layer 3 handoff [7,8,9,10]. To provide appropriate connection, we use AR's status information: capacity, current load, depth of network hierarchy [8]. And we extend prefix information option on router advertisement message [7] to obtain AR's status information. To analysis reachable ARs and decide appropriate AR, we design decision engine (DE). By proposed AR discovery process, the handoff MH selects appropriate AR automatically. And, using discovery process, the load balance can be provided on ARs and the network topology can be more efficient.

The remaining of the paper is organized as follows. We present problem define and survey the related works in section 2. In section 3, the necessity of AR status information is presented. In section 4, we introduce optimal AR discovery algorithm. Then, section 5 shows simulation experiment. Finally, we conclude the article in section 6.

2 Related Work

Most of the related work of the 802.11 discovery process concerns the optimization of the scanning latency during a layer 2 handover, when a MH roams from one AP to another. One simply method to reduce the full scanning latency is to use selective scanning which allows to only scan a subnet of channels, instead of probing each of them. Another proposed method is focused on reducing the value of the scanning timers (MinCT and MaxCT) [2]. [3] fixed the potential best values for both timers presenting theoretical considerations and simulation results. The smooth handover [4] and the periodic scanning [5] methods are based on splitting the discovery phase into multiple sub-phases. The objective of this division is to allow a MH to alternate between data packet exchange and the scanning process. A MH builds a list of target APs maintaining only information of channel and SSID.

These methods are only focused on reducing latency to minimize disconnected time of MH. However, they are not sure that the MH connects an appropriate AP. To provide appropriate connection, more information is needed. The information could not be served by probe signal at layer 2. In order to collect information, we focus on neighbor discovery protocol at layer 3. The router advertisement message is one of message format from neighbor discovery protocol in IPv6 [7].

To collect more information, we focus on router advertisement message. However, there are not enough fields on reserved field in the router advertisement message to collect more information. So, we focus on prefix information that must use in hierarchy architecture in IPv6.

3 Extended Prefix Information Option

To access a new AR in an L2 handoff, the AR provides only SSID and authentication. However, the information provided by L2 is insufficient to determine a suitable AR. In this paper, we focus on an L3 handoff for suitable AR selection. There are four information elements needed to determine the suitable AR. First, the status of the AR is necessary. Second, the maximum capacity of the AR is necessary. Third, the expected AR load is necessary; because several MHs use the same AR in a cell, the network bandwidth is limited. The MH should find a free AR. The last point is the depth of the network hierarch, which represents the logical location with respect to a border gateway in a subnet. To prevent frequent handoffs and receive faster transmission in a subnet, the MH should connect with a higher-level AP in a subnet hierarchy. The additional attributes in the extended router advertisement message are as follows: Status: status of the AR(stationary/portable) , Cap: maximum capacity of AR, Load: current load of AR, Depth: depth of network hierarchy. Fig. 1 shows a format of the extended prefix information option. However, we use the minimum bits in the reserved field because the signaling overhead on wired/wireless links is an important issue.

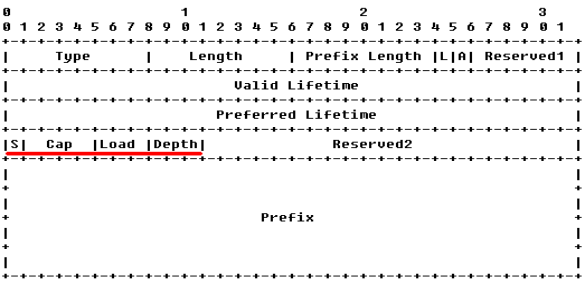


Fig. 1. Extended prefix information option format

4 AR Discovery Process

In section 4, to decide the appropriate AR, we proposed decision engine (DE) that analyzes router advertisement message as we shown in section 3 and decides appropriate AR. Proposed AR discovery process consists of three parts. First part is that information is received by router advertisement message and extracts them by neighbor list. Then, second part predicates all information for decision making. The decision making part is divided two parts. One is active state for the MH that moves

frequently, the other is idle state for the MH that moves rarely. Table 1 shows decision making algorithm of AR discovery process in the DE. The last part of process is when the MH loses its connection. The MH broadcasts a neighbor solicitation message and then connects a new AR by decision making.

Table 1. The Decision Making Algorithm of AR Discovery Process

```
// For Active state.
// Priority: status > bandwidth > signal > depth > load
predicator Idle : BinaryComparator {
    bool operator () ( NeighborItem l, NeighborItem r )
    {
        if (l.status < r.status)              return true;
        if (l.status > r.status)              return false;
        if(l.bandwidth>r.bandwidth)           return true;
        if (l.bandwidth < r.bandwidth)        return false;
        if (l.signal > r.signal)              return true;
        if (l.signal < r.signal)              return false;
        if (l.depth < r.depth)                return true;
        if (l.depth > r.depth)                return false;
        if (l.loadRatio < r.loadRatio)        return true;
        if (l.loadRatio > r.loadRatio)        return false;
        return false;
    }
}
```

```
// For Idle state.
// Priority: status > depth > signal > load > bandwidth
predicator Active : BinaryComparator {
    bool operator () ( NeighborItem l, NeighborItem r )
    {
        if (l.status < r.status)              return true;
        if (l.status > r.status)              return false;
        if (l.depth < r.depth)                return true;
        if (l.depth > r.depth)                return false;
        if (l.signal > r.signal)              return true;
        if (l.signal < r.signal)              return false;
        if (l.loadRatio < r.loadRatio)        return true;
        if (l.loadRatio > r.loadRatio)        return false;
        if(l.bandwidth > r.bandwidth)         return true;
        if (l.bandwidth < r.bandwidth)        return false;
        return false;
    }
}
```

5 Performance Evaluation

5.1 Simulation Environment

We evaluated the performance of proposed discovery process with the optimal AR discovery algorithm and compared it with the 802.11 discovery process. In order to evaluate the performance, we have developed a simulator by JAVA. The DE is incorporated in the simulator. The simulator has the capability to emulate the random mobility and dynamic traffic pattern of MH. Our simulation based studies provide a means to conduct experiments and compare the efficiency of various strategies. Fig. 2 shows the simulation environment used to evaluate the proposed discovery process.

There are three subnets. Each subnet has a border router (BR) to connect the Internet. Subnet A is composed of three hierarchy. BRA is consists of an AR and six routers. Each router on subnet A has six ARs. Also, subnet B is composed of three hierarchy. BRB is consists of an AR and two routers. Each router on subnet B has three ARs. Subnet C is composed of two hierarchy. BRC is consists of an AR and a router. A router on subnet C has three ARs. On this architecture, we generate 2 kind

of mobile devices. One is MR that changes its point of attachment by its random movement. The other is MH. In this simulation, 70 MRs and 150 MHs are generated. These mobile devices are randomly located at initial state and each of them has random mobility. Also we consider traffic pattern of mobile user. The 3% of MHs are specified heavy user and the 10% of MHs are specified light user. The traffic pattern is randomly changed within ratio.

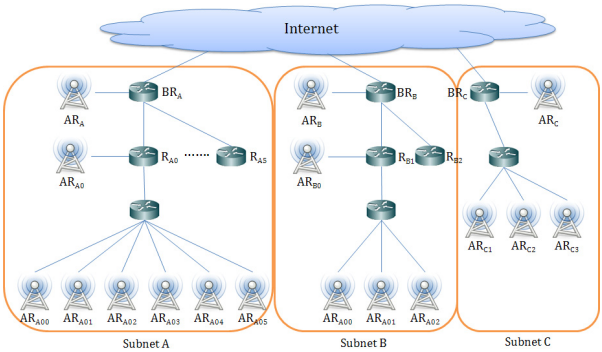


Fig. 2. Simulation Environment

5.2 Experimental Results and Analysis

In our first experiment, we compare link down between discovery process with the DE and 802.11 discovery process. Fig. 3 shows the number of link down until 20 minutes. Both discovery processes have little difference on count of link down. Total count of link down in 20 min: Proposed = 35690, 802.11 = 37540 (2% increased).

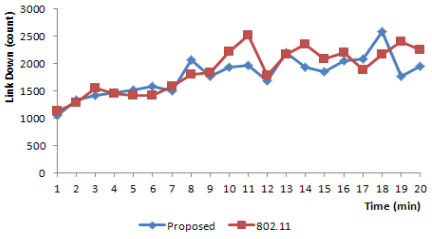


Fig. 3. Number of link down until 20 min

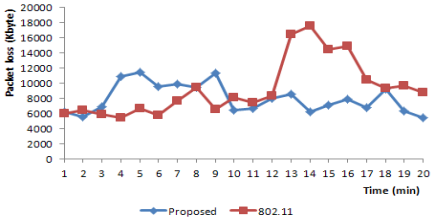


Fig. 4. Packet loss until 20 min

Fig. 4 shows the packet loss until 20 minutes. Even both discovery processes perform same probe scanning method in 802.11, proposed discovery process has better result than 802.11 discovery process. Total packet loss in 20 min: Proposed = 159847Kbytes, 802.11 = 185475Kbytes (3.8% increased). Almost 30 Mbytes data are lost in 20 min. It is the second significant result.

Fig. 5 shows the average ratio of router bandwidth until 20 minutes. Because we need more specific experiment, the time unit is changed min to 1/25 tick (25 tick = 1 sec, eg. 25 tick x 60 x 20 = 1,200). In figure 12, proposed discovery process is evenly distributed except initial state. In 802.11, the load of router is rapidly increased 1 time between 100-200, 1 time between 500-600, and 4times between 600 and 700. Through the comparison of both discovery processes at fig. 5, we probe the load balance of ARs and topological stability.

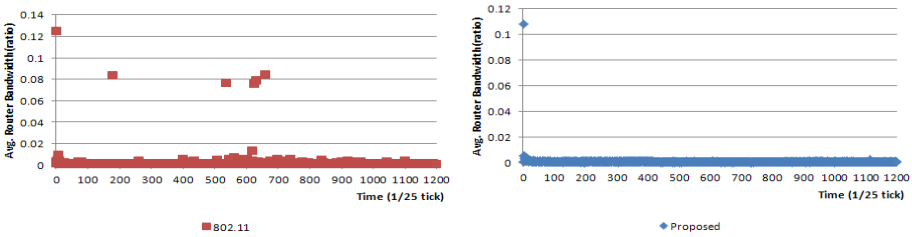


Fig. 5. Average ratio of router bandwidth until 20 min

6 Conclusion

In this paper, we focus on AR selection problem of mobile host when entering into a new cell. Therefore, we proposed decision engine to find appropriate AR on 802.11. The decision only with the AR's SSID and signal strength could not provide the appropriate connection to the MH. Choosing appropriate AR, we extend prefix information option on neighbor discovery protocol on layer 3 to obtain the AR's status information such as status, capacity, current load, and depth of network hierarchy. Also, we proposed the decision engine (DE) on MH and proposed the optimal AR discovery algorithm that analyzes ARs and decides the optimal AR automatically using AR's status information. Through the performance analysis, we probe that the load balance of AR and efficiency of network topology is presented. In this experiment, because we need huge computing performance, we simulate basic scanning method on L2. In future work, we should consider about reducing full scanning latency and compare with scanning methods on L2.

References

1. RFC 5416: Control and Provisioning of Wireless Access Points (CAPWAP) Protocol Binding for IEEE 802.11, `http://www.rfc-editor.org/rfc/rfc5416.txt`
2. Shin, S., Forte, A.S., Rawat, A.S., Schulzrinne, H.: Reducing MAC Layer Handoff Latency in IEEE 802.11 Wireless LANs. In: 2nd International Workshop on Mobile Management and Wireless Access Protocols, pp. 19–26 (2004)
3. Velayos, H., Karlsson, G.: Techniques to Reduce the IEEE 802.11b Handoff Time. In: 2004 IEEE International Conference on Communications, pp. 3844–3846 (2004)

4. Liao, Y., Gao, L.: Practical Schemes for Smooth MAC Layer Handoff in 802.11 Wireless Networks. In: 2006 International Symposium on World of Wireless, Mobile and Multimedia Networks, pp. 181–199 (2006)
5. Montavont, J., Montavont, N., Noel, T.: Enhanced Schemes for L2 Handover in IEEE 802.11 Networks and Their Evaluations. In: IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications, pp. 1429–1434 (2005)
6. Koutsopoulos, I., Tassiulas, L.: Joint Optimal Access Point Selection and Channel Assignment in Wireless Networks. IEEE/ACM Transactions on Networking 15(3), 521–532 (2007)
7. RFC 4861: Neighbor Discovery for IP version 6 (IPv6), http://www.rfc-editor.org/rfc/rfc4861.txt
8. RFC 5380: Hierarchical Mobile IPv6 Mobility Management (HMIPv6), http://www.rfc-editor.org/rfc/rfc5380.txt
9. Prahmkaew, S.: Performance Evaluation of Convergence Ad Hoc Networks. Journal of Convergence 1(1), 101–106 (2011)
10. Castignani, G., Arcia, A., Montavont, N.: A study of the discovery process in 802.11 networks. Presented at Mobile Computing and Communications Review (January 2011)

Development and Usability Assessment of Tablet-Based Synchronous Mobile Learning System

Jang Ho Lee

Dept. of Computer Eng., Hongik University, 72-1 Sangsu, Mapo, Seoul 121-791, South Korea
janghol@cs.hongik.ac.kr

Abstract. We present a tablet-based synchronous mobile learning system that allows remote students to join a distance learning session with their tablets in real time. The system enables students to watch the live video of an instructor as well as slide with annotation. It also allows students to ask questions with text during a lecture. We performed a usability assessment of the system by having a group of students test the prototype and conducting a survey. The result shows that the students feel that usefulness and ease of use of the system is above the normal level.

Keywords: Mobile learning system, Tablet, Usability.

1 Introduction

As smartphones and tablets have become widespread, the researchers‘ interest has shifted from traditional desktop-based distance learning system to mobile learning system running on those mobile devices that can provide as much computing power as old desktop computers[1] and enable students to participate in the lecture from anywhere with their mobile devices. Compared to the most of the existing mobile learning systems that are asynchronous, there are a few synchronous mobile learning systems that support synchronous learning. MLVLS [2] is a synchronous mobile learning system that allows students to watch live video and slides on Symbian OS-based smartphone, although it doesn’t support real-time interaction between an instructor and students.

We have an experience of developing a smartphone-based real-time mobile learning system that enables students not only to watch video and slide with annotation but also to send text feedback to an instructor in real time on a smartphone [3]. But the problem with smartphone-based synchronous mobile learning system is that the display is so small for a user to look at the lecture content. The slides are often made with a presentation application on a desktop PC or a notebook equipped with a 13 to 24 inch display. However, those figures and texts on a slide as well as annotation that are large enough to look at comfortably on the display of a PC or a notebook are not easily recognizable on the smartphone with 4 to 5 inch display. And it is often difficult to type questions with small on-screen keyboard of the smartphone and recognize answers from an instructor in real time. These problems can be more serious to people who has poor eyesight.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_39,

These problems of the smartphone-based system can degrade the student's learning experience. Classroom Presenter [4] is a tablet-based synchronous learning system that allows sharing of lecture slides and annotation between an instructor and students in a classroom. However, since it assumes an instructor and students are in the same classroom, it doesn't support video and audio of the instructor nor questions from students.

Therefore, we propose a tablet-based synchronous mobile learning system that enables students not only to watch the video and slides with annotation but also to send questions to an instructor in real time with a tablet. The slide with annotation and the questions and answers on the proposed system's display is comfortable enough for students to look at. Typing questions on a large on-screen keyboard of the proposed system became also easy and quick.

This paper consists of the following sections. Section 2 shows the design and implementation of the proposed tablet-based synchronous mobile learning system. In section 3, we describe the usability assessment of the prototype system. Finally, section 4 shows the conclusions and future work.

2 Tablet-Based Synchronous Mobile Learning System

Fig. 1 shows the communication architecture of the presented tablet-based synchronous mobile learning system consisting of a desktop server for instructor running on Windows PC and a tablet client for student running on iOS iPad.

The desktop server acts as a encoding and broadcasting server as well as a client for instructor. The desktop server encodes lecture video, audio, and slide with annotation. The video is encoded with H.263 and the audio is encoded with G.723.1.

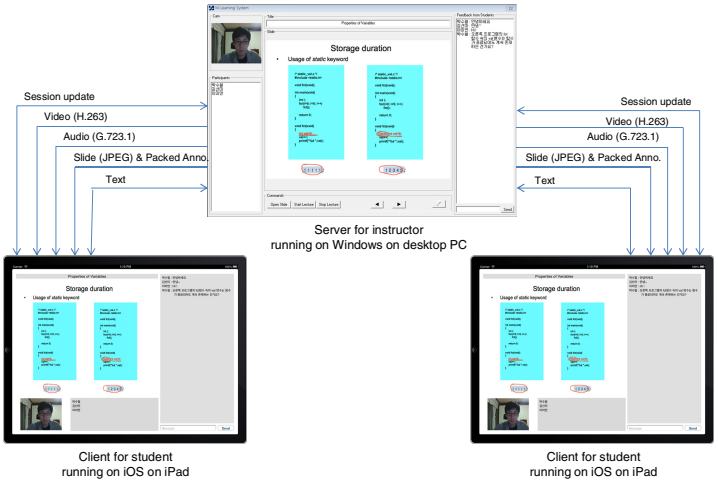


Fig. 1. Communication architecture of the tablet-based synchronous mobile learning system

The slide is encoded with JPEG. Annotation events that occurred for a short period of time is grouped and packed into a packet. The desktop server also broadcasts to tablet clients various data: encoded lecture video, encoded audio, encoded slides with annotation, text feedback from a tablet client and session update (e.g., joined a lecture session). Each type of data is sent to a client through a separate socket. As a client for instructor, the desktop server renders the video of the instructor from a webcam, plays the audio from a microphone, renders the slide with annotation, and displays the text feedback.

When a tablet client receives lecture data from the desktop server, it decodes them and renders the video and slide with annotation or plays audio. The client can send the text feedback and session update to the desktop server which, in turn, broadcasts them to all the clients and displays the feedback and updates its own session state.

The development platform for the presented system is as follows. The tablet client is being developed in Objective-C with the Xcode 4.6 integrated development environment [5] and iOS SDK 6.2 [6] on Mac OS X Mountain Lion 10.8.3. The desktop server is being developed in Microsoft Visual C++ 2010 with MFC on Windows 7.

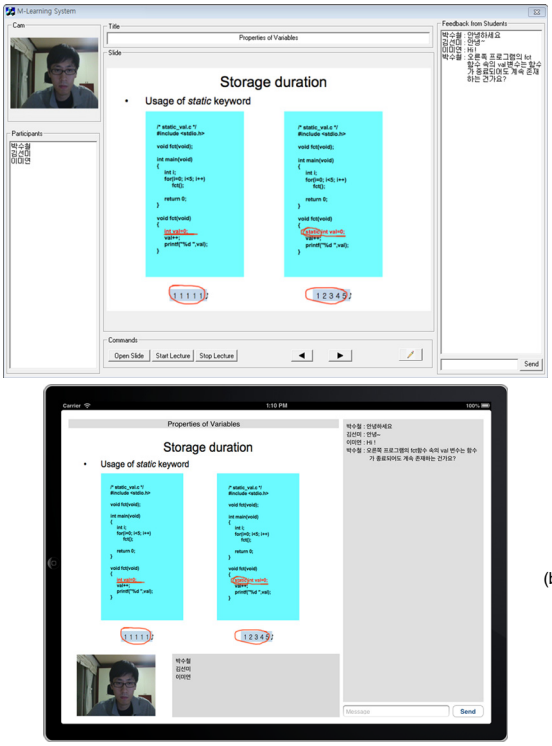


Fig. 2. (a) User interface of the server for instructor running on Windows on a desktop PC and (b) user interface of the client for student running on iOS on an iPad tablet

The rendering of video in the desktop server is done by using Video for Windows library. Audio in the server is handled with winmm.kib library. The network communication in the server is implemented with CAsyncSocket class.

The rendering of video at the tablet client is done by using Core Graphics library of the Media Framework supported by iOS. The playing of audio at the client is done with AudioToolbox Framework and AVFoundation Framework. The sending of data at the client is implemented with CFSocket library in Core Foundation of iOS.

Fig. 2(a) shows the user interface of the desktop server for instructor. The UI of the instructor's desktop server consists of the panels for video of the instructor, participant list, lecture title, slide, slide control command panel and feedback from students.

Fig. 2(b) shows the user interface of the iPad tablet client for student. The UI of the student's tablet client consists of the panels for slide, video of the instructor, participant list, and feedback from students.

3 Usability Assessment

We conducted the usability assessment for the prototype of the proposed tablet-based mobile learning system with undergraduate students of the Department of Computer Engineering at Hongik University. The theory behind our assessment method is the Technology Acceptance Model(TAM) [7]. TAM is a theory of measuring user acceptance of information technology based on scales for two specific variables, perceived usefulness and perceived ease of use, which are hypothesized to be fundamental determinants of user acceptance. These two variables are considered to affect the intent to use, which in turn has an influence on the usage behavior.

The scale of perceived usefulness and perceived ease of use ranges from 1 (lowest) to 5 (highest). The assessment was conducted with user surveys that were collected from 15 students who responded among 18 students who were asked. Among 15 students, 9 were female and 6 were male.

In our assessment, one variable called the perceived usefulness means that using the proposed mobile learning system would enable a student to learning something more quickly. Table 1 shows the perceived usefulness of our prototype system that can be further categorized into four aspects. Those aspects are how useful the size of the screen real estate of client is to a student in terms of understanding the lecture in overall (mean: 3.93), how useful the combination of the components (video, audio, slide, annotation) of client are to a student in understanding the lecture content(mean: 3.87), how useful the text feedback mechanism for asking an instructor a question in real time is to a student who didn't understand a certain part of a lecture (mean: 4.00) and how useful the whole UI is to a late student who tries to figure out the context of the current lecture and catch up with it (mean: 3.80).

The perceived usefulness of the client screen size and the text feedback are relatively high. This result suggests that the big screen size and the big virtual onscreen keyboard of tablet results in user's high satisfaction. The average of the means of the four aspects of the perceived usefulness is 3.90, which indicates that users who tested our prototype feel that the proposed tablet-based mobile system is somewhat useful.

Table 1. Perceived usefulness

Aspect	Frequency					Mean	SD
	1(lowest)	2	3	4	5(highest)		
Client screen size	0	0	4	8	3	3.93	0.68
Component combination	0	0	5	7	3	3.87	0.72
Text feedback	0	0	4	7	4	4.00	0.73
UI for late participant	0	2	3	6	4	3.80	0.98
Average	0	0.5	4	7	3.5	3.90	0.78

The other variable called the perceived ease of use means that it would be easy for a student to become skillful at using the proposed mobile learning system. Table 2 shows the perceived ease of use of our prototype system that can be further categorized into three aspects. Those aspects include how easily a student can recognize the video and audio of the instructor (mean: 3.60), how easily a student can recognize the texts and figures in the lecture slide as well as the annotation (mean: 3.93) and how easily a student can type a feedback message such as question to the instructor quickly without typing errors using virtual on-screen keyboard (mean: 3.93).

The perceived ease of use of slide/annotation recognition and text-feedback typing are relatively high. This suggests that the big display size of the client enables students to easily recognize the text and figures in slide and annotation and to easily type questions on the virtual onscreen keyboard that is bigger than the smartphone counterpart with uncomfortably small virtual keyboard. The perceived ease of use of video/audio recognition is relatively low. This could be that the transmission of the video and audio is sometimes not continuous depending on the condition of the wireless network and the optimization of the decoding is not complete yet. The average of the means of the three aspects of the perceived ease of is 3.82, which suggests that testers of the prototype feel that the proposed tablet-based mobile system is easy to use to a certain level that is above the normal level.

Table 2. Perceived ease of use

Aspect	Frequency					Mean	SD
	1(lowest)	2	3	4	5(highest)		
Video/Audio recognition	0	1	6	6	2	3.60	0.8
Slide/Anno. recognition	0	0	4	8	3	3.93	0.68
Text-feedback typing	0	1	3	7	4	3.93	0.85
Average	0	0.67	4.33	7.00	3.00	3.82	0.78

As described above, the perceived usefulness and the perceived ease of use in using our prototype are between 3 and 4, which suggests that the students feel that the proposed tablet-based mobile learning system is somewhat useful and easy to use to a certain extent above the normal level.

4 Conclusions

We propose a tablet-based synchronous mobile learning system that enables students to participate in a live distance learning session with their tablets. The students can watch the live video of instructor and slide with annotation. They can also send questions to an instructor with text in real time. The system allows students to recognize the texts and figures of a slide and annotation more easily and to type questions to an instructor more comfortably compared to the smartphone-based counterpart. We performed an assessment of the proposed system by having a group of students use the prototype. The survey result shows that perceived usefulness and ease of use of the proposed system by the students is higher than the normal level. We plan to develop more effective encoding/decoding mechanism of video, audio, slide and annotation. We hope that the proposed tablet-based mobile learning system will help develop more effective mobile learning system for students.

References

1. Wains, S.I., Mahmood, W.: Integrating M-learning with E-learning. In: 9th ACM SIGITE Conference on Information Technology Education, pp. 31–38. ACM (2008)
2. Ullrich, C., Shen, R., Tong, R., Tan, X.: A Mobile Live Video Learning System for Large-Scale Learning-System Design and Evaluation. IEEE Transactions on Learning Technologies 3(1), 6–17 (2010)
3. Lee, J.: Real-Time Mobile Distance Learning System for Smartphone. In: Luo, Y. (ed.) CDVE 2012. LNCS, vol. 7467, pp. 24–32. Springer, Heidelberg (2012)
4. Anderson, R., Anderson, R., Davis, P., Linnell, N., Prince, C., Razmov, V., Videon, F.: Classroom Presenter: Enhancing Interactive Education with Digital Ink. IEEE Computer 40(9), 56–61 (2007)
5. Xcode 4, `https://developer.apple.com/xcode`
6. iOS Dev Center, `https://developer.apple.com/devcenter/ios`
7. Davis, F.D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 13(3), 319–340 (1989)

Bidirectional Multichannel Beacon Management for TVWS WPAN

Young-Ae Jeon[1], Dae-Young Kim[2], In Jang[3], and Kwang-Il Hwang[3,*]

[1] IT Convergence Technology Research Laboratory, ETRI, Daejeon, Rep. of Korea
yajeon@etri.re.kr
[2] Dept. of Information & Communication, Chungnam National University, Daejeon, Rep. of Korea
dykim@cnu.ac.kr
[3] Department of Embedded Systems Engineering, Incheon National University, Incheon, Rep. of Korea
hkwangil@incheon.ac.kr

Abstract. Unlike general ISM bands, TVWS bands have unique requirements. In particular, to deploy WPAN systems with extremely restricted resources on a TVWS band, cooperative channel management strategy among devices are required. In this paper, we propose a Bidirectional Multichannel Beacon Management $(BM)^2$ which fulfils the strict requirements, such as PAN operation independency, periodic channel monitoring capability, minimum overhead, for TVWS WPAN. The experimental results demonstrates that $(BM)^2$ shows high superframe utilization ratio and small recovery time.

Keywords: Beacon Scheduling, IEEE802.15.4e, IEEE802.15.4m, LR-WPAN, Multichannel Utilization, Superframe Duration Allocation, TV White Space.

1 Introduction

The television white space (TVWS) refers to television channels that are not used by any licensed services at a particular location and at a particular time. To exploit this unused TVWS spectrum for improved spectrum efficiency, regulatory agencies have begun developing regulations to permit its use this TVWS. Therefore, regulatory requirements for white space communications are being identified based on the rules from the FCC (refer to FCC 10-174) and ECC (refer to ECC Report 159).

In order to deploy various wireless systems on TVWS bands, several TVWS standardization activities in the world, such as IEEE 802.22[1], IEEE 802.11af [2] etc, are being rolled out. In addition, there are many instances in large area device command and control applications where infrastructure requirements need to be minimized for effective deployment. These needs are effectively served by the ability to operate 802.15.4 class networks in the TVWS spectrum [3].

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*, Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_40,

However, communications in TVWS bands has more strict requirements than in general ISM bands. According to the FCC regulation, four types of white space devices are defined: Fixed, Mode 2, Mode 1, Sensing-only Device. Fixed and Mode 2 devices has geolocation and database access features. Mode 1 device does not have geolocation and database access feature. Sensing-only device relies only on sensing to protect incumbent users. In particular, Fixed and Mode 2 devices must access a TV band database over the Internet to determine the TV channels that are available at their geographic coordinates prior to their initial service transmission at a given location. Furthermore, they shall access the database each time it is activated from power-off as well as at least a day to verify that the operating channels continue to remain available.

Based on the above basic requirements, TVWS WPAN should be designed taking account of the followings.

- PAN operational Independency
- Periodic channel monitoring capability
- Minimum overhead

Therefore, this paper deals with a method for multiple PC(PAN Coordinators) to efficiently operate and manage channels utilizing multiple WPAN channels in available TV channel.

Even though there are several researches [4 - 5] related to multichannel based WPAN, these are not satisfied with the TVWS WPAN requirements.

In this paper, we propose a bidirectional multiple beacon management method for TVWS WPAN meeting the TVWS requirements.

2 $(BM)^2$: Bidirectional Multichannel Beacon Management

According to the regulation, in TVWS WPAN a master device which plays a role in accessing a TV database over the Internet and managing the TVWS WPAN channels, is required. Therefore, we define an additional device, SPC(Super PAN Coordinator), which is more powerful than general WPAN PC. The SPC has the Internet access capability and obtains an available TV channel from the database. Then it maps the TV channel to corresponding WPAN channel set. As a matter of fact, since WPAN requires smaller bandwidth than other wireless systems, a TV channel can be divided into a number of available WPAN channels. (e.g., the number of available WPAN channels in case 2FSK with 200KHz channel spacing is 638 at TV channel 0 of 470.2MHz)

The proposed idea is to enable each PC operate in an independent channel by allocating different channels to different PCs. Furthermore, the SPC has the ability to listen to each periodic beacon of child PCs and to monitor each channel state through the periodic beacon listening. It results in no additional transactions to monitor the state of channels allocated to child PCs. Therefore, this architecture constructs bidirectional beacon control and thus it is called $(BM)^2$ (Bidirectional Multiple Beacon Management).

Figure 1 illustrates the extended superframe structure for TVWS WPAN. We define additional period, BOP (Beacon Only Period), at the end of basic superframe, in which SPC can listen to different beacons of child PCs operating on different channels. The BOP consists of DBSs. Each DBS is composed of one or more base slots, which are *aBaseSlotDuration* in length.

The PCs that channel is not yet allocated transmit *DBS request* frame on the common channel (a base channel in which SPC transmits a periodic beacon) and on the reception of the request, the SPC allocates a vacant channel out of available channel set and a DBS(Dedicated Beacon Slot) within the managing BOP. The requester is informed of the allocation information (channel and DBS) through *DBS allocation* frame of SPC. Upon the reception of *DBS allocation* frame, the requester (PC) schedules its own DBS time, switches from current SPC channel to the allocated channel at the corresponding time, sends its beacon, and maintains its own superframe on that channel for its superframe duration.

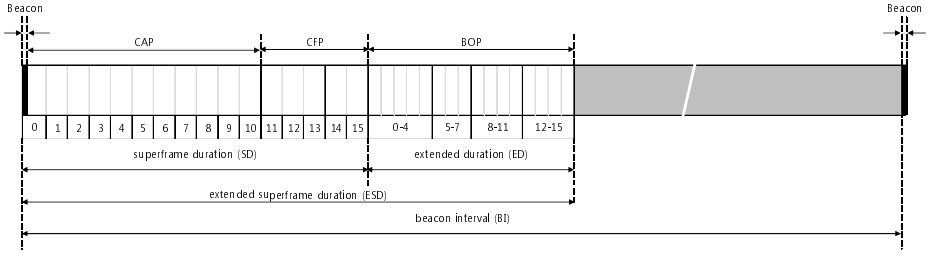


Fig. 1. Extended superframe structure

Figure 2 shows an example of bidirectional multiple beacon management of a TVWS WPAN in which DBS allocation is already completed. As shown in Fig. 2, the SPC can monitor the state of the allocating channels without additional channel verification procedure, by listening to beacons of its child PCs switching to the corresponding channel at the start of each DBS time for the total BOP duration. This management method in which the SPC is capable of monitoring multichannel beacons of child PCs using BOP, can eliminate the overhead of addition transactions to manage allocated channels, and guarantee independency of child PANs as much as possible.

Even though TVWS bands are regulated by regional agency, there still exist interference possibilities on certain channel by various wireless systems. Moreover, current TV channel that the SPC is using should be able to be changed according to the result of periodic verification transaction of SPC to the TVWS server. In the case of interference problem, since each PC that transmits beacon frames periodically without CCA (Clear Channel Assessment) cannot recognize some problems on its own channel, it is required that SPC is capable of identifying problems on the allocated channels, and reallocating other channel to the problematic PC.

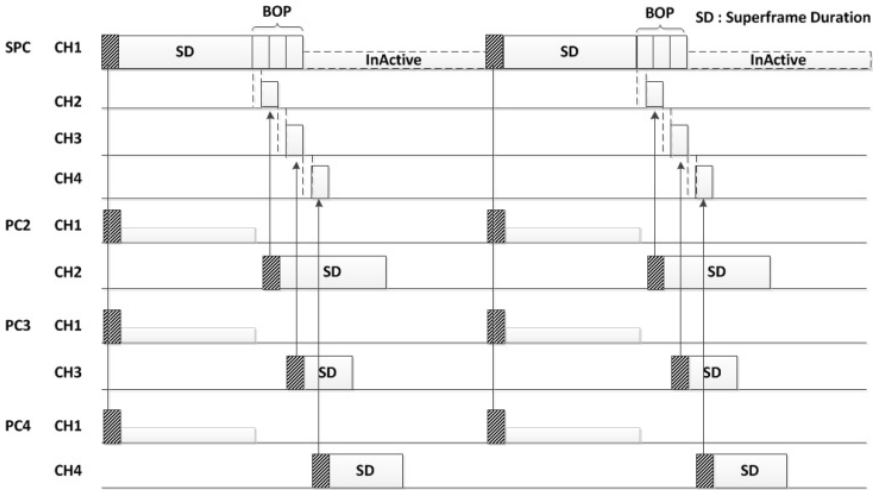


Fig. 2. An example of bidirectional multi beacon management

The $(BM)^2$ is capable of coping well with these problems. As shown in Fig. 3, in the case of listen failure of three consecutive beacon frames of a specific child PC on the corresponding channel, the SPC considers that the channel has some problem, and makes the PC change the problematic channel into other channel immediately by sending DBS allocation frame for a new channel allocation at upcoming superframe duration of SPC. Since the PC already maintains its own superframe and follows the schedule of the SPC, DBS slot for the PC in BOP is not changed. Furthermore, even in the case that entire channel that a SPC is managing should be changed, each PC can re-operate changing only the current channel without reallocation procedure of DBS slot.

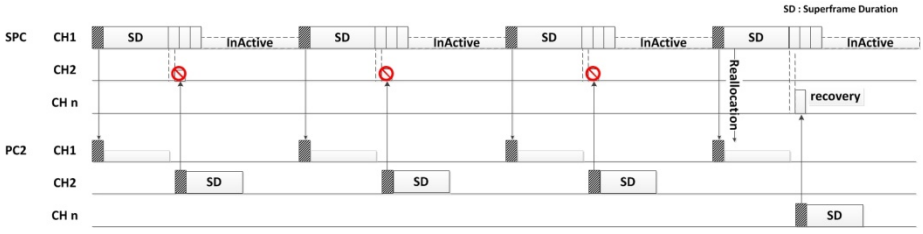


Fig. 3. An example of reactive channel recovery

3 Experimental Results

In order to evaluate functionality and performance of $(BM)^2$, we implemented $(BM)^2$ modifying lr-wpan module in ns-3 [6]. We basically experimented with star topology, as shown in Fig. 4, having 1 SPC and the different number of child PCs from 1 to 12. Table 1 shows major parameters used in our experiment.

Table 1. Simulation Parameters

Parameter	Value
aBaseSuperframeDuration	*aBaseSlotDuration* × *aNumSuperframeSlots*
aNumSuperframeSlots	*16*
aBaseSlotDuration	*60symbols*
SD(Superframe Duration)	*aBaseSuperframeDuration* × 2^{SO}
BI (Beacon Interval)	*aBaseSuperframeDuration* × 2^{BO}
ED (Extended Duration) = BOP Duration	*aBaseSuperframeDuration* × 2^{EO}
The number of Super PAN	*1*
The number of coexist PANs	*1~12*
The number of available Channels	*16 Channels (various)*
BO(Beacon Order)	*0~14*
SO(Superframe Order)	*0~13*

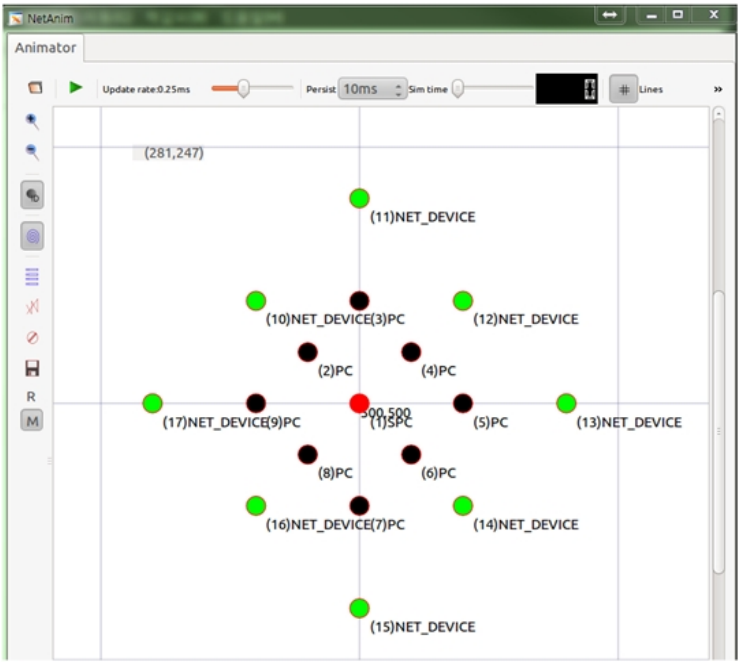


Fig. 4. Simulation Topology

First, we observed superframe utilization, which represents a utilization ratio of available maximum superframe duration of child PCs with respect to varying beacon interval of SPC. That is, superframe utilization is used to measure an independency of child PCs on SPC. For a comparative performance evaluation, we used

IEEE802.15.4e DSME [7] which is capable of multiple superframes scheduling within a single channel. Figure 5 shows the result of SU (superframe utilization) ratio at the different number of child PCs. The available maximum number of superframe of child PCs is increased proportional to the increment rate of beacon interval size. Therefore, the result shows that SU ratio of both methods maintains the constant value respectively without regard to the beacon interval size, which is determined by BO value, in steady state. However, single channel shows rapid decline as the number of child PCs increases. It maintains reasonable SU ratio (0.25) at 2 child PCs but shows very low SU ratio (0.0625) at 12 child PCs. On the other hand, $(BM)^2$ shows constant SU ratio (0.5) without regard to the number of child PCs. That is, for example, beacon interval size with 10 BO value is 2^{10}*$BaseSuperframeDuration$, and in that case all child PCs can utilize superframe duration bounded by $(0 \le SO \le 9)$.

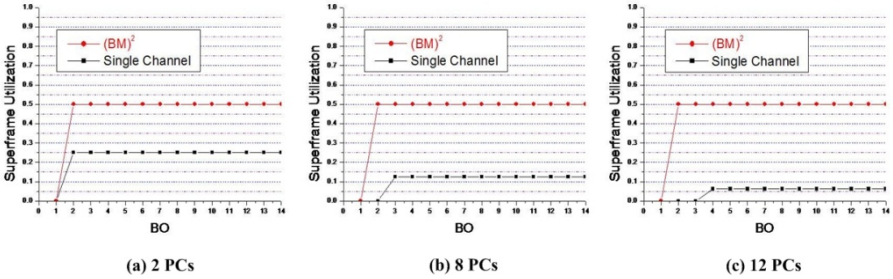


Fig. 5. A comparison of superframe utilization

This results from the fact that while in the case of single channel, the beacon interval of SPC should be divided by the number of child PCs plus 1, $(BM)^2$ allows for each child PCs to operate on different channel at the same time so that available superframe duration is more extended than in single channel. In addition, small portion out of remaining 50% in beacon interval of SPC is used for extended superframe duration for managing BOP, and the remainder can be used for inactive duration to save energy.

Next observation is recovery time from interference environment. We measured the total time required to be completely recovered switching to other channel from the time when the SPC identifies problems on the channel, in the case that interference on a channel of an arbitrary child PC occurs. Figure 6 shows the result of total recovery time divided by detection time and recovery time. The SPC monitors allocated channels through periodic beacon listening so that if the SPC fails to listen to three consecutive beacons, it considers the channel has some problem. That is, the detection period is the time length to identify three consecutive beacon failures so that the variation of the time is proportional to the BO size. On the other hand, it has also been shown that pure recovery time is considerably small (max. 754msec). It results from the fact that the SPC sends *DBS allocation* frame for a new channel allocation to the problematic PC at the SPC superframe duration right after problem detection, and thus the SPC can listen to refreshed beacon of the problematic PC on the new channel at the corresponding DBS within the immediate upcoming BOP.

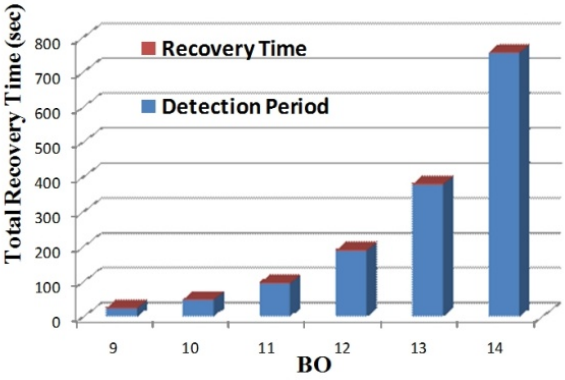


Fig. 6. Recovery Time

4 Conclusion

In this paper, we proposed a Bidirectional Multichannel Beacon Management $(BM)^2$, and its performance was evaluated. Through the experiment, it is demonstrated that the proposed $(BM)^2$ supports better PAN operation independency than single channel beacon scheduling methods, and is capable of recovering the channel quickly through periodic beacon listening. In addition, the overhead of additional transactions to manage allocated channels is eliminated.

In particular, since communication systems on the TVWS band should have the ability to cope well with channel diversity characteristics, it is expected that the proposed $(BM)^2$ will be able to contribute to designing new TVWS PAN utilizing multi-channels.

Acknowledgement. This work was supported by the IT R&D program of MKE/KEIT [10041864, Development on Spectrum Efficient Multiband WPAN System for Smart Home Networks], and also supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2012R1A1A2041271).

References

1. IEEE P802.22-2011, Standard for Wireless Regional Area Networks Part 22: Cognitive Wireless RAN Medium Access Control (MAC) and Physical Layer (PHY) specifications (July 2011)
2. IEEE P802.11af™/D1.02 Draft Standard Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 3: TV White Spaces Operation U.S. (June 2011)
3. http://www.ieee802.org/15/pub/TG4m.html

4. Abdeddaim, N., Theoleyre, F., Rousseau, F., Duda, A.: Multi-Channel Cluster Tree for 802.15.4 Wireless Sensor Networks. In: IEEE Proceedings of PIMRC 2012, pp. 590–595 (2012)
5. Wu, Y., Stankovic, J.A., He, T., Lin, S.: Realistic and Efficient Multi-Channel Communications in Wireless Sensor Networks. In: INFOCOM, pp. 1193–1201 (2008)
6. ns-3, http://www.nsnam.org
7. IEEE Std 802.15.4e, Part 15.4: Amendment 1: MAC sublayer. IEEE Standard for Information Technology (2012)

An Information Architecture for Preventing and Tracing Information Leakage in the Age of Micro Devices

Jong Uk Choi[1] and Joo Won Cho[2]

[1] Sangmyung University, Seoul, Korea
juchoi@smu.ac.kr
[2] Markany, Seoul, Korea
jwcho@markany.com

Abstract. Micro devices, such as high-quality camera and voice recorder module, are becoming essential part of future computer systems in Google Glass, Apple's iWatch, Samsung's WatchPhone, Olympus's MEG 4.0, Baidu's Baidu Eye, and others. In addition, separate products of tiny devices, micro camera or digital voice recorders are already in the market for sale.

Wide-spread use of micro devices, whether in smart phones or separate modules, has brought serious issue of internal information leakage problem. Especially, 'smart work' computing environment in which each employee collaborate with other people through mobile devices, out of office will lead to serious security holes in future business. In this paper addressed is a new information security architecture for information leakage prevention.

In this paper, discussed are new ways of information protection technologies which utilizes E-DRM for prevention, and watermark technologies for tracing. Watermark technologies have been employed in copyright protection area to trace illegal copy and distribution in the cyber space. We applied those watermark technologies to tracing information leakage at private organizations and government agencies who are trying to protect sensitive information.

Keywords: Enterprise DRM, watermarking, access control, micro-device, internal security.

1 Introduction

Internal information protection is becoming one of the most serious issues with a wide spread use of smart phones and micro devices. As shown in Fig.1 and Fig.2, there are so many micro devices available in the market at a price range of 40 – 60 USD. Considering that most smart phones are equipped with high-quality camera and voice recorder, there are no ways to protect internal information from malicious actions of employees in private sectors or government organizations. In other words, those technologies which have provided so far very effective ways to prevent illegal copy from computer systems and transfer to third parties through networks, wired or wireless, or storage devices, cannot provide any protection mechanism in the age of micro devices and smart phones. Even if very confidential information can be made

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_41,

copy but cannot be properly displayed on the third party's computer, because of encryption mechanism of traditional DRM (Digital Right Management) technology, it can be properly displayed at employee's terminal and easily taken photo copies by smart phones or micro devices.

Fig. 1. Mini DV Pen Spy Video Hidden Spy Camera Recorder support max 16GB TF MicroSD 10PC Lot (http://www.dhgate.com/wholesale)

Fig. 2. Wholesale - Black Mini Wireless SPY Hidden DV DVR U8 USB DISK HD Camera Motion Detector (http://www.dhgate.com/wholesale)

In this paper, suggested is an architecture for internal information security system which is comprised of prevention technology and tracing technology. Prevention technology in internal information security area has heavily relied on encryption and control of user activities, which is called 'Enterprise DRM'. Tracing technology suggested in this paper utilizes various watermarking mechanism which can embed information into digital text, image, audio, and video. Traditionally, DRM and watermarking technologies have been employed in protecting copy right of digital contents.

In the age of wired network and even in the age of wireless networks, E-DRM can effectively prevent illegal copy and transferring confidential information. However, with wide spread use of smart phones and micro devices, internal information security

faces very serious challenge: photo copy of displayed computer terminal and voice recording in a very confidential meeting. Traditional technologies cannot prevent photo copying and voice recording. Based on the observation, suggested is a new architecture of internal information security system in which traditional E-DRM and watermarking technologies are employed.

Fig. 3. Architecture of Internal Information Security System

In the following section, traditional prevention technology, called E-DRM, is described. In section 3 tracing mechanism based on watermarking technology is explained while conclusion and future research is described in section 4.

2 Prevention: E-DRM for End-User Access Control

E-DRM technology has evolved from DRM (Digital Right Management) technology which was developed to deliver digital contents and to prevent illegal copies and distribution of the digital contents. DRM technology's copyright protection capabilities primarily rely on data encryption and access control functions. Generally, encrypted contents such as songs or movies are delivered to a user with an encryption key which may be in the same content package, or sent through a separate channel. At the rightful applicant's terminal or user's device, the contents are decrypted and played. Because the contents are encrypted with a key delivered to the user, when it is illegally copied and sent to other users, it cannot be properly played. A DRM agent decrypts packaged contents and enforces access control to the contents based on the payment conditions, such as the number of plays, whether to be able to send the contents to a third party, etc. Currently, the access control of DRM systems in playing digital contents is very simple: 'play' or 'not play'.

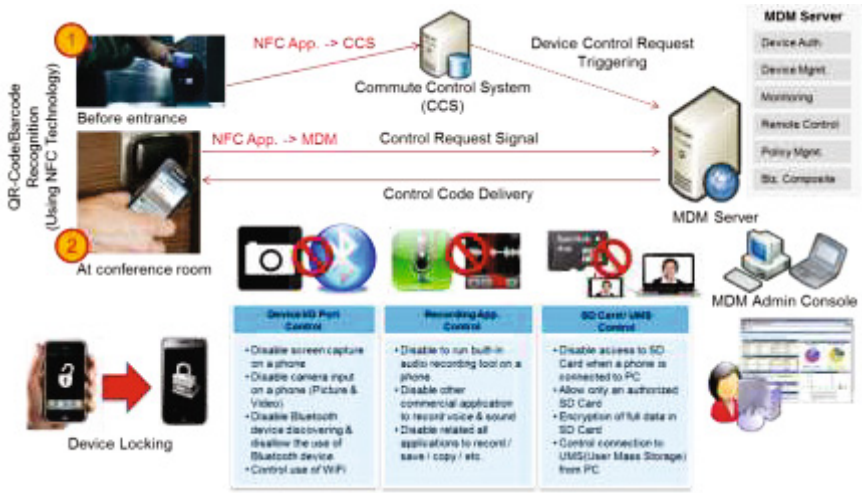


Fig. 4. Functions of MDM device control

Mobile Device Management (MDM) is getting popular in IT community, encompassing from a simple device control technology to a very comprehensive control system. As shown in Figure-5, device control is an important part of the MDM Technology, including control of camera, recorder, and SD storage. When an employee comes into office area (secure zone) and touches NFC reader, the functions of camera, recorder, and storage will be turned off. On the contrary, when the employee gets out of the office or secure zone, the function of important devices, with a touch to NFC reader, will be turned on. There are more functions which are added to the device control in MDM technology.

3 Tracing: Watermarking Technologies

Watermarking technology embeds imperceptible signal into digital content for authentication and forensic purpose. In the early of 1990s, the technology was developed to authenticate authors of digital contents, for example, image, audio, or video by hiding author's information. Because the technology hides imperceptible signal to naked eyes or ears without digital distortion, the embedded marks can be copied and transferred to third parties. That is why digital watermarking technology has been called 'passive technology' or 'tracing technology, while DRM is called 'active technology' or 'prevention technology'.

Recently, watermarking technology is being employed as a 'forensic technology' to identify illegal distributors. Currently music service providers are using audio watermarking technology to trace illegal copy and distribution when they send new music album to radio stations for broadcast. For example, when Universal Music International (UMI) has a new song in the market and tries to send it to radio station in the States using internet file transfer function, UMI is not sure whether digital file

of the song can be illegally copied and distributed in the internet. For the purpose of tracing illegal copy and distribution, UMI embeds imperceptible signal, ID number of radio station, into the music which cannot be audible to even golden ears and then sends the 'watermarked music file' to radio stations. Every music file sent by UMI has different unique number inside the music which can be detected by UMI when the music file is found to have been illegally distributed.

3.1 Watermarking Technology for Tracing

Image watermark technology and audio watermark technology have been successfully applied to copyright protection and internal information protection [www.markany.com] through tracing and authentication. Likewise, text watermarking technology is very effective in tracing illegal leakage of digital document. When an employee takes photo of displayed document or data on his/her computer terminal, how can it be traced?

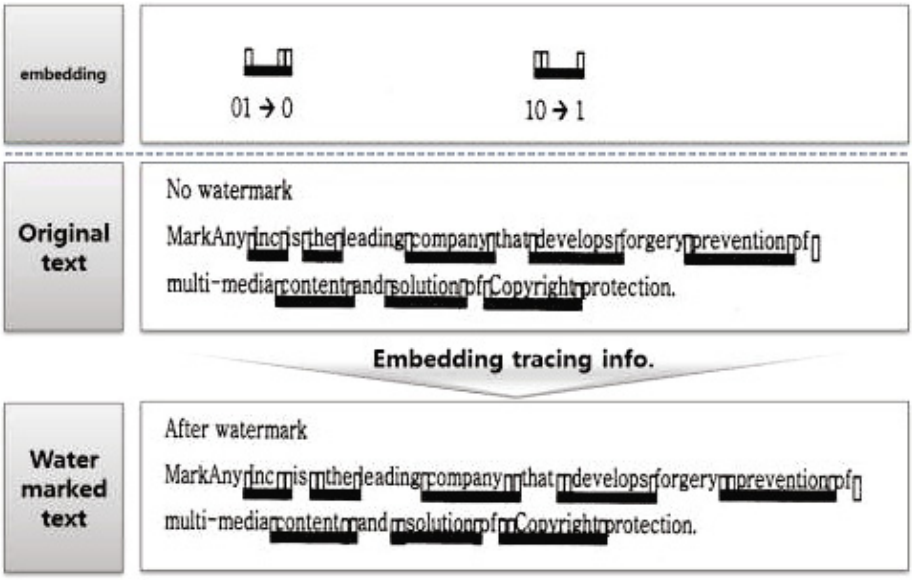


Fig. 5. Text Watermarking

3.2 Audio Watermarking: Tracing Illegal Voice Recording in a Security Zone

A user can attempt recording conversation in a meeting room with smart phones or tiny voice recorder disguised as glasses, foundation pen, or USB. As shown in Fig.*, there are so many micro devices which are being sold in the market, usually having a big memory capacity, and Bluetooth function. They are basically equipped with camera function and voice recording function with wireless transmission mechanism.

Open Space Audio Tracing (OSAT) technology was developed, based on audio watermarking which can embed imperceptible data signal into the digital file and which can detect hidden information later. In a security zone, such as military camp or government top office, this technology can be usefully employed to trace who and when the conversation is recorded. As depicted in Figure-8, the system is composed of WM (watermark) signal generator and WM detector.

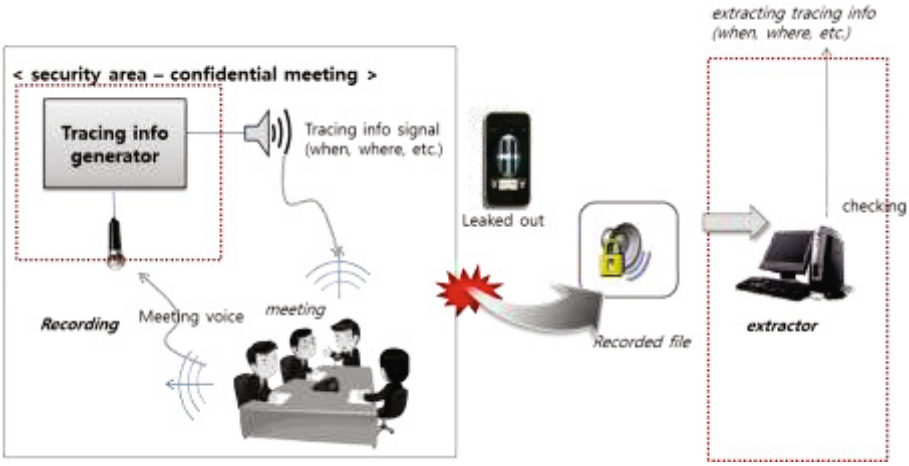


Fig. 6. Audio Tracing in a Security Zone

As depicted in Figure-8, conversations in a security zone can be secretly recorded without notice of the administration by one of the meeting participants. However, in the OSAT technology, information of 'meeting time' and 'location' are broadcast in the meeting room by signal generator and embedded into the recorded digital file. OSAT, even if it cannot prevent illegal recording, can disclose 'when' and 'where' the conversation was recorded by detecting recorded conversation. Because number of the participants is usually limited to a quite few members in the organization, the information of the meeting (location and time) can reveal who has made illegal recording.

4 Conclusion and Future Research

In this paper, we suggested a new architecture for internal information security system in which E-DRM technology and watermarking technology play important role in effectively preventing information leakage and tracing illegal photos taken at user terminal, and voice recording of meeting conversation in a security zone. Those technologies have been available in the market for a quite long time, and therefore cannot claim novelty of the system. However, we claim that applying those old technologies to new area might very meaningful in solving problems which newly have come up.

We believe that watermarking technology suggested in this paper as an information security solution, instead of copyright protection technology in the age of micro devices, might be powerful mechanism in protecting confidential information in military organizations, top government organizations, and private sectors.

References

1. Arnab, A., Hutchison, A.: Requirement Analysis of Enterprise DRM Systems. In: Proceedings Information Security, South Africa, Hotel Balalaika Sandton, Johannesburg (2005)
2. Elisabeth, H.: How enterprise DRM works: Everything you need to know about information rights management. Computer World (April 27, 2010)
3. Chongthammakun, R., Jackson, S.J.: Boundary Objects, Agents, and Organizations: Lessons from E-Document Systems Development in Thailand. In: Proceedings of 2012 45th Hawaii International Conference on Systems Sciences, Hawaii (2012)
4. Markany, Method and Apparatus of watermark embedding and extraction from text documents, patent, filed at October 2002, and registered at January 14, 2005, No. 10-0467930, Korea Patent Office
5. Cheng, W., Feng, H., Yang, C.: A robust text digital watermarking algorithm based on fragments regrouping strategy. In: 2010 IEEE International Conference on Information Theory and Information Security (ICITIS), pp. 600–603 (2010)
6. Yu, Z., Liu, X.: A New Digital Watermarking Scheme Based on Text. In: International Conference on Multimedia Information Networking and Security, MINES 2009, vol. 2, pp. 138–140 (2009)
7. Jalil, Z.: A Review of Digital Watermarking Techniques for Text Documents. In: International Conference on Information and Multimedia Technology, ICIMT 2009, December 16-18, pp. 230–234 (2009)
8. Bharati, P.D., Nitin, P.N.: Text watermarking algorithm using structural approach. In: 2012 World Congress on Information and Communication Technologies (WICT), October 30-November 2, pp. 629–633 (2012)

Automatic Images Classification Using HDP-GMM and Local Image Features

Wanhyun Cho[1], Seongchae Seo[2], In-Seop Na[2,*], and Soonja Kang[3]

[1] Department of Statistics, Chonnam National University, Gwang-Ju, Korea
[2] School of Electronic & Computer Engineering, Chonnam National University, Gwang-Ju, Korea
[3] Department of Mathematics Education, Chonnam National University, Gwang-Ju, Korea
{whcho,scseo,kangsj}@chonnam.ac.kr, ypencil@hanmail.net

Abstract. In this paper, we propose a new method based on the probability model that can classify automatically various images by subjects without any prior exchange of information with users. First, we introduce the hierarchical Dirichlet processes Gaussian mixture model (HDP-GMM) that can be applied in images classification, and consider the variational Bayesian inference method to estimate the posterior distribution for the hidden variables and parameters required by this model. Second, we examine the extraction method of various local patches features from given image, which can accurately represent the colors and contents of images. Next, we have trained the HDP-GMM using the extracted patch features, and then present a scheme to classify a given image into the appropriate category or topic by using trained model. Finally, we have applied our model to classify various images datasets, and we have showed the superiority of the proposed method using several evaluation measures for classification method.

Keywords: Automatic images classification, Hierarchical Dirichlet processes Gaussain mixture model, Variational Bayesian inference algorithm, Local patch features, Histogram feature and descriptor feature.

1 Introduction

Today, the explosive usage of internet system or smart phone is able to yield a large amount of visual data. Unfortunately, since these data given from its instruments are both scattered and unorganized, making search and retrieval of related image is very difficult. But, large digital libraries, which are built by collecting resources from different locations, can make searching for specific images or video shots. They would also like to browse and navigate through the image corpus. Such requirements have created great demands for effective and flexible systems to manage or classify digital images or videos. Therefore, automatic image classification algorithm is one of the most important components of machine vision, and it is also the main challenge of

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*, Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_42,

artificial intelligent. But, it is so difficult to find comprehensive solution of machine vision problems [1].

Recently, various methods that can be used to classify documents or images were proposed at the machine learning area. These methods mostly use the statistical models, such as Latent Dirichlet Allocation (LDA), Gaussian Mixture Model (GMM), Hidden Markov Model (HMM), Dirichlet Process Mixture (DPM), and Hierarchical Dirichlet Process (HDP).

Blei et al. [2] proposed LDA, a generative probabilistic model for allocation a collections of discrete date such as text corpora. And they present efficient approximate inference techniques based on variational methods and an EM algorithm for empirical Bayes parameter estimation. Hu [3] discuss algorithms that extent LDA to accomplish tasks like document classification for text, object location for images, and automatic harmonic analysis for music. For each domain, he also emphasizes approaches that go beyond LDA's traditional bag-of-words representation to achieve more realistic models that incorporate order information. Permuter et al. [4] published the review paper about a study of Gaussian mixture models (GMMs) of color and texture features for image classification and segmentation. But, LDA or GMM model used in the papers that reviewed so far has a problem that must assume in advance the number of hidden topic or components. In fact, we don't know how many number of hidden topic or components in advance.

As this time, a new model to solve these problems, the Dirichlet process mixture model has been proposed. Blei and Jordan [5] present a variational inference algorithm for Dirichlet process mixtures (DPM). They also present experiments that compare the algorithm to Gibbs sampling algorithms for DPM of Gaussian and present an application to a large-scale image analysis problem. Gorur and Rasmussen [6] have proposed the choice of conjugate and non-conjugate base distributions on a particular class of DPM models which is widely used in application, the Dirichlet process Gaussian mixture model (DPGMM). They compare computational efficiency and modeling performance of DPGMM defined using a conjugate and a conditionally conjugate distributions. Yang [7] has designed a shared parts latent topic model with Dirichlet process to share mixture components between categories in image classification. In his paper, the number of components is unknown and is to be inferred from the train set, the DP is introduced into the model to provide a nonparametric prior for the number of mixture components within each category. However, when classifying a document or image, the DP model does not satisfy the properties that each category should be shared the whole topics with each other. In other words, we consider the problem of the modeling of relationships among sets of documents or images. Each document or image can be represented by a number of latent clusters or topics, where a topic is generally modeled as a probability distribution on words chosen from some basic vocabulary or feature vectors extracted from image dataset. When we consider a clustering subject of such documents or images, we may wish to allow topics to be shared among the documents in the corpus.

Hence, to address this problem, the hierarchical Dirichlet process (HDP), a new stochastic model that meets all properties considered until now has been proposed. Teh et. al. [8] consider set of Dirichlet process, one for each group, where the

well-known clustering property of the Dirichlet process provides a nonparametric prior for the number of mixture components within each group. They describe two Markov chain Monte Carlo sampling schemes for the HDP mixture model. Teh et. al. [9] derived the first variational algorithm to deal with the HDP and deal with hyper-parameters of Dirichlet variables. Fox et. al. [10] proposed a Bayesian nonparametric approach to speaker dialogrization that builds on the HDP hidden Markov model. They develop a sampling algorithm that employs a truncated approximation of the DP to jointly resample the full state sequence, greatly improving mixing rates. Wang et. al. [11] proposes an online variational inference algorithm for the HDP, an algorithm that is easily applicable to massive and streaming data.

In this paper, we will consider the HDP-GMM model that is most suitable for images clustering. First, we introduce the probabilistic structure of HDP-GMM, and consider the variational Bayesian inference method to estimate the posterior distribution for the hidden variables and parameters required by this model. Second, we examine the extraction method of various local patches features from given image, which can accurately represent the colors and contents of images. Next, we have trained the HDP-GMM using the extracted patch features, and then present a scheme to classify a given image into the appropriate category or topic by using trained model. Finally, we have applied our model to classify various images datasets, and we have showed the superiority of the proposed method using several evaluation measures for classification method.

2 HDP-GMM and VB-Inference

2.1 Hierarchical Dirichlet Processes and Gaussian Mixture Model

A two-level hierarchical Dirichlet Process (HDP) is a collection of Dirichlet processes that share a base distribution G_0, which is also drawn from a DP. The process defines a set of random probability measures $(G_j)_{j=1}^{\infty}$, one for each group, and a global random probability measure G_0. The global measure G_0 is distributed as a Dirichlet process with concentration parameter γ and base probability measure H:

$$G_0 | \gamma, H \sim DP(\gamma, H)$$

and the random measures $(G_j)_{j=1}^{\infty}$ are conditionally independent given G_0, with distributions given by a Dirichlet process with base probability measure G_0:

$$G_j | \alpha, G_0 \sim DP(\alpha, G_0).$$

Hence, the hyper-parameters of the HDP consist of the baseline probability measure H, and the concentration parameters γ and α_0. The baseline H provides the prior distribution for the parameters $(\phi_j)_{j=1}^{\infty}$. The distribution G_0 varies around the prior H, with the amount of variability governed by γ. The actual distribution G_j over the parameters ϕ_j in the jth group deviates from G_0, with the amount of variability governed by α.

Here, the construction of HDP is actually formed by twice applying Sethuraman's stick-breaking construction of the Dirichlet process (DP). For the top-level (corpus-level) DP draw, since the global measure G_0 is distributed as a Dirichlet process, it can be expressed by using the following stick-breaking representation:

$$\beta'_k \mid \gamma \sim \text{Beta}(1,\gamma),\ \beta_k = \beta'_k \prod_{l=1}^{k-1}(1-\beta'_l)$$
$$\theta_k \mid H \sim H\ ,\quad G_0 = \sum_{k=1}^{\infty}\beta_k \delta_{\theta_k}.$$

Thus, G_0 is discrete and has support at the atoms $\theta = (\theta_k)_{k=1}^{\infty}$with weight $\beta = (\beta_k)_{k=1}^{\infty}$. For the bottom-level (document-level) DP draw, since the each measure G_j is distributed as a Dirichlet process, it can be also expressed by using the following stick-breaking representation:

$$\pi'_{jt} \mid \alpha_0, G_0 \sim \text{Beta}\left(\alpha_0\beta_k, \alpha_0\left(1-\sum_{l=1}^{k}\beta_l\right)\right)$$
$$\pi_{jt} = \pi'_{jt}\prod_{l=1}^{t-1}(1-\pi'_{jl})$$
$$G_j = \sum_{t=1}^{\infty}\pi_{jt}\delta_{\theta_k}\ ,$$

where $\theta = (\theta_k)_{k=1}^{\infty}$ are the same atoms as global measure G_0.

A HDP can be used as the prior distribution over the factors for grouped data. For each j, let $(\phi_{jn})_{n=1}^{N_j}$ be i.i.d random variables distributes as G_j. Eachϕ_{jn} is a factor corresponding to a single observation x_{jn}. The likelihood is given by:

$$\phi_{jn} \mid G_j \sim G_j\ ,\ x_{jn} \mid \phi_{jn} \sim F(\phi_{jn}).$$

Furthermore, since each random variable ϕ_{jn} is distributed as a distribution G_j, it takes on atom θ_k with probability π_{jk}. Here, we may denote this using an indicator variable z_{jn} , which takes on positive integer values and is distributed according to $\pi_j = (\pi_{jt})_{t=1}^{\infty}$. Then, we have $\phi_{jn} = \theta_{z_{jn}}$, and the likelihood function for the observed values is given as

$$z_{jn} \mid \pi_j \sim \pi_j\ ,\ x_{jn} \mid z_{jn},\ (\theta_k)_{k=1}^{\infty} \sim F(\theta_{z_{jn}})\ .$$

Here, we assume the distribution for a mixture component of observed values x_{jn} as a Gaussian distribution with a full covariance. That is,

$$F(x_{jn} \mid z_{jn} = k, \theta_k = (\mu_k, \Lambda_k^{-1})) \sim \mathcal{N}(x_{jn} \mid \mu_k, \Lambda_k^{-1}),$$

where Λ_k is the inverse covariance. And the parameters μ_k and Λ_k^{-1} are further specified by a Gaussain distribution prior and a Wishart distribution prior with additional hyper-parameters, respectively. That is,

$$p(\mu_k) \sim \mathcal{N}(m_0, (\beta_0\Lambda_0)^{-1}),$$
$$p(\Lambda_k) \sim \mathcal{W}(W_0, \upsilon_0).$$

Finally, we also place Gamma priors over the other hyper-parameters α and γ of the Hierarchical Dirichlet process mixture model as follows.

$$\alpha \mid a_\alpha, b_\alpha \sim \text{Gamma}(\alpha \mid a_\alpha, b_\alpha)$$
$$\gamma \mid a_\gamma, b_\gamma \sim \text{Gamma}(\gamma \mid a_\gamma, b_\gamma).$$

In this paper, we focus on an application of the HDP-GMM to modeling a collection of images. For the top-level, the global topics θ_k are distributions on a collection of M cluster mean vectors of features extracted from all images. The global topic weights $\beta = (\beta_k)_{k=1}^{\infty}$ are still drawn from a stick-breaking prior. For each image j, image-specific topic frequencies are drawn $\pi_j = (\pi_{jt})_{t=1}^{\infty}$ given from $DP(\alpha, G_0)$. Then for each mean vector index n in image j, a topic indicator is drawn z_{jn} from Multi (π_j), and finally a feature vector is drawn x_{jn} from a distribution $F(x_{jn} \mid \theta_{z_{jn}})$. The graphical representation of HDP-GMM considered up to now is given as Figure 1.

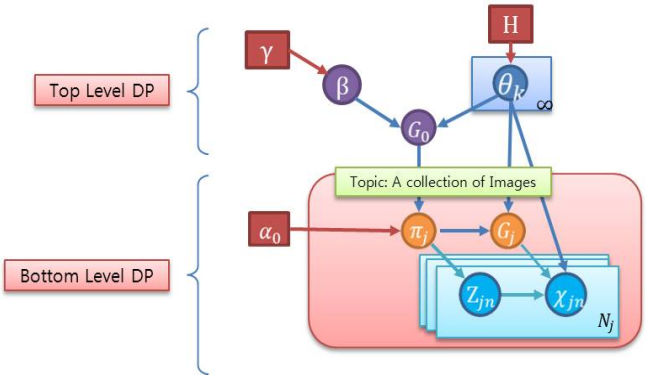


Fig. 1. Graphical representation of HDP-GMM

2.2 Variational Bayesian Inference

Here, we will consider the variational Bayesian inference method to approximate the posterior of the latent variables for HDP-GMM. Variational methods have their root in the early work of calculus of variations, which concerns functional derivatives about how the value of functional changes with respect to infinitesimal changes of the input functions. Variational inference approximates the posterior over the hidden variables by positing a simpler distribution which is optimized to be close in Kullback-Leibler (KL) divergence to the true posterior. This problem is approximately solved by optimizing a function equal up to a constant to the KL of interest [11, 12].

3 Image Classification

Here, we will consider image representation as a collection of features extracted from local image patches. And then we will explain how to classify images by using HDP-GMM.

3.1 Extraction of Local Patch Features

In general, it is very difficult to extract feature vectors that accurately represent the color or the contents of the image. We handle this issue by extracting large feature set, so that many regions are likely to be salient. In our model, we will consider three types features; region based feature, corner-like feature and transform-based feature. The region-based features can represent the color characteristics of the image, and the corner-based features can express the structure of the image. And the transform-based features can express the texture characteristics of image [13-15].

First, the region-based features are generated as follows. We divide an image into hundreds of uniform grids with 8×8 or 16×16 size pixels. Then, we measure intensity values for pixels belonging to each uniform grid, and the histogram feature is constructed with proper number of bins by occurring of intensity values.

Second, the corner region describes the geometry properties around key points in the image. First, we detect salient points or interest points that deliver shape and orientation of a local region. The currently most popular distinguished detectors, which give sufficient performance results, are Harris or Hessian point based detectors, Difference of Gaussian Points (DoG) detector, Entropy Based Salient Region (EBSR) detector, Maximally Stable Extremal Regions (MSER) detector, and Edge Based Regions (EBR) detector and Intensity Based Regions (IBR) detector. Here, we use Harris-Laplace detector to localize points in scale space. Next, we extract the interest feature descriptors which describe the characteristics of the neighboring area around detectors obtained from an image. Here, we have used two types of descriptors such as SIFT or GLOH features which are widely used to express the geometric characteristics of the local region. And then PCA dimensionality reduction technique is applied respectively for two types of descriptor's datasets to reduce the large dimensionality of the data sets, keeping 36 coefficients.

Third, the transform-based feature describes the contrast, uniformity, coarseness characters in the image. Here, we apply the 2D DWT with each square patch to get wavelet coefficients. Next, we take the coefficients in LL sub-band, and then we arrange these coefficients according to their magnitude. Finally, we use proper number of coefficients as texture features.

3.2 Image Classification

We have used two steps in order to perform the image classification. In the first step, we represent each image by a collection of feature vectors extracted from large patches. We have trained the HDP-GMM using these feature vectors as the input vector, and by the trained model, we construct Topics and Sub-topics for a collection of large images. From a given results, we describes each image as a simplex of these discovered topics as well as a multinomial representation of Sub-topics. We have used this topic-simplex and sub-topic multinomial representation of the image at next step to perform image classification.

The first process of image classification can be summarized as follows:

1. For each image, segment it into large number of patches.
2. For each patch, compute a set of feature vectors.
3. Feed this information as input to HDP-GMM.
4. Obtain topic-simplex and sub-topic multinomial representation for image from estimated HDP-GMM.

In the second step, we have used the Bayesian inference method for Dirichlet-Multinomial to classify the j-th image into the k-th topic. We first assume that the j -th image is represented with the occurrence counts of Topics and Sub-topics by the following form:

Table 1. Occurrence matrix of Topics and Sub-topics

	1	2	…	K	Total
	β_1	β_2	…	β_K	1
1	n_{11}	n_{12}		n_{1K}	
2	n_{21}	n_{21}		n_{2K}	
…					
T	n_{T1}	n_{T2}		n_{TK}	
Total	$n_{.1}$	$n_{.2}$		$n_{.K}$	N

We assume the prior distribution for the probability random vector $\theta = (\theta_1, \cdots, \theta_K)$ that an j-th image belongs to one of K topics is the Dirichlet distribution with multinomial parameter vector $\beta = (\beta_1, \cdots, \beta_K)$, and its density function is given by

$$p(\theta = (\theta_1, \cdots, \theta_K)|\ \beta) = \frac{\Gamma(\sum_{k=1}^{K} \beta_k)}{\prod_{k=1}^{K} \Gamma(\beta_k)} \prod_{k=1}^{K} \theta_k^{\beta_k - 1} \quad .$$

Here, the value β_k is a positive real number that determines the form of the distribution, and in our case, we take this value as the probability that j-th image belongs to k-th topic.

And, suppose that a discrete random variable X_i is multinomial distributed with a parameter vector $\theta = (\theta_1, \cdots, \theta_K)$ for $i = 1, \cdots, N$ and $n_{.k}$ is the number of times topic k occurs in N total observations $X = (X_1, \cdots, X_N)$. Then the likelihood function of a discrete random vector $D = (n_{.1}, \cdots, n_{.K})$ given a total number of observation N and a parameter vector $\theta = (\theta_1, \cdots, \theta_K)$ is multinomial distributed with parameters N and θ, and its function is given by

$$p(D = (n_{.1}, \cdots, n_{.K})|\ N, \theta) = \frac{N!}{\prod_{k=1}^{K} n_{.k}!} \prod_{k=1}^{K} \theta_k^{n_{.k}} \quad .$$

Therefore, after observing data with counts $\{n_{\cdot k}\}$, we have the posterior distribution for the parameters as being

$$p(\theta|\, D,\beta) \propto p(\theta|\beta)p(D|\,\theta) = \frac{\Gamma(\sum_{k=1}^{K}\beta_k)}{\prod_{k=1}^{K}\Gamma(\beta_k)}\prod_{k=1}^{K}\theta_k^{\beta_k-1}\frac{N!}{\prod_{k=1}^{K}n_{\cdot k!}}\prod_{k=1}^{K}\theta_k^{n_{\cdot k}} \propto \prod_{k=1}^{K}\theta_k^{\beta_k+n_{\cdot k}-1},$$

which is the form of a Dirichlet distribution. Hence we know the constant of proportionality immediately from the form of Dirichlet distribution and so we have

$$p(\theta|\, D,\beta) = \frac{1}{B(\beta,n)}\prod_{k=1}^{K}\theta_k^{\beta_k+n_{\cdot k}-1},$$

where $B(\beta,n)$ is the normalized constant of Dirichlet distribution.

Finally, we want to infer the probability of some new data point x. Then, the predictive distribution for the new multivariate observation x is given as

$$p(x|\, D,\beta) = \int\left(\prod_{k=1}^{K}\theta_k^{x_k}\frac{1}{B(\beta,n)}\prod_{k=1}^{K}\theta_k^{\beta_k+n_{\cdot k}-1}\right)d\theta .$$

Hence, the probability of new image x belonging to topics k is given by

$$p(x=k|\, D,\beta) = \int\left(\theta_k\frac{1}{B(\beta,n)}\prod_{k=1}^{K}\theta_k^{\beta_k+n_{\cdot k}-1}\right)d\theta = \frac{\beta_k+n_{\cdot k}}{\sum_{k=1}^{K}\beta_k+n_{\cdot k}}.$$

Thus, we classify a new observation image into topic which make to maximize the posterior probabilities. That is,

$$\text{if } \hat{k} = \max_{1\le k\le K}\frac{\beta_k+n_{\cdot k}}{\sum_{k=1}^{K}\beta_k+n_{\cdot k}}, \text{ then new image classify topic } \hat{k}.$$

4 Experimental Results

Here, we have conducted various experiments to evaluate the classification performance of our method using COIL 20 and Caltex image data sets.

4.1 Coil Image Data

First, we carried out the experiment using the COIL-20 image data in order to evaluate the performance of proposed object detection system. Figure 2 shows example image of the COIL-20 images data set.

In the coarser step, we build initial classifications by considering the local features of images and K-means clustering method. And then in the dense step, we construct fine classification using a Gaussian Dirichlet process mixture model taking initial model as an coarser classification results. Our experimental results show that we can classify properly an example image collection into 10 groups. Our classification result is shown in Fig 3.

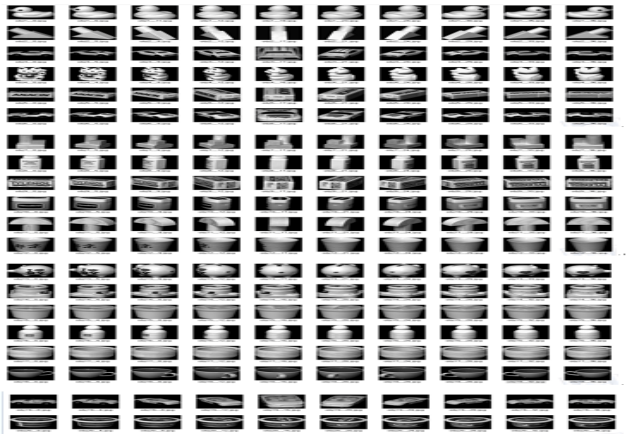


Fig. 2. Columbia University Image Library:COIL-20 dataset: 200 images of twenty categories in each category having 10 similar images

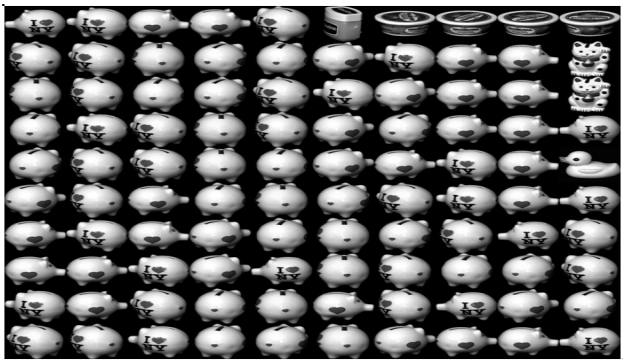


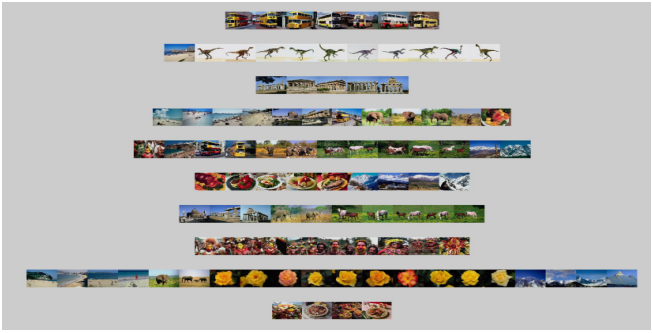
Fig. 3. Classification results for 100 real images of ten categories in each category having 10 similar images

4.2 Caltex Image Data

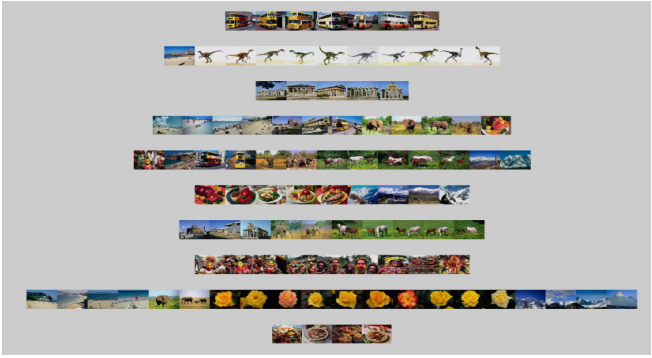
The second real images dataset contains 100 images of 10 objects with 10 images for each class that is collected from Caltex dataset. This data set is showed in the Figure 4.

Here, we test our classification algorithm using local features generated by SIFT, visual word and local image histogram. In the coarser step, we build initial classifications by considering the local features of images and K-means clustering method. And then in the dense step, we construct fine classification using a Gaussian Dirichlet process mixture model taking initial model as an coarser classification results. Our experimental results show that we can classify properly an example image collection into 10 groups. Our classification result is shown in Fig 5.

Fig. 4. Various real images of ten categories in each category having 10 similar images



(a) The coarser classification result: K-Means



(b) The final classification results: K-Means + GDPM

Fig. 5. Classification results for 100 real images of ten categories in each category having 10 similar images

5 Conclusion

In this paper, we present new methods for automatic images classification using the bag of visual and Gaussian Dirichlet process mixture model. Our approach has significant advantage over existing techniques. Considering local content information extracted from image patches, our approach can classify automatically real images without some assumption about number of categories or requiring a priori information about initial structure and it can also generalize better to different image collections.

Acknowledgements. This work was supported in part by the Research Foundation Grant by the Chonnam National University (CNU-2012) and Following are results of a study on the "Leades INdustry-university Cooperation" Project, supported by the Ministry of Education, Science & Technology (MEST).

References

1. Zneit, R.A., Jazar, A.A., Ayyoub, B.: Automatic Color Images Classification Algorithm. International Journal of Computer Science Issues 9(2), 305 (2012)
2. Blei, D.M., Ng, A.Y., Jordan, M.L.: Latent Dirichlet Aloocation
3. Hu, D.J.: Latent Dirichlet Allocation for Text, Images, and Music
4. Permuter, H., Francos, J., Jermyn, I.: A study of Gaussian mixture models of color and texture features for image classification and segmentation. Pattern Recognition 39, 675 (2006)
5. Blei, D.M., Jordan, M.I.: Variational Inference for Dirichlet Process Mixtures. Bayesian Analysis 1(1), 121–144 (2006)
6. Gorur, D., Rasmussen, C.E.: Dirichlet Process Gaussian mixture models: choice of the base distribution
7. Yang, B.: Image classification by latent topic model with Dirichlet Processes
8. The, Y.W., Beal, M.I., Blei, D.M.: Hierarchical Dirichlet Processes. Journal of the American Statistical Association 101(476), 1566–1581 (2006)
9. The, Y.W., Kurihara, K., Welling, M.: Collapsed Variational Inference for HDP
10. Fox, E.B., Sudderth, E.B., Jordan, M.I., Willsky, A.S.: The sticky HDP-HMM: Baysian Nonparametric Hidden Markov Models with persistent states
11. Wang, C., Paisley, J., Blei, D.M.: Online variational inference for the hierarchical Dirichlet processes
12. Sun, S., Xu, X.: Variational Inference for Infinite Mixtures of Gaussian Process with Applications to Traffic Flow Prediction
13. Yang, B.: Image Classification by Latent Topic Model with Dirichlet Process. Journal of Computational Information Systems 7(11), 3803–3810 (2011)
14. Tuytelaars, T., Mikolajczyk, K.: Local Invariant Feature Detectors: A survey. Computer Graphics and Vision 3, 177–280 (2007)
15. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. on PAMI 27(10), 1615–1631 (2005)

A Novel Construction for PEKS Scheme Using Matrix Group*

Tin Q. Phan[1], Van H. Dang[2], and Thuc D. Nguyen[2]

[1] University of Information Technology, VNU HCM City
tinpq@uit.edu.vn
[2] University of Science, VNU HCM City
{dhvan,ndthuc}@fit.hcmus.edu.vn

Abstract. Public key Encryption with Keyword Search (PEKS) scheme enables a sender to send encrypted data to a third party using the receiver's public key. The receiver who owns the private key is able to give the third party the ability of search for some relevant data items by keywords without revealing the content of data and keywords. In this paper, we study the problem of PEKS construction without using bilinear maps. Our construction is based on the action of a matrix group on a set of vectors. The security of our system relies on the difficulty of discrete logarithm problem in the matrix group. Experimental results we obtained indicated that our construction over $GF(2)$ is very efficient in terms of the time complexity.

1 Introduction

Storing data on remote servers offers users much better connectivity than storing data on local computers. Users can access their data from anywhere and at any time. To protect the privacy, data must be encrypted before sending to the database service providers. The problem raised is how to retrieve or search for some data items of this encrypted database. In 2004, Dan Boneh et al. [1] firstly proposed a non-interactive Public key Encryption with Keyword Search scheme called PEKS as an original solution for this problem. A non-interactive PEKS involves four polynomial time randomized algorithms:

1. KeyGen(s): takes the security parameter s and returns a pair of keys (A_{pub}, A_{priv}), where A_{pub}, A_{priv} are respectively public and private key.
2. PEKS(A_{pub}, W): produces a searchable encryption S_W of keyword W.
3. Trapdoor(A_{priv}, W'): returns a trapdoor $T_{W'}$ of keyword W' using the private key A_{priv}.
4. Test(A_{pub}, S_W,$T_{W'}$): returns 1 if $W = W'$ and 0 otherwise.

* This work was supported by the project "Design and implementation of FPGA-cryptography IP cores" (No. B2012-18-02TĐ).

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*, Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_43, © Springer-Verlag Berlin Heidelberg 2014

In PEKS scheme, any user can generate PEKS ciphertext and send it to Alice by using A_{pub} but only Alice who owns A_{priv} can generate a trapdoor to search. Boneh et al. have proved that PEKS implies Identity Based Encryption (IBE) and given some constructions for this scheme. The most significant construction of PEKS uses bilinear maps and its security is based on the Bilinear Diffie-Hellman problem.

1.1 Related Work

Philippe Golle et al. [2] proposed two schemes for the problem of conjunctive keyword search in which single keywords are combined using boolean relations. Both of their constructions are based on bilinear maps. Michel Abdalla et al. [3] fixed some gaps of PEKS scheme in [1] regarding to computational and statistical consistency. In 2007, Boneh and Waters [6] continued to fix the restriction of single keyword search by constructing public key systems which support range queries (e.g. $x \geq a$) as well as subset queries (e.g. $x \in S$). Moreover, these systems allow arbitrary conjunctive queries.

The security of PEKS was also discussed in many studies. Jin Wook Byun et al. [14] assumed that PEKS scheme is vulnerable to offline keyword guessing attack in case of small keyword set. Joonsang Baek et al. [7] cited the problem of curious servers in which the server administrators intentionally keep the copies of trapdoors. Quiang Tang [13] showed that the curious servers can generate a tag for each keyword and do the test with the trapdoor at hand with the assumption such that the keyword set is public and polynomial size.

There were still many studies on public key searchable encryption such as the integration of public key data encryption and public key searchable encryption [4], a design for multi-user system [5], delegated search [11], etc. Most of works in this area are based on bilinear maps except [9, 10, 17]. Giovanni Di Crescenzo et al. [9] firstly (to our knowledge) constructed PEKS under an intractability assumption related to quadratic residuosity modulo Blum-Williams integers. Dalia Khader [10] made a construction of PEKS not using pairings but using K-Resilient IBE which was introduced in [15]. Nevertheless, both Giovanni Di Crescenzo and Dalia Khader have not mentioned about the performance of their construction yet. Thuc D. Nguyen and Van H. Dang firstly introduced another approach using Quasi-Inverse to construct PEKS scheme [17]. In this study, we aim to construct a matrix-based PEKS scheme in the hope that using the linear matrix operations makes our systems efficient in practical applications.

1.2 Our Contribution

We propose a novel implementation for PEKS scheme in [1] without using bilinear maps. Our construction is based on the action of a matrix group on a set of vectors. The security of our system relies on hardness assumptions related to the subset sum and discrete logarithm problem. We also evaluate the performance of our construction and show that it is very efficient in terms of the time complexity.

The paper is divided into five sessions: Session 1 for brief introduction, Session 2 for some backgrounds related to matrix group and the discrete logarithm problem. We define our construction in Session 3 and discuss about its security and performance in Session 4. Session 5 is for the conclusion.

2 Preliminaries

2.1 Action of the Matrix Group on a Nonempty Vector Set

In algebra, the action of a group on a set is defined as below:

Definition 1. *Let G be a group and S be a nonempty set. Then G is said to act on S if there is a map from* $G \times S$ *to S, written* $(g, s) \mapsto gs$*, satisfying:*

1. $es = s$ *for all* $s \in S$.
2. *For all* $g, h \in G$ *and* $s \in S$, $(gh)s = g(hs)$.

Proposition 1. *Let G be a group of* $n \times n$ *matrices and S be a nonempty set of* n *-dimensional vectors. Let* $(A, x) = xA$ *or* $(A, x) = Ax^T$ *be the normal multiplication of* $x \in S$ *and* $A \in G$*. Then* (A, x) *is a group action.*

Proof. $(I, x) = xI = x, \forall x \in S$ and I is $n \times n$ indentity matrix.
Let $y = (B, x) = xB = (y_1, y_2, \dots, y_n)$, where $y_i = \sum_{j=1}^{n} x_j b_{ji}$.
Let $z = (A, y) = yA = (z_1, z_2, \dots, z_n)$, where $z_i = \sum_{k=1}^{n} y_k a_{ki} = \sum_{k=1}^{n} \left(\sum_{j=1}^{n} x_j b_{jk}\right) a_{ki} = \sum_{j=1}^{n} x_j \sum_{k=1}^{n} b_{jk} a_{ki} = \sum_{j=1}^{n} x_j (ab)_{ji}$. Thus, $z = (AB, x) = x(AB)$. The proof for group action $(A, x) = Ax^T$ is similar to $(A, x) = xA$.

Proposition 2. *Let* $(A, x) = xA$ *and* $(B, y) = By^T$ *be the group actions. Then* $(xA)(By^T) = x(AB)y^T$.

Proof. Let $z = xA = (z_1, \dots, z_n)$, where $z_i = \sum_{j=1}^{n} x_j a_{ji}$.
Let $t = By^T = (t_1, \dots, t_n)^T$, where $t_i = \sum_{k=1}^{n} b_{ik} y_k$.
Then $zt = \sum_{l=1}^{n} z_l t_l = \sum_{l=1}^{n} \left(\sum_{j=1}^{n} x_j a_{jl}\right) \left(\sum_{k=1}^{n} b_{lk} y_k\right) = \sum_{j=1}^{n} \sum_{k=1}^{n} x_j \left(\sum_{l=1}^{n} a_{jl} b_{lk}\right) y_k$ $= \sum_{j=1}^{n} \sum_{k=1}^{n} x_j (ab)_{jk} y_k$. Thus, $(xA)(By^T) = x(AB)y^T$. This completes the proof.

2.2 Group Generated by a Companion Matrix

We recall some basic concepts from linear algebra. Let $q = p^m$ be a prime power then $\mathbb{F}_q$ will denote the finite field of order q. Let $f(x) = c_0 + c_1 x + \cdots + x^n$ be a monic polynomial of degree n in $\mathbb{F}_q[x]$. The companion matrix C of f has the form:

$$C = \begin{bmatrix} 0 & 0 & \dots & 0 & -c_0 \\ 1 & 0 & \dots & 0 & -c_1 \\ 0 & 1 & \dots & 0 & -c_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -c_{n-1} \end{bmatrix}$$

It is well-known that the characteristic polynomial of C is $f(x)$. C is non-singular if and only if $f(0) = \det(A - 0 \cdot I) = \det(A) = c_0 \neq 0$.

We denote by $\lambda_1, \ldots, \lambda_s$ the s distinct *eigenvalues* of C and by $n_1, \ldots, n_s$ their respective *algebraic multiplicities*. This means

$$f(x) = (x-\lambda_1)^{n_1}(x-\lambda_2)^{n_2}\cdots(x-\lambda_s)^{n_s} \text{ with } n_1 + \cdots + n_s = n.$$

Let E denote the *splitting field* of $f(x)$ over $\mathbb{F}_q$ and $GL(n,E)$ is the set of all non-singular $n \times n$ matrices over E under the matrix multiplication. A Jordan matrix is a direct sum (the $\oplus$ operator) of Jordan blocks. Let $J_{n_i}(\lambda)$ denote the Jordan block of size $n_i \times n_i$ corresponding to λ_i. There always exists a matrix $Q \in GL(n,E)$ such that $Q^{-1}CQ = J_C$, where J_C is the Jordan matrix corresponding to C.

$$J_C = J_{n_1}(\lambda_1) \oplus J_{n_2}(\lambda_2) \oplus \ldots \oplus J_{n_s}(\lambda_s)$$

We are interested in the case that $f(x)$ is irreducible. Hence, $f(x) = 0$ has n distinct roots $\lambda, \lambda^q, \ldots, \lambda^{q^{n-1}}$ in E. Thus, the Jordan matrix J_C has the form:

$$J_C = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

From those results, we obtain the following:

Prososition 3. *Let $f(x) = c_0 + c_1x + \cdots + x^n$ be a monic polynomial of degree n in $\mathbb{F}_q[x]$. If $f(x)$ is irreducible and $q^n - 1$ is a prime, the companion matrix of f will have the order of $q^n - 1$.*

Proof. It follows the group theory, if $Q^{-1}CQ = J_C$ then $ord(J_C) = ord(C)$. Since J_C is a the diagonal matrix, $ord(J_C) = lcm(ord(\lambda_1), ord(\lambda_2), \ldots, ord(\lambda_n))$. Furthermore, all the roots of $f(x)$ has the same order then $ord(J_C) = ord(\lambda_1)$. The smallest extension field containing all the roots of f(x) is $E = \mathbb{F}_{q^n}$. Consequently, $ord(\lambda_1) \mid (q^n - 1)$. If $q^n - 1$ is prime, $ord(C) = q^n - 1$. This completes the proof.

Let G denote the cyclic group generated by the companion matrix C, then G has the order of q^n–1.

$$G = \{I, C, C^2, \ldots, C^{q^n-2}\}$$

Proposition 4. *G is an abelian group under the matrix multiplication.*

Proposition 5. *Let G be a group generated by $n \times n$ companion matrix and $S \setminus \{0^{1\times n}\}$ be a nonempty set of n-dimensional vectors. Let $(A, x) = xA$ be the matrix multiplication of $x \in S$ and $A \in G$. Then action (A, x) of G on S is faithful.*

Proof. We need to prove that there are no group elements $X \neq I$ such that $xX = x$ for all $x \in S \setminus \{0^{1\times n}\}$. Suppose that there exists $X \in G$ such that $xX = x$. It implies that $xX^k = x$ for any integer k. Since G is a cyclic group, X is also a generator. Thus, for all $Y \in G$ then $xY = x$. One can easily point out an example to make above assumption fail. Therefore, only $I \in G$ such that $xI = x$ for all $x \in S \setminus \{0^{1\times n}\}$. This completes the proof.

Equivalently, Proposition 5 implies that there do not exist $A, B \in G$ and $A \neq B$ such that $xA = xB$ for all $x \in S \setminus \{0^{1\times n}\}$. Indeed, suppose that $xA = xB$ then $x = xBA^{-1}$. Since $BA^{-1} \neq I$, this violates the Proposition 5.

2.3 The DLP in GL(n, p)

The problem of determining a given α and $\beta = \alpha^a$ is called the *discrete logarithm problem* (DLP). The best algorithms that are known for solving the DLP in an arbitrary group G are the exponential square root attacks. In this paper, we are interested in a matrix group G with the size of N. The exponential square root attacks have the running time in $O(N^{\frac{1}{2}}\log(N))$ and space requirement in $O\left(N^{\frac{1}{2}}\right)$ group elements. The weakness of these attacks in a matrix group is the problem of storing approximately $N^{1/2}$ $(n \times n)$ matrices. Therefore, in a matrix group of the large size, exponential square root attacks can be avoided.

It is very well-known that the DLP in $GL(n, q)$ can be reduced in probabilistic polynomial time to the DLP in extensions of $\mathbb{F}_q$ [12]. The core of algorithms in [12] is based on the factorization of characteristic polynomial that

$f(x) = {f_1}^{e_1}{f_2}^{e_2} \ldots {f_s}^{e_s}$ (${f_i}^{e_i}$ is an irreducible polynomial of degree e_i).

The discrete logarithms are solved individually on extension fields $\mathbb{F}_{q^{e_i}}$, $1 \leq i \leq s$. After that, the Chinese remainder theorem is used to obtain the exponent. If we use group G which is generated by $n \times n$ companion matrix of an irreducible polynomial, algorithms in [12] can only reduce the DLP in G to the DLP in $\mathbb{F}_{q^n}$ whose size is analogous to the size of G ($q^n - 1$ by Proposition 3). Note that the DLP in $\mathbb{F}_{q^n}$ is nontrivial since $\mathbb{F}^*_{q^n}$ is a cyclic group of prime order. Suppose that there do not exist dominant algorithms to solve the DLP in $\mathbb{F}_{q^n}$ and let G be a group generated by $n \times n$ companion matrix of an irreducible polynomial, our scheme relies on the following assumption:

Assumption 1. *Solving DLP by reducing it into $\mathbb{F}_{q^n}$ is not easier than solving it in G.*

3 Public-Key Encryption with Keyword Search Construction Using Matrix Group

3.1 Proposed Construction

Given a finite field $\mathbb{F}_q$ with $q = p^m$ elements, where p is a prime. In this setting, cryptographic applications are almost related with either the case $p = 2$ or the case $m = 1$. To construct a group as described in session 2.2, we use an irreducible polynomial of degree n with coefficients in $\mathbb{Z}_p$ (the case $m = 1$). We also use a hash function H whose hash values are in $\mathbb{Z}_{p^n-1}$. Our construction consists four algorithms that work as follows:

— KeyGen(n): The input parameter determines the size of the companion matrix. The algorithm picks a random $d \in \mathbb{Z}_{p^n-1}$ and a random $g \in \mathbb{Z}_p^{\lceil\sqrt{n}\rceil \times n}$. It then computes $h = gC^d$ and outputs $A_{pub} = [g, h]$ and $A_{priv} = d$.
— PEKS(A_{pub}, W): First compute $w = H(W)$ and pick a random $r \in \mathbb{Z}_{p^n-1}$. Output $S_W = [hC^w C^r, gC^r g^T]$.
— Trapdoor(A_{priv}, W'): Compute $w' = H(W')$ and output $T_{W'} = C^{-d}C^{-w'}g^T$.
— Test($S_W, T_{W'}$): Let $S_W = [D, E]$. Test if $DT_{W'} = E$. If so, output 1; if not, output 0.

3.2 Consistency

Since we use the same hash function H, if $W = W'$ then $w = w'$. From the Test function, $DT_{W'} = (hC^w C^r)(C^{-d}C^{-w}g^T)$. By Proposition 2, $T_{W'} = h(C^w C^r C^{-d} C^{-w})g^T$. By Proposition 4, the matrix multiplication in G is commutative then $DT_{W'} = h(C^{-d}C^r)g^T$. Since $h = gC^d$, $DT_{W'} = gC^r g^T$.

If $W \neq W'$ the $w \neq w'$. Hence, $DT_{W'} = g(KC^r)g^T$ with $K \neq I$. By Proposition 5, $gKC^r \neq gC^r$ then $g(KC^r)g^T = g(C^r)g^T$ with the negligible probability of p^{-n}.

4 Security and Performance

4.1 Security Analysis

Definition 2. (*SUBSET-SUM*) *Given a pair* (S, t), *where* S *is a set* $\{s_1, s_2, \ldots, s_N\}$ *of integers and* t *is an integer. Whether there exists a subset* $S' \subseteq S$ *such that* $t = \sum_{s \in S'} s$.

Proposition 6. *Subset sum problem is NP-Complete.* [16]

Definition 3. (*HP-1*) *Given a pair* (x, y), *where* $x \in \mathbb{Z}_p^{1\times n}$ *and* $y = xA$ *with* $A \in \mathbb{Z}_p^{1\times n}$, *find* A.

Proposition 7. *HP-1 is a hard problem.*
Proof. Let $x = (x_1, x_2, \ldots, x_n)$ and $A = [a_1 \quad a_2 \quad \cdots \quad a_m]^T$. Then $xA = x_1a_1 + x_2a_2 + \cdots + x_na_n$, where $a_i \in \mathbb{Z}_p^{1\times n}$. If we present a_i as a number with radix p then $a_i \in \mathbb{Z}_{p^n}$.

Now let HP-1-VARIANT be the problem that given a triplet (S, t, HW), where $S = \{s_i = i \mid 0 \leq i \leq p^n - 1\}$, $t = y_0 s_0 + y_1 s_1 + \cdots + y_{p^n-1} s_{p^n-1}$ and $HW = n$ is the Hamming weight of y. Find binary vector $y = (y_0, y_1, \ldots, y_{p^n-1})$.

Since solving *HP-1-VARIANT* is finding n elements subset $\{s_i \in S \mid y_i = 1\}$, *HP-1-VARIANT* implies *SUBSET-SUM*. *HP-1* is event harder than *SUBSET-SUM* because we are required to solve the permutation problem of n elements to obtain A. Hence, *HP-1* is at least NP-Complete. This completes the proof.

Definition 4. (*HP-2*) *Given a triplet* (C, x, y), *where* C *is* $n \times n$ *companion matrix of irreducible polynomial* $f(x)$ *over* $\mathbb{Z}_p$, $x \in \mathbb{Z}_p{}^{1\times n}$ *and* $y = xA$ *with* $A = C^l$, *find* l.

Proposition 8. *HP-2 can be converted to the DLP in the splitting field of* $f(x)$.

Proof. We have $J_C^l = Q^{-1}C^lQ = Q^{-1}AQ \Leftrightarrow QJ_C^lQ^{-1} = A$.

Then $xA = xQJ_C{}^lQ^{-1}$ and hence, $xAQ = xQJ_C{}^l = x[\mu_1 \quad \cdots \quad \mu_n]\begin{bmatrix} \lambda_1{}^l & 0 & \cdots & 0 \\ 0 & \lambda_2{}^l & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n{}^l \end{bmatrix}$

Clearly, $\lambda_1{}^l$ can be obtained by $(xAQ)_1\,(x\mu_1)^{-1}$. Now we can solve the DLP in the splitting field of $f(x)$ to obtain l. This completes the proof.

Equivalently, we can also prove Proposition 7 and Proposition 8 in case $y = Ax^T$. Therefore, we will mention *HP-1* and *HP-2* in both $y = xA$ and $y = Ax^T$. In our construction, we have $\sqrt{n}$ equations from $y = xA$ (or $y = Ax^T$), where $x \in \mathbb{Z}_p^{[\sqrt{n}]\times n}$. Since $\sqrt{n} < n$, finding A also depends on *HP-1* and *HP-2*.

Assumption 2. *Given public key* $A_{pub} = [g, h = gC^d]$ *and PEKS ciphertext* $S_W = [hC^wC^r,\ gC^rg^T]$, *it is infeasible to discover private key and the information related to keyword* W.

If *HP-1* and *HP-2* are hard, this assumption will hold.

Assumption 3. *Given trapdoor* $T_{W'} = C^{-d}C^{-w'}g^T$, *it is infeasible to discover private key and the information related to keyword* W'.

If *HP-1* is hard, this assumption will hold.

4.2 Performance Evaluation

For the experimental purpose, our construction was implemented over $\mathbb{Z}_2$. The performance evaluated below focuses on the running time of four algorithms. Experiments were done on a personal computer with Core i3 2.4 GHz processor, 4 GB RAM and 64-bit Operating System installed. With $n \in \{\ldots,31,61,89,107,127,521,\ldots\}$ is prime numbers and then $2^n - 1$ is Mersenne primes. We chose respectively $n = 89,107,127$ as the security parameter of our system with the assumption that the computational limitation of current computers is 2^{80}. The running time of KeyGen algorithm is respectively 89, 156, 238 miliseconds; 173, 308, 463 miliseconds for PEKS algorithm; 91, 157, 257 miliseconds for Trapdoor algorithm; and 0.018, 0.022, 0.025 miliseconds for Test algorithm.

PEKS algorithm has the highest running time because it requires two matrix exponentiations which are C^r and C^W. If the keyword set was fixed and small in size, we could precompute C^W for all keyword W. We can also privately store C^r for a set of random r and retrieve C^r randomly to use. These improvements are considerable in removing the matrix exponentiation in PEKS and Trapdoor algorithm. Test is the most efficient algorithm in our construction because it requires only one vector-matrix multiplication. This is meaningful in cases server must serve a lot of clients and the database is large.

5 Conclusion

We have proposed a novel construction for PEKS without using bilinear maps but using the action of a matrix group on a set of vectors. The security of proposed construction is based on the difficulty of discrete logarithm problem in the matrix group. Our experiments have indicated that the proposed construction provides an efficient searching ability. Beside those achievements, we are also facing with drawbacks such as the limitation in choosing system parameter as primes and the ability of reducing the DLP in a matrix group to the DLP in a splitting field.

Acknowlegement. Authors would like to thank Prof. Ton That Tri, Sai Gon University and Prof. Tran Dan Thu, University of Science, HCM VNU for their useful analyses and ideas in our work.

References

[1] Boneh, D., Di Crescenzo, G., Ostrovsky, R., Persiano, G.: Public-key encryption with keyword search. In: Cachin, C., Camenisch, J.L. (eds.) EUROCRYPT 2004. LNCS, vol. 3027, pp. 506–522. Springer, Heidelberg (2004)

[2] Golle, P., Staddon, J., Waters, B.: Secure conjunctive keyword search over encrypted data. In: Jakobsson, M., Yung, M., Zhou, J. (eds.) ACNS 2004. LNCS, vol. 3089, pp. 31–45. Springer, Heidelberg (2004)

[3] Abdalla, M., Bellare, M., Catalano, D., Kiltz, E., Kohno, T., Lange, T., Malone-Lee, J., Neven, G., Paillier, P., Shi, H.: Searchable Encryption Revisited: Consistency Properties, Relation to Anonymous IBE, and Extensions. In: Shoup, V. (ed.) CRYPTO 2005. LNCS, vol. 3621, pp. 205–222. Springer, Heidelberg (2005)

[4] Baek, J., Safavi-Naini, R., Susilo, W.: On the integration of public key data encryption and public key encryption with keyword search. In: Katsikas, S.K., López, J., Backes, M., Gritzalis, S., Preneel, B. (eds.) ISC 2006. LNCS, vol. 4176, pp. 217–232. Springer, Heidelberg (2006)

[5] Hwang, Y.-H., Lee, P.J.: Public key encryption with conjunctive keyword search and its extension to a multi-user system. In: Takagi, T., Okamoto, T., Okamoto, E., Okamoto, T. (eds.) Pairing 2007. LNCS, vol. 4575, pp. 2–22. Springer, Heidelberg (2007)

[6] Boneh, D., Waters, B.: Conjunctive, Subset, and Range Queries on Encrypted Data. In: Vadhan, S.P. (ed.) TCC 2007. LNCS, vol. 4392, pp. 535–554. Springer, Heidelberg (2007)

[7] Baek, J., Safavi-Naini, R., Susilo, W.: Public key encryption with key-word search revisited. In: Gervasi, O., Murgante, B., Laganà, A., Taniar, D., Mun, Y., Gavrilova, M.L. (eds.) ICCSA 2008, Part I. LNCS, vol. 5072, pp. 1249–1259. Springer, Heidelberg (2008)

[8] Tang, Q., Chen, L.: Public-Key Encryption with Registered Keyword Search. In: Martinelli, F., Preneel, B. (eds.) EuroPKI 2009. LNCS, vol. 6391, pp. 163–178. Springer, Heidelberg (2010)

[9] Di Crescenzo, G., Saraswat, V.: Public key encryption with searchable keywords based on Jacobi symbols. In: Srinathan, K., Rangan, C.P., Yung, M. (eds.) INDOCRYPT 2007. LNCS, vol. 4859, pp. 282–296. Springer, Heidelberg (2007)

[10] Khader, D.: Public Key Encryption with Keyword Search based on K-Resilient IBE. In: Gavrilova, M.L., Gervasi, O., Kumar, V., Tan, C.J.K., Taniar, D., Laganá, A., Mun, Y., Choo, H. (eds.) ICCSA 2006. LNCS, vol. 3982, pp. 298–308. Springer, Heidelberg (2006)

[11] Ibraimi, L., Nikova, S., Hartel, P., Jonker, W.: Public-Key encryption with Delegated Search. In: Lopez, J., Tsudik, G. (eds.) ACNS 2011. LNCS, vol. 6715, pp. 532–549. Springer, Heidelberg (2011)

[12] Menezes, A., Wu, Y.-H.: The discrete logarithm problem in GL(n, q). Ars Combinatorica 47, 23–32 (1997)

[13] Tang, Q.: Revisit the Concept of PEKS: Problems and a Possible Solution. Technical Report TR-CTIT-08-54 (2008) ISSN 1381–3625

[14] Byun, J.W., Rhee, H.S., Park, H.-A., Lee, D.-H.: Offline keyword guessing attack on recent keyword search schemes. In: Jonker, W., Petković, M. (eds.) SDM 2006. LNCS, vol. 4165, pp. 75–83. Springer, Heidelberg (2006)

[15] Heng, S.-H., Kurosawa, K.: k-Resilient Identity-Based Encryption in the Standard Model. In: Okamoto, T. (ed.) CT-RSA 2004. LNCS, vol. 2964, pp. 67–80. Springer, Heidelberg (2004)

[16] Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C.: 34.5.5: The subset-sum problem. In: Introduction to Algorithms, 3rd edn. MIT Press and McGraw-Hill (2009) ISBN 0–262–03293–7

[17] Nguyen, T.D., Van Dang, H.: Quasi-inverse Based Cryptography. In: Murgante, B., Misra, S., Carlini, M., Torre, C.M., Nguyen, H.-Q., Taniar, D., Apduhan, B.O., Gervasi, O. (eds.) ICCSA 2013, Part IV. LNCS, vol. 7974, pp. 629–642. Springer, Heidelberg (2013)

An Efficient Outlier Detection Technique in Wireless Sensor Networks

Hongyeon Kim and Jun-Ki Min

School of Computer Science and Engineering,
Korea University of Technology and Education,
Byeongcheon-myeon, Cheonan, Chungnam, Republic of Korea, 330-708
{zenweird,jkmin}@koreatech.ac.kr

Abstract. An outlier in wireless sensor networks is a sensor reading generated by malfunctioned sensor nodes or external environmental effect. Since an outlier is an important indicator of environment change or malfunctioned sensors, many techniques to detect the outliers have been proposed. However, the previous outlier detection techniques mainly depend on the size of data clusters and consume much energy to obtain the outliers. In this paper, we propose an efficient outlier detection technique based on data clustering. In order to construct data clusters of the sensor readings, we adapt the *Pigeonhole Principle*, and then extract the outliers based on the clusters. In our experiment, we demonstrate the efficiency of our proposed technique compared to other outlier detection techniques.

Keywords: Outlier, Monitoring, WSN.

1 Introduction

A wireless sensor network (WSN) is typically composed of a large number of tiny and inexpensive sensors which are scattered in the sensor field to measure quantitative data. Sensor nodes in a WSN generate a large amount of data that must be communicated to the the base station using radio transmission.

Sensor nodes are severely constrained in terms of the computation power, communication bandwidth, and battery power. Among these limitations, the power is of utmost importance since replacing the battery of sensor nodes is either too expensive or impossible. Thus, the energy preservation is a major research issue since it directly impacts the life time of the network. Recent researches have shown that the radio communication is more expensive than the computation or sensing. Thus, many techniques [1–4] have been proposed in order to reduce the communication overhead.

Of particular, since sensor nodes are placed in outdoor for the applications such as disaster monitoring and habitat monitoring, a sensor node can be malfunctioned or sensor readings may be incorrect due to external impact, the severe external environments. In addition, due to the sudden change in external environments, some sensor readings may deviate significantly from the norm sensor readings. These abnormal sensor readings are called the *outliers*.

For example, assume that a wireless sensor network consisting of several sensor nodes is dispatched in a mountain to monitor forest fires. When an outlier value is

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_44,

detected and sent to a forest guard, the forest guard can identify the actual forest fires or he/she can initialize the sensor node if the outlier is generated by the malfunctioned sensor. Thus, the outlier detection is quite important task to detect an event or maintain sensor networks harmoniously.

In this paper, we focus on an energy-efficient outlier detection technique in WSNs. To construct data clusters, we adapt the *Pigeonhole Principle*. By applying cluster width, called the *permission range*, obtained by the *Pigeonhole Principle*, we partition the data space into several clusters. However, if we partition the data space evenly, some sensor readings are identified as outliers even though similar sensor readings of them are detected by sensor nodes. Thus, we partition the data space unevenly based on the location of sensor readings. Then, we identify the outliers according to the user-defined threshold θ which is related to the number of sensor readings in clusters.

2 Related Work

In wireless sensor network, a lot of outlier detection techniques have been proposed. In [5], an outlier detection technique was proposed to collect the outliers with respect to the neighbor sensor nodes. Each sensor node calculates the median of the sensor readings received from the neighbors as well as its readings. Each sensor node computes the mean and μ the standard deviation σ of the differences between its sensor readings and the calculated median. If a sensor reading v's standardized value (= $(v - \mu)/\sigma$) is greater than or equal to the user-defined threshold, it is regarded as an outlier.

In [6], an outlier detection technique based on data clustering, called DC, was proposed. Given a user-defined threshold ε, if the distance between any two sensor readings is less than ε, they become a cluster. Similarly, when the distance between any two clusters is less than ε, they are merged. The inter-cluster distance (ICD) of a cluster to k-nearest clusters is computed to detect outlier clusters. When ICD of a cluster is quite different from the mean of ICDs, it is regarded as an outlier clusters. Then, the information of outlier clusters is broadcasted into WSNs. Thus, it takes much energy.

In [7], an outlier detection technique based on the the Epanechnikov kernel function in order to identify the outliers. Given a value v, each sensor node estimates the number of values around v using the kernel function. If the number of values around value v is less than a user-defined threshold p, a value v is regarded as an outlier. However, to make the kernel function, each sensor node transmits its estimation model to the base station along the routing path. Thus, it consumes a lot of energy.

In [8], an outlier detection technique based on the distance between sensor readings and the estimation deviation was proposed. Each sensor node computes the expected deviation and the average distance between pairs of sensor readings detected within k. If the average distance is greater than the expected deviation, the sensor reading at the current time becomes an outlier. However, when the expectation model is frequently updated, the communication overhead increases.

In [9], an outlier detection technique based on an ellipsoid was proposed. Each sensor node constructs an ellipsoid boundary of sensor readings using the mean and covariance of sensor readings and transmits its ellipsoid boundary to the base station. Then, the base station merges all received ellipsoids, computes the global ellipsoid boundary, and

broadcasts the global ellipsoid to all sensor nodes. With respect to the global ellipsoid boundary, each sensor node identifies the outliers among its sensor readings.

3 Preliminary

In this chapter, we present the basic model of sensor networks and the *Pigeonhole Principle* briefly.

3.1 Sensor Networks

We consider a sensor network consisting of n stationary sensor nodes $\{s_1, s_2, ..., s_n\}$ deployed in a field of interest and the powered base station serving as an access point for users to pose ad hoc queries. We use a routing tree [10] which is frequently used as a primitive to collect sensor nodes. Two nodes capable of bi-directional wireless communication directly are referred to as the neighbors for each other. Each sensor node can broadcasts a message to all of its neighbors (or from a parent to its child nodes) at a time. A simple sensor network using a tree routing is shown in Figure 1. In Figure 1, s_1 to s_3 are the intermediate sensor nodes that have the child sensor nodes, and s_4 to s_9 are the leaf sensor nodes that have no child node.

Sensor nodes generate its readings periodically. A sampling period is known as an *epoch* [10]. To agree on a global time base that allows sensor nodes to start and end each epoch simultaneously, each sensor node executes SMACS protocol [11] or a global time synchronization protocol [12]. Based on global time synchronized, nodes sleep for a certain period of time in each epoch to minimize energy consumption and each sensor node awakes to sample and receive results when its neighbors try to propagate a message.

3.2 Pigeonhole Principle

The *Pigeonhole Principle* [13] addresses that if n pigeons are put into m pigeonholes with $n > m$, then at least one pigeonhole must contains greater than one pigeon. This principle is mainly used when a simple mathematical proof, and it is utilized a lot in every day life. For instance, as shown in Figure 2, if you throw 9 darts on a dartboard which is divided into 8 sectors, at least a sector of the dartboard includes two darts.

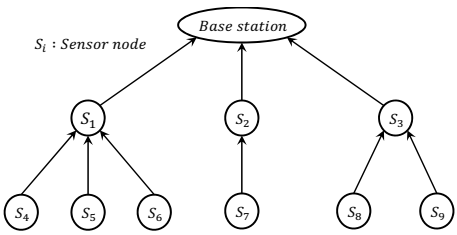


Fig. 1. A simple sensor network

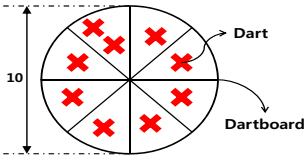


Fig. 2. Example of the Pigeonhole Principle

4 Outlier Detection Based on Clustering

In this section, we present our proposed outlier detection technique which is based on the data clustering in WSNs. In addition, we introduce an efficient data transmission scheme for our outlier detection technique.

4.1 Clustering Technique

Generally, in the outlier detection techniques based on data clustering, the width of clusters is the most important since the number of sensor readings in a cluster is affected by the width of clusters. If the width of clusters is too large, all sensor readings may belong to a single cluster. Otherwise, each cluster may have only a single sensor reading. Thus, the outliers cannot be identified in these cases.

To solve the above problem, in this paper, we adapt the *Pigeonhole Principle* to determine the width of clusters. We regard that pigeonholes and pigeons are the domain of sensor readings and the sensor readings respectively. Thus, when we partition this domain into x number of sub-domains where x is less than the number of sensor readings, at least one sub-domain contains more than two sensor readings.

Suppose given a set of sensor readings $D = \{d_1, d_2, ..., d_n\}$ ($|D| = n$), we can acquire the domain of D as $[min_{d_i \in D}(d_i),\ max_{d_i \in D}(d_i)]$. Then, we obtain the *permission range* (*PR*) using Equation 1.

$$PR = \left(\frac{MAX(D) - MIN(D)}{n - 1} \right) \tag{1}$$

We use *PR* as the width of clusters and identify the outliers. If the number of sensor readings in a cluster is less than or equal to a user-defined threshold θ, we call that a cluster is the *outlier cluster* and the sensor readings in outlier clusters are considered as the outliers.

For instance, given a set of sensor readings D shown in Figure 3, *PR* of D is obtained as $1.7 = (MAX(D) - MIN(D))/(n-1) = (19-2)/(11-1)$. When we partition the

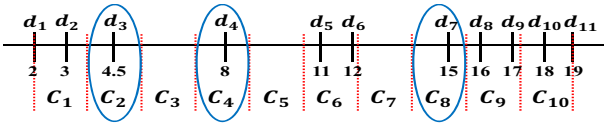


Fig. 3. Equi-partitioning based on permission range

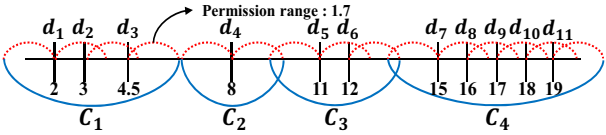


Fig. 4. Non-Equi partitioning based on permission range

domain of D evenly, the clusters represented by dotted lines are presented in Figure 3. If the number of sensor readings in a cluster is less than or equal to θ, a cluster is regarded as an outlier cluster. Assume that a user-defined threshold θ is 1. In Figure 3, C_2, C_4, and C_8 are the outlier clusters (ellipse) and the sensor readings in each outlier cluster are the outliers. Although a sensor reading d in a cluster C_4 is definitely an outlier since a sensor reading having similar values do not exist, the sensor readings c and g are not the outliers since there are b and h which are similar to c and g respectively. In other words, in spite of the difference between any two sensor readings is less than the differences of the others, these sensor readings may belong to separated clusters.

To solve this problem, we propose the *non-equi partitioning* based on the permission range (*PR*). In the non-equi partitioning, a set of clusters is constructed with respect to Definition 1.

Definition 1. *If the difference between two sensor readings d_i and d_j in D is less than or equal to the permission range* PR *obtained by Equation 1, we say d_i* is close to *d_j.*

In our proposed technique, if d_i and d_j are close each other, a single cluster contains d_i and d_j. For instance, as shown in Figure 4, since the difference between sensor readings d_1 and d_2 are close, a cluster C_1 for them is constructed. And then, a sensor reading d_3 is inserted into C_1 since the difference between d_2 and d_3 is less than *PR*. But d_4 is not inserted into C_1. The result of non-equi partitioning is presented in Figure 4. Each cluster consists of $C_1 = \{d_1, d_2, d_3\}$, $C_2 = \{d_4\}$, $C_3 = \{d_5, d_6\}$, and $C_4 = \{d_7, d_8, d_9, d_{10}, d_{11}\}$ respectively. When a user-defined threshold θ is 1, a cluster C_2 is an outlier cluster and a sensor reading d_4 in C_2 is an outlier.

4.2 Clustering Scheme for WSNs

If each sensor node transmits its readings to the base station blindly at each epoch and the base station computes the outliers, each sensor node consumes much energy energy. In this section, we present an efficient data transmission scheme for our outlier detection algorithm. We assumed that all sensor nodes take sensor readings periodically and keep these readings into their local storage for a time window w. Note that when w is 1, each sensor node transmits data at each epoch.

At first, according to the Definition 1, each sensor node constructs the clusters using its sensor readings detected within w. If the number of sensor readings in a cluster is greater than a user-defined threshold θ, all sensor readings in this cluster cannot be the outliers (i.e., non-outlier cluster (*NOC*). Otherwise, all sensor readings in a cluster may be the outliers, and then we call such the clusters *outlier candidate clusters* (*OCC*s).

Along the routing path to the base station, each sensor node transmits *NOC*s and *OCC*s. For *NOC*s, *cluster range*s (*CR*s) is transmitted where *CR* consist of minimum and maximum values of sensor readings in a *NOC*. For *OCC*s, each *OCC* with its sensor reading is sent to the parent node.

When a parent node p received *CR*s of *NOC*s and *OCC*s from its child nodes, p merges them with its clusters. To merge the clusters, we use the following definition.

Definition 2. *Given two cluster ranges* $CR_i = [min_i, max_i]$ *and* $CR_j = [min_j, max_j]$ *where* $(min_i < min_j)$*, if* $min_j - max_i \leq PR$*, we say* CR_i *and* CR_j overlap within PR.

If any two cluster range overlap within *PR*, there are at least two sensor readings which are close and contained in different clusters. Thus, if two clusters C_i and C_j whose cluster ranges CR_i and CR_j overlap within the permission range PR, two clusters are merged into a new cluster whose CR is $[min(min_i, min_j), max(max_i, max_j)]$.

Note that, when two clusters are merged into a new cluster where at least one of them is a *NOC*, the merged cluster cannot be an *OCC* since the number of sensor readings in a *NOC* is already greater than θ. Therefore, we need only *CR* when at least one cluster is a *NOC*.

In contrast, when two *OCC*s are merged into a new cluster, we check whether the number of sensor reading in the new cluster is greater than θ or not. Since each sensor node sends all sensor readings in each *OCC*, we can easily count the number of sensor readings in the new cluster.

Along the routing path from each sensor node to the base station, *CR*s of *NOC*s and *OCC*s are merged and transmitted. Finally, the base station can determine the outliers among the received *OCC*s. Recall that the cluster ranges (*CR*s) rather than the senor readings for *NOC*s are transmitted along the routing paths. Thus, we can reduce the energy consumption of each sensor nodes.

5 Experiments

5.1 Experimental Environments

To evaluate the performance of our proposed algorithm compared with the state-of-the-art algorithm, we used a set of real data which is provide by Intel Berkeley Research Lab [14]. A sensor network consists of 54 sensor nodes, and each sensor node is deployed in 40.5m $\times$ 31m area as shown in Figure 5. The base station is located at the center of the area. We set the default communication distance 7m. The maximum depth and the maximum width of a routing tree in sensor network are 5 and 3 respectively. Sensor readings consist of temperature (Celsius), humidity (%), and illumination (Lux). In three attribute, we used temperature.

As competitors, we implemented the Brute-Force (BF) and an outlier technique based on data clustering (DC) [6]. In BF, each sensor node transmits its readings at each time. In DC, the cluster information are transmitted and merged along the routing paths from each sensor node to the base station. When the distance of two clusters is less than ε, two clusters are merged. The base station computes the inter-cluster distance

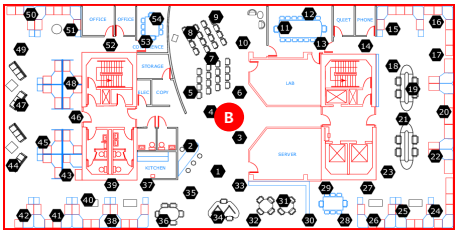


Fig. 5. Placement of sensor nodes [14]

Table 1. Parameters

Parameter	Default value	Comments	Range
p	40	Packet Size	40
w	8	Window size	2, 5, 8, 11, 14
θ	2	Threshold	1, 2, 3, 4, 5

ICD of each cluster using its k-nearest clusters and broadcasts the means of *ICD*s to identify the outlier clusters. We set ε and k for DC are 0.26 and 4 respectively. We named our algorithm as PC.

To compute the energy consumption, we used free space channel model [15]. Under this model, to transmit a l-bits message and a distance c, a sensor expends $E_T(l,c) = E_{T-elec}(l) + E_{T-amp}(l,c) = l * E_{elec} + \xi_{amp} * l * c^2$. And, to receive this message, a sensor expends $E_R(l) = E_{R-elec}(l) = l * E_{elec}$. In this experiment, we set 50 nJ/bit to the electronic circuit constant (E_{elec}) and 100pJ/bit/m^2 to the transmit amplifier constant (ξ_{amp}). We set the size of packet as 40 byte. The parameters used in our experiment are summarized in Table 1.

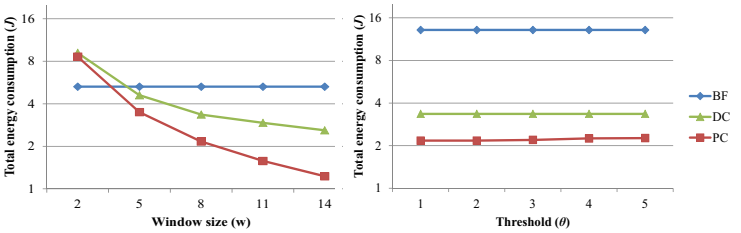
5.2 Experimental Result

To evaluate the energy consumption of each outlier detection algorithms, we run our own simulator for 1000 epoches and plot the total energy consumption.

Figure 6(a) shows the energy consumption of each algorithm varying the window size w. When the size of a window is small (i.e., $w = 2$), BF shows the best performance since BF transmits sensor readings, but DC and PC transmit the information of clusters.

However, as w increases, the performance of PC and DC are improved since the cluster information rather than sensor readings are transmitted. Furthermore, the performance gap between PC and DC increases with increasing w. As w increases, the number of non-outlier clusters (*NOC*s) increases in PC. Thus, the size of data to be transmitted decreases. But, in DC, although the energy consumption is reduce since DC is also based on the data clustering, DC is worse that our proposed PC since PC use the permission range to determine the cluster width. Additionally, in DC, the mean of *ICD*s needs to be broadcasted to obtain the outliers in each sensor node. Our proposed algorithm PC is better than the DC about 58% on the average.

Figure 6(b) shows the energy consumption varying θ. As shown in Figure 6(b), the performances of all techniques are stable in spite of varying θ. The energy consumption



(a) Varying w (b) Varying θ

Fig. 6. Total energy consumption

of DC is constant, since the inter-cluster information has a fixed size. But the energy consumption of PC increases slightly, since the number of sensor readings in an outlier candidate cluster increases slightly. Nonetheless, our proposed technique PC shows the best performance in terms of energy-efficiency. In this experiment, our proposed technique is better than the DC about 51%.

6 Conclusion

In this paper, we present an efficient outlier detection technique in wireless sensor networks. To obtain the appropriate width of clusters, we adapt the *Pigeonhole Principle*. In our proposed technique, each sensor node in WSNs constructs and merges the clusters based on the permission range *PR*. Then, our proposed technique uses two kinds of clusters (*NOC* and *OCC*) in order to detect the outliers and reduce the energy consumption of each sensor. In our experiments with a set of real-life data, we show that our proposed technique outperforms existing techniques significantly.

Acknowledgements. This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (2012R1A1B3003060).

References

1. Abadi, D.J., Madden, S., Lindner, W.: Reed: Robust, efficient filtering and event detection in sensor networks. In: Proceedings of the 31st International Conference on Very Large Data Bases (VLDB), pp. 769–780 (August 2005)
2. Yang, X., et al.: In-network execution of montering queries in sensor networks. In: Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data, pp. 521–532 (June 2007)
3. Madden, S., Franklin, M.J., Hellerstein, J.M., Hong, W.: Tag: A tiny aggregation service for ad-hoc sensor networks. In: 5th Symposium on Operating System Design and Implementation (OSDI) (December 2002)
4. Trigoni, N., Guitton, A., Skordylis, A.: Routing and processing multiple aggregate queries in sensor networks. In: Proceedings of the 4th International Conference on Embedded Networked Sensor Systems(SenSys), pp. 391–392 (October 2006)

5. Weili, W., Xiuzhen, C., Min, D., Kai, X., Fang, L., Ping, D.: Localized outlying and boundary data detection in sensor networks. IEEE Transactions on Knowledge and Data Engineering 19(8), 1145–1157 (2007)
6. Rajasegarar, S., Leckie, C., Palaniswami, M., Bezdek, J.C.: Distributed anomaly detection in wireless sensor networks. In: 10th IEEE Singapore International Conference on Communication systems (ICCS 2006), pp. 1–5. IEEE (2006)
7. Palpanas, T., Papadopoulos, D., Kalogeraki, V., Gunopulos, D.: Distributed deviation detection in sensor networks. ACM SIGMOD Record 32(4), 77–82 (2003)
8. Ghaddar, A., et al.: Algorithm for temporal anomaly detection in wsns. In: 2011 IEEE Wireless Communications and Networking Conference (WCNC), pp. 743–748. IEEE (2011)
9. Suthaharan, S., Leckie, C., Moshtaghi, M., Karunasekera, S., Rajasegarar, S.: Sensor data boundary estimation for anomaly detection in wireless sensor networks. In: 2010 IEEE 7th International Conference on Mobile Adhoc and Sensor Systems (MASS), pp. 546–551. IEEE (2010)
10. Madden, S., Franklin, M.J., Hellerstein, J.M., Hong, W.: Tinydb: an acquisitional query processing system for sensor networks. ACM Transactions on Database Systems (TODS) 30(1), 122–173 (2005)
11. Sohrabi, K., Gao, J., Ailawadhi, V., Pottie, G.J.: Protocols for self-organization of a wireless sensor network. IEEE Personal Communications 7(5), 16–27 (2000)
12. Sundararaman, B., Buy, U., Kshemkalyani, A.D.: Clock synchronization for wireless sensor networks: A survey. Ad Hoc Networks 3, 281–323 (2005)
13. Discrete, P.G.R.: Discrete and Combinatorial Mathematics: An Applied Introduction, 4th edn. Addison Wesley Publishing Company (1998)
14. Lab, I.B.R.: Intel berekely research lab data (2004), `http://db.csail.mit.edu/labdata/`
15. Heinzelman, W.R., Chandrakasan, A., Balakrishnan, H.: Energy-efficient communication protocol for wireless microsensor networks. In: Proceedings of Annual Hawaii International Conference on System Sciences (HICSS) (January 2000)

Discovering High-Quality Paths with Load Balancing in Wireless Sensor Networks

Anh Tai Tran and Myung Kyun Kim

School of Electrical Engineering, University of Ulsan, Korea
taitrananhvn@gmail.com, mkkim@ulsan.ac.kr

Abstract. This paper proposes a reliable routing protocol called LBR for wireless sensor networks, which find high-quality paths with considering load balance among nodes. The LBR protocol transmits messages through a reliable path to reduce the packet loss rate. To the problem of increasing data transmission reliability, our routing protocol uses ETX as a metric for estimating link quality. By using ETX, the LBR protocol can choose good links in the network. In addition to the above, the problem of load balance is also considered in our protocol. By taking this into account, the LBR protocol avoid routing through the congested nodes and balance load among nodes better. This results in increasing the lifetime of the network. The simulation results using the Qualnet simulator showed that the proposed protocol outperforms other protocols in terms of packet delivery ratio, end-to-end delay and the efficiency of load balance.

Keywords: Wireless sensor networks, load balance, link quality, routing protocol.

1 Introduction

Routing in wireless networks has been an interesting research area for many years. In which, the problem of reliable data transmission has become significant. The most commonly used link metric in routing protocols is minimum hop-count [1, 2]. This metric selects a path minimizing hop-count between the source and destination without considering the lossy ratio of wireless links. In static wireless networks, several studies have shown the poor performance of minimum hop-count [3, 4]. Several estimators in terms of link quality have been proposed such as ETX [3], RTT (Per-hop Round Trip Time) [5] and BW (Bandwidth) [6, 7]. Among them, the ETX estimator is shown to be the best metric in static wireless networks [4]. To deal with the problem of reliable data transmission, our routing protocol uses ETX as a link metric to select high-quality routes.

Additionally, the problem of balancing load in wireless networks is one of the appealing topics studied recently [8, 9]. Routing protocols using shortest path can make the center of the network become “crowded”, because many paths transmit data through the center than through the periphery of the network. This issue causes increased congestion and energy consumption for nodes near the center. Our routing

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_45,

protocol aims to increase the lifetime of network and has a good load distribution among nodes. The rest of the paper is organized as follows. Section 2 gives a detailed description of LBR protocol for both reliable communication and load balancing. Section 3 presents the performance evaluation of the proposed protocol and Section 4 concludes the paper.

2 LBR – A Load Balanced Reliable Routing Protocol

2.1 Expected Transmission Count (ETX)

The ETX of a link is the expected number of data transmissions required to send a data packet over that link, including retransmissions [3]. Each node measures the probability that a data packet is successfully delivered to the receiver, known as the forward delivery ratio and denoted as d_f. The probability that an ACK is successfully received by the sender, which is called the reverse delivery ratio, is denoted as d_r. The ETX value of the link is given by $ETX = \frac{1}{d_f \times d_r}$. To compute ETX, each node broadcasts periodically a probe packet of a fixed small size (usually every second). The probe contains the count of probes received from each neighboring node in the last time window (usually in previous 10 seconds). Based on these probes, a node can calculate the delivery ratios of probes on the links to and from its neighbors (d_r and d_f). Since the 802.11 MAC does not retransmit broadcast packets, these counts allow the sender to estimate the number of transmission times to send successfully a unicast packet to the receiver. To illustrate the calculation of ETX, consider two nodes A and B. Node A needs to calculate the probability that a data packet will be successfully transmitted from A to B. Assume that node A received 6 probe packets from B in the previous 10 seconds, while node B received 9 probe packets from A. Thus, the delivery ratio of packets from B to A is 0.6, while the delivery ratio of packets from A to B is 0.9. In the last probe packet, B reported that it received 9 probe packets from A in the previous 10 seconds. Thus, node A knows the forward delivery ratio from A to B is 0.9 (d_f = 0.9). Node A also knows the reverse delivery ratio from B to A is 0.6 (d_r = 0.6). A successful unicast data transfer in 802.11 involves sending the data packet and receiving a link-layer acknowledgment from the receiver. Thus, node A can calculate the probability that the data packet will be successfully transmitted from A to B is 0.9×0.6 = 0.54. The expected number of transmissions before the packet is successfully delivered is 1/0.54 = 1.85. This is the value of the ETX metric for the link from A to B.

2.2 Definitions

The LBR protocol considers both the end-to-end reliability of routes and the load of each node when selecting a path between a source and a destination. The LBR uses ETX as a link quality estimator to measure the packet transmission reliability on the

link. The end-to-end ETX of a path from node S to node D, e2eETX(x,y), is defined as the sum of link ETX values in the path. We define the load of node n, *load(n)*, as the number of data packets that the node transmits and receives in a unit time during the previous window w from the current time. The load factor of node n, *lf(n)*, is defined as the ratio between *load(n)* and max_load. In which, max_load denotes the maximum number of data packets that a node can transmit and receive in a unit time, which is calculated as (*max_data_transmission_rate /data_packet_size*). The remaining load factor of node n, rlf(n), is then defined as follows.

$$rlf(n) = 1 - lf(n) = 1 - load(n)/max\_load \quad (1)$$

The higher remaining load factor of node n means that it has smaller congestion. The LBR protocol uses *rlf(n)* and link-ETX values to find a path which has the smallest end-to-end ETX value and the minimum remaining load factor of the nodes in the path is the highest. Node *n* keeps its *rlf(n)* by calculating the number of data packets it received and transmitted during the previous window.

2.3 Route Discovery

The LBR protocol uses ETX as a link quality estimator and considers the remaining load factor of a node as a criterion for balancing load when selecting a path from a source to a destination.

The route discovery process is performed by exchanging RREQs and RREPs between a source and a destination. When a source node S wants to transmit messages to a destination node D, it tries to set up a path by broadcasting a RREQ packet. A RREQ packet carries (*id, etx, mrf*), where id denotes the ID of the path, and etx denotes the route metric value, which is the sum of link metric values over which the RREQ packet traversed. mrf indicates the minimum remaining load factor of nodes in the path through which the RREQ has been traveled. Each intermediate node maintains a reverse route entry RE[*id*, *p*Hop, *src*ETX, *src*MRF] for a message flow id to store reverse route towards the source originating RREQ. In which, *src*ETX denotes route-ETX value between the source and the intermediate node, and pHop is a previous node towards the source on the path. *src*MRF denotes the minimum remaining load factor of nodes in the path id. Initially, node S broadcasts RREQ[*id, etx* = 0, *mrf* = *rlf*(S)] packet. When a node Y with the remaining load factor rlf(Y) receives a route request packet RREQ[*id*, *etx*, *mrf*] from its neighbor X, it handles as described in Algorithm 1.

When node Y receives RREQ[*id, etx, mrf*] packet from X for a new path *id*, it creates a reverse routing entry for the path id as RE[*id*, *p*Hop = X, *src*ETX = *etx*+ETX(X,Y), *src*MRF = *mrf*]. After that, if node Y receives another RREQ[*id*, *etx*, *mrf*] packet from Z, it compares the routing information in the new RREQ packet with that of the path id in the routing table. If the routing information in the new RREQ packet is better than that of the path id in the routing table, it updates the reverse routing entry for the path id. We determine the better path in terms of two factors. The first one is the load balancing factor (Condition 1). That is, if the minimum remaining load factor of the new RREQ packet (*mrf*) is greater than the existing minimum

remaining load factor in the routing table (*src*MRF) by more than H1 threshold, the path of the new RREQ packet is better than the existing path in the routing table. The second factor is the route-ETX values of the paths (Condition 2). If Condition 1 is not true, LBR compares the route-ETX value of the new RREQ packet with the route-ETX value in the routing table and select the better (smaller) one. In the case of Condition 2, however, we accept the new path in the RREQ packet only when it has smaller route-ETX value and the minimum remaining load factor in the RREQ packet is not lower than the existing one by more than H2 threshold.

Algorithm 1 LBR Route Request

1: When Y receives RREQ[*id*, *etx*, *mrf*] packet from X
2: **If** It is the first RREQ packet of path *id* **then**
3: Y creates a reverse route entry of path *id*
4: RE[*id, p*Hop:= X,
*src*ETX := *etx* + ETX(X,Y), *src*MRF = *mrf*]
5: Y updates RREQ as
6: RREQ[*id*, *etx* := *etx* + ETX(X,Y), *mrf* = min{*mrf*, *rlf*(Y)}]
7: Y rebroadcasts the updated RREQ[*id*, *etx*, *mrf*] and sets the timer ΔdelayRREQ to collect multiple RREQs
8: **else if** a reverse entry RE[*id*, *p*Hop, *src*ETX, *src*MRF] exists **then**
9: **if** (*mrf* – *src*MRF ≥ H1) OR ▷ *Condition 1*
((*etx*+ETX(X,Y) < *src*ETX) ▷ *Condition 2*
AND (*mrf* + H2 ≥ *src*MRF)) **then**
10: Y updates the reverse entry of path *id*
11: RE[*id*, *p*Hop:=X,
*src*ETX := *etx*+ETX(X,Y), *src*MRF := *mrf*]
12: Y updates and keeps RREQ
13: RREQ[*id*, *etx* := *etx* + ETX(X,Y),
mrf := min{*mrf*, *rlf*(Y)}]
14: When ΔdelayRREQ becomes 0, Y rebroadcasts RREQ.
15: **else**
16: Y drops the new RREQ packet
17: **end if**
18: **end if**

If the path of the new RREQ packet is better than the existing one in the routing table in terms of Condition 1 or Condition 2, node Y updates its routing table. After that, it has to rebroadcast the new RREQ packet to its neighbors to notify the better route. However, if node Y rebroadcasts the new RREQ packet immediately, the number of RREQ broadcast packets can be large. Thus, we use ΔdelayRREQ timer to reduce the number of RREQ broadcast packets. After broadcasting the first RREQ packet, node Y sets ΔdelayRREQ timer and collects all the RREQ packets arriving during that time, and selects the best route and rebroadcasts it when ΔdelayRREQ timer expires. The value of ΔdelayRREQ timer can affect the performance of the LBR

protocol. If it is set to small value, each node cannot receive enough number of RREQ packets during that time, so the number of RREQ packets can increase. If it is given a large value, each node can select the best route by collecting enough number of RREQ packets, but the route establishment delay can increase. On receiving the first RREQ packet, the destination sets ΔdelayDEST timer, and collects all the RREQ packets arriving during that time, and selects the best route and transmits a RREP packet through the reverse of the selected path when ΔdelayDEST timer expires.

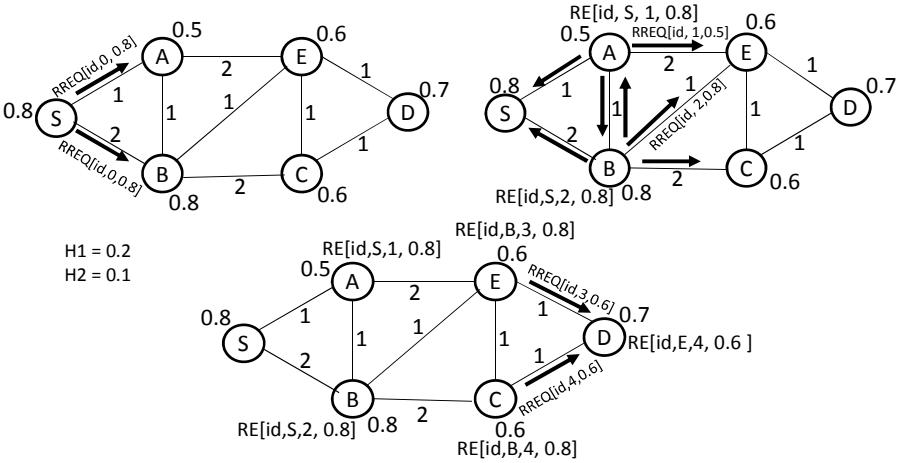


Fig. 1. LBR route discovery example

2.4 Route Discovery Example of LBR

A network with ETX value in each link and the remaining load factor in each node is given as shown in Fig. 1. For example, the ETX value of link (A-E) is 2 and the remaining load factor of node A is 0.5. The parameters H1 and H2 are chosen to be 0.2 and 0.1, respectively. The source S broadcasts a RREQ[id, *etx* = 0, *mrf* = 0.8] packet to find a route to the destination D. In which, the route-ETX value is initiated to be 0 (*etx* = 0) and the minimum remaining load factor is 0.8 (equal to the remaining load factor of node S). Based on the Condition 1 and Condition 2, LBR chooses routes that the route-ETX value (*etx*) is small and the minimum remaining load factor (*mrf*) is large. This means the routes have high throughput and low congestion.

When node A and B receive RREQ[*id*, 0, 0.8] from S, A and B create the reverse route entries for the path *id*, RE[*id*, S, 1, 0.8] (*p*Hop := S, *src*ETX := *etx* + ETX(S-A) = 0 + 1 = 1, *src*MRF := *mrf* = 0.8) and RE[*id*, S, 2, 0.8], respectively. After that, A and B update and rebroadcast the RREQ packet immediately, RREQ[*id*, 1, 0.5] (*etx* := *src*ETX = 1, *mrf* := min(*mrf*,*rlf*(A)) = min(0.8, 0.5) = 0.5) and RREQ[*id*, 2, 0.8] respectively.

At node E, assume that it first receives the RREQ[*id*, 1, 0.5] packet from node A and then the RREQ[*id*, 2, 0.8] from node B. For the first RREQ, node E create a reverse entry RE[*id*, A, 3, 0.5]. When node E receive the RREQ from B, it checks

Condition 1 and Condition 2 to compare the quality of the two paths, the path of the new RREQ packet (S-B-E) and the existing path in the routing table (S-A-E). Because the Condition 1 is true (the minimum remaining load factor of the new route is larger than that of the old route by H1 threshold, 0.8 – 0.5 > H1= 0.2), node E updates new reverse route RE[*id*, B, 3, 0.8].

The process of handling the route request packet at other nodes is similar above. Finally, node D chooses the best route S-B-E-D with the route-ETX value of 4 and the minimum remaining load factor of 0.6.

3 Performance Evaluation

In this section, we used the Qualnet simulator [10] to evaluate the performance of our protocol and compared with the AODV [2] and MRFR [11] protocols. The AODV protocol uses shortest paths to transmit data.

The MRFR protocol is also a reliable routing protocol, which uses ETX as a link quality metric. MRFR uses routes that have the smallest route-ETX value without considering load balance. Our simulations were implemented on a static 125-node network of 1500m*1500m. In all simulations, we set up five simultaneous flows with sources and destinations chosen randomly. The results of almost all simulations were calculated as the average of the five flows. The data rate of each node is 5.5 Mbps. The data packet size is 512 bytes. The MAC layer model is 802.11 and the physical layer model is 802.11b. We use the transmission power of 15dBm (32mW) for all of the nodes. At the application layer, the constant bit rate (CBR) traffic is used. Source nodes of each flow send data packets from 1 CBR (1 packet per second) to 14 CBR (14 packets per second), depending on the simulation. During the first 30 seconds, the nodes just broadcast probes to measure link metrics and establish network stability. At the time of 30 seconds, the LBR protocol uses ETX to discover routes for the five flows. After routes have been established, the five flows start to transmit data. In most of the simulations, the total running time is 240 seconds. The parameters H1 and H2 of the LBR protocol were chosen as 0.2 and 0.1, respectively.

3.1 End-to-End Packet Delivery Ratio

Fig. 2 compare the packet delivery ratio of the routing protocols. The simulation result was calculated as the average of five flows. We measured the delivery ratio at different data loads, varying from 2 CBR to 14 CBR.

The results show that LBR has the best performance compared to the two other protocols. Both our protocol and MRFR obtain high delivery ratios (almost above 98%). Because both MRFR and LBR use ETX as a link metric for finding high quality paths, routes discovered by the two protocols are reliable. While our protocol and MRFR have high delivery ratio at the different cases of load, the AODV protocol has a low delivery ratio when the data load is increased. For example, at a load of 14 CBR, MRFR and LBR have approximately delivery ratios of a 0.97, while delivery ratio using AODV is 0.83. This presents an performance improvement of 14%. AODV uses shortest paths for routing; however, it does not consider link loss ratios.

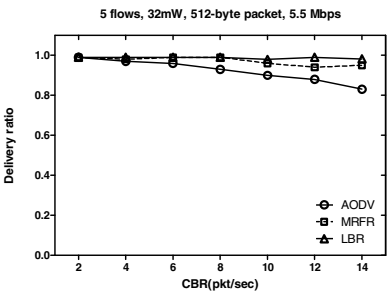


Fig. 2. The end-to-end packet delivery ratio

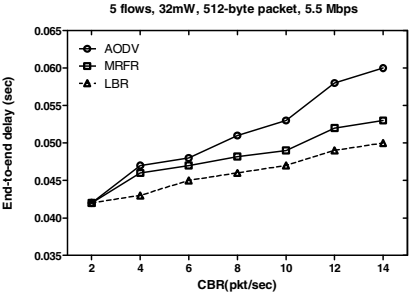


Fig. 3. Average end-to-end packet delay

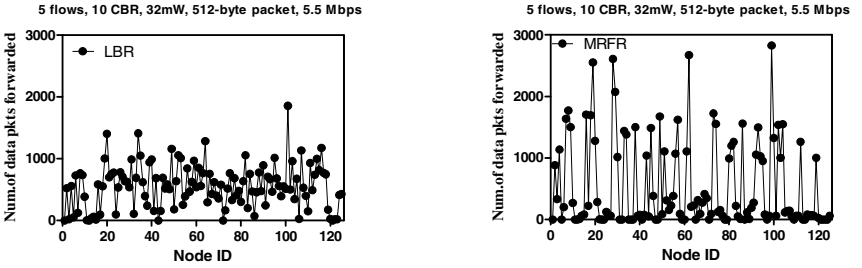


Fig. 4. Comparison of load distribution of LBR and MRFR

Therefore, AODV can use routes with a high probability of route failure. In contrast to AODV, MRFR and LBR take into account the link loss ratios. This is why they find high quality routes and obtain a good performance.

3.2 Average End-to-End Packet Delay

We compared the average end-to-end packet delay of the three protocols. Fig. 3 shows the simulation result. As the data traffic is increased, the average end-to-end packet delay of the three protocols is increased. Although AODV protocol uses the shortest path routing, the average end-to-end packet delay is shown to be the highest in most cases. This is because, in the case of AODV, the route fault happens more frequently, which incurs the large average end-to-end packet delay. Besides using the ETX link quality metric, our protocol uses the load balancing factor when selecting paths. As a result, congestion avoidance is better than that of MRFR. Therefore, as shown in Fig. 3, our protocol has a little smaller average end-to-end delay than that of MRFR.

3.3 Load Balancing Efficiency

Our protocol uses an end-to-end reliable path while considering load balance to transmit data packets. To compare the load balancing efficiency of LBR and MRFR,

we measured the number of data packets that each node forwarded during the simulation. In our simulation, source nodes send data at rate of 10 CBR.

Fig. 4 shows the distribution of the number of data packets forwarded at each node. As shown in Fig. 4, the load distribution of our protocol is more even than that of MRFR. This result indicates that our protocol takes paths that avoid congested nodes near the network center while providing reliable data transmission.

4 Conclusions

We have proposed an effective reliable routing protocol while considering load balance. For load balanced reliable communication, our protocol selects a path that has the small route-ETX value and the minimum remaining load factor of the nodes in the path is large. We have evaluated the performance of our protocol using Qualnet and compared with AODV and MRFR protocols. The simulation results show that our protocol has better performance than the comparing protocols.

Acknowledgments. This work was supported by University of Ulsan, School of Excellence in Electrical Engineering.

References

1. Perkins, C.E., Bhagwat, P.: Highly dynamic destination-sequenced distance-vector routing (dsdv) for mobile computers. In: SIGCOMM 1994, pp. 234–244. ACM, New York (1994)
2. Perkins, C.E., Royer, E.M.: Ad-hoc on-demand distance vector routing. In: 2nd IEEE Workshop on Mobile Computing Systems and Applications, pp. 90–100 (1997)
3. De Couto, D.S.J., Aguayo, D., Bicket, J., Morris, R.: A high-throughput path metric for multi-hop wireless routing. In: Mobicom 2003, pp. 134–146. ACM (2003)
4. Draves, R., Padhye, J., Zill, B.: Comparison of routing metrics for static multi-hop wireless networks. In: SIGCOMM 2004, pp. 133–144. ACM (2004)
5. Adya, A., Bahl, P., Padhye, J., Wolman, A., Zhou, L.: A multi-radio unification protocol for IEEE 802.11 wireless networks. In: First International Conference on Broadband Networks, pp. 344–354 (2004)
6. Keshav, S.: A control-theoretic approach to flow control. SIGCOMM Comput. Commun. Rev. 25(1), 188–201 (1995)
7. Das, S.M., Pucha, H., Papagiannaki, K., Hu, Y.C.: Studying wireless routing link metric dynamics. In: 7th ACM SIGCOMM Conference on Internet Measurement, pp. 327–332. ACM (2007)
8. Sankar, A., Liu, Z.: Maximum lifetime routing in wireless ad-hoc networks. In: INFOCOM 2004. IEEE (2004)
9. Luo, J., He, Y.: Geo quorum: Load balancing and energy efficient data access in wireless sensor networks. In: INFOCOM 2011. IEEE (2011)
10. SCALABLE: The network simulator - qualnet, `http://web.scalable-networks.com/content/qualnet`
11. Ngo, H.P., Kim, M.K.: Mrfr - multipath-based routing protocol with fast-recovery of failures on manets. TI IS 6(12), 3081–3099 (2012)

Secure NFC Authentication Protocol Based on LTE Network

Ebrahim AL- Alkeem, Chan Yeob Yeun, and Joonsang Baek

Khalifa University, Electrical and Computer Engineering Department,
PO Box 127788, Abu Dhabi, UAE

Abstract. NFC (Near Field Communication) has a good adaptable structure that it can be easily combined with any wireless network. Since NFC can be used to communicate without using a proper wire, all the transactions can be done remotely without any physical connections. In this paper, we propose a new authentication protocol based on LTE network to secure the NFC. Our protocol enhances the security level provided by the LTE. Our approach is new in a sense that it covers LTE in contrast to old networks like GSM & 3G substantially treated in the literature. Moreover, both GSM and 3G have several drawbacks when they are combined with the NFC technology, which has potential weakness in confidentiality, integrity, and authentication. Hence our new approach will resolve the security of the new LTE system. We expect that our protocol will result in new secure applications for the smart phone markets.

Keywords: NFC, EAP- AKA, ProVerif, KDF, MME, HSS.

1 Introduction

The need of NFC (Near field Communication) security becomes important by increasing the number of applications that supports. NFC can be used in many useful fields. Many people are using NFC technology in their daily activity as it can be used for money transactions, downloading applications and so. NFC protocol is an efficient technology that comes with major security challenges. The security of mobile transactions is not an easy subject after looking at the characteristics of this system, starting from the price checking process and ending with execution the transaction. The security requirements of NFC protocols mainly depend on the network security which has been discussed on many pervious literature [1, 2]. The proposed protocol provides the desired security requirements to handle all attacks that are related to confidentiality, integrity, availability and authentication. Security analysis with formal verification using a tool based on pi-calculus is also provided for the proposed protocol. The aim of the proposed protocol is to ensure that the majority of the security contributions will be tackled by applying this new scheme which overcomes of the most security vulnerability or attacks. NFC with LTE get its robustness from the two layers provided by the LTE network which will be discussed in Section 3, However, GSM and 3G use one layer only [3, 4]. This paper is organized as follow.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_46,

Section 2 introduces the NFC system and how its work. In Section 3, our proposed NFC protocol is proposed. Section 4 shows the security analysis of the proposed scheme. Finally, Section 6 concludes this paper.

2 How NFC System Works

NFC is a niche technology to exchange data between a smart phone and a tag or between two phones, one of which will act like a reader (phone) and the other will act like a target (tag or another phone). Usually, the target will not require any power since it can be a passive component as the reader will do the scanning process. NFC basically is a subsidiary of RFID (Radio Frequency Identification). However, NFC has shorter reading range about 20 cm and is used in many applications nowadays, such as performing contactless payment through the mobile phone. Also, it can be used in advertising and downloading applications etc. Mainly, the NFC phone reads an NFC tag and communicates with the backend server. In some cases the NFC phone could be both the tag and the reader or there would be no need to contact the backend server. NFC tag or POS (Point of Sale) can be found in many items such as smart posters, the POS, electronic devices, etc. It usually comes with small chip hidden behind a sticker with NFC logo. The payment processing server can be communicated through the NFC by different communication technologies. Smart transactions provided by the backend server and might be vary according to applications such as a web page for reserving movie ticket, issuing receipts, or highly secure financial transaction service which requires high secure connection. The NFC can be used in many applications such application download, ticket purchasing, tourist guide and money transactions. In this paper, we will go through the payment scenario employing the NFC technology. Figure 1 shows a complete smart retail environment based on item level tagging and NFC applications. The following model can increase the efficiency of the retails operation and will smooth the purchasing process.

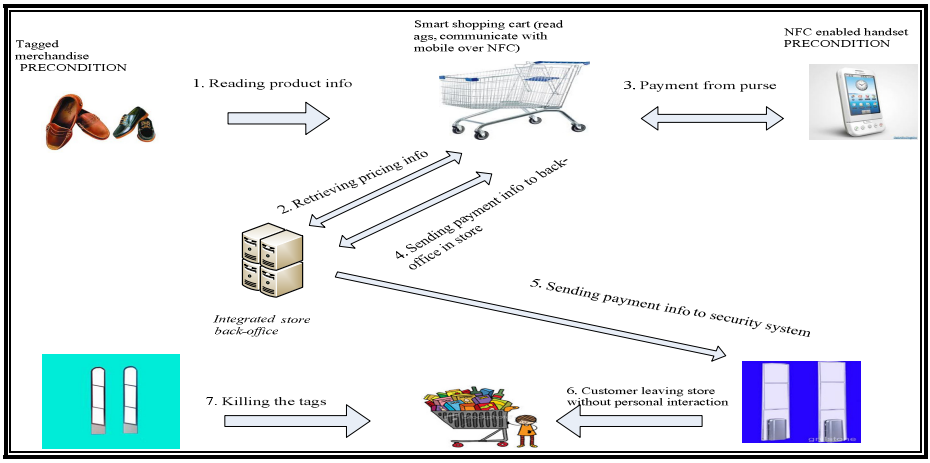


Fig. 1. NFC-based payment processes in retail environment

This efficient model could minimize both operational time and cost as it reduces the human interaction in the operation as a user will be more self-dependent and can conduct most of the operations by himself. There is no need for an employee to tell the client about the price and to go through payment process. This is a very smart approach which minimizes the time requires and efficient.

3 The Proposed NFC Protocol

The proposed design shows an example of a user start selecting an item and checking the price through the NFC POS. The client will make a decision based on the receiving price and proceed with the payments through the LTE network.

We understood that the GSM and 3G have many limitations in terms of security, which can be addressed by using LTE network. Also we noticed that the NFC technology is depending heavily on the network itself as it gains the same security level which network has. Therefore, combining NFC technology with the LTE network will increase the overall level of security. Also, our novel approach has been introduced for the first time. The previous works only covered networks such as GSM [5] and 3G [6]. The proposed scheme provides new approach which has not been discussed in the previous literature, as our solution depends on the LTE infrastructure which provides two security level that can provide isolation to the system when attacks occur.

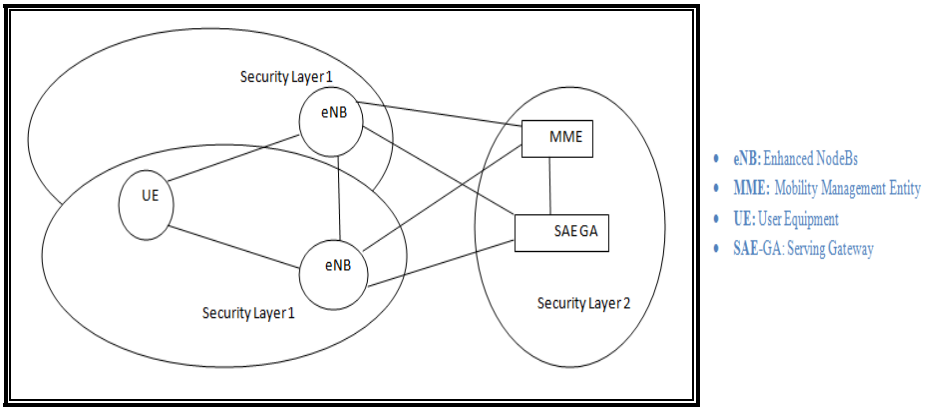


Fig. 2. Security layer of LTE

As Figure 2 illustrates, there are two layers of security in LTE. The reason behind it is that the eNB is not capable to be a fully trusted zone as attacker can gain access and we could add additional layer in order to increase the level of security. That's why the LTE have two security layers: Security of AS (Access Layer) and NAS

(Non-Access Layer). Security layer one is usually between the UE and the eNB which provides encryption and integrity protection for the AS. However, the NAS security is between the MME (Mobile Management Entity) and the UE (User Equipment) which provide encryption and integrity protection for NAS signaling. The NFC payment process through the LTE network can be divided into five major parts: Price Checking, Authentication, NAS Security, AS Security, and transaction execution. The proposed scheme explains the internal process of a customer transaction using the NFC mobile to perform a purchase in a store through the LTE network.

Stage 1 Price Checking

There are many requirements that need to be fixed in order to perform a success transaction. Our proposed design is to achieve the transaction in a physical store environment. Both the POS (Point of Sale) and the client phone should be NFC enabled. The Mobile phone should support the LTE technology and afford high calculation capability. The store area should be well covered by the LTE network and both should operate under the same LTE cell. Figure 3 explains this stage.

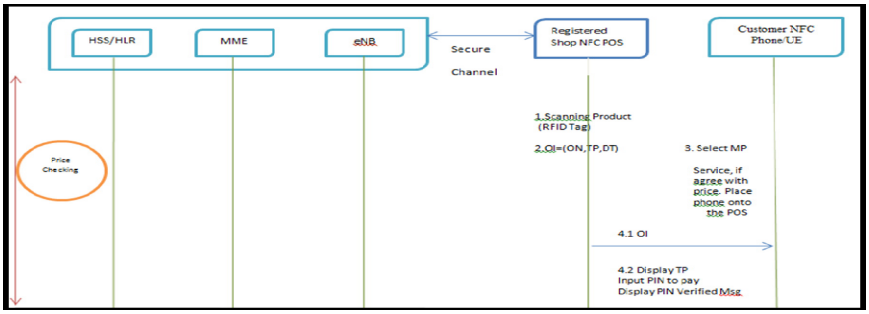


Fig. 3. Price checking stage

Stage 2 Identification & Authentication

We assume that the communication between the payment gate way and the shop POS is over a secure channel. The customer phone should be on and authenticated with the LTE network. Both client phone and the server should share the secret key (KSI_{ASME}). The LTE uses the AKA algorithm, which is a challenge and response protocol that offers a mutual authentication. This means that both the client and the network exchange the challenge and response which allows proving the identity of the client. Therefore, the possibility of MITM (man in the middle attack) attack is not possible with the mutual authentications. The mutual authentication stage is where the mobile reader will agree on the TP(Total Price) with the POS as it will validate the user to guarantee secure transaction. Figure 4 explains this stage.

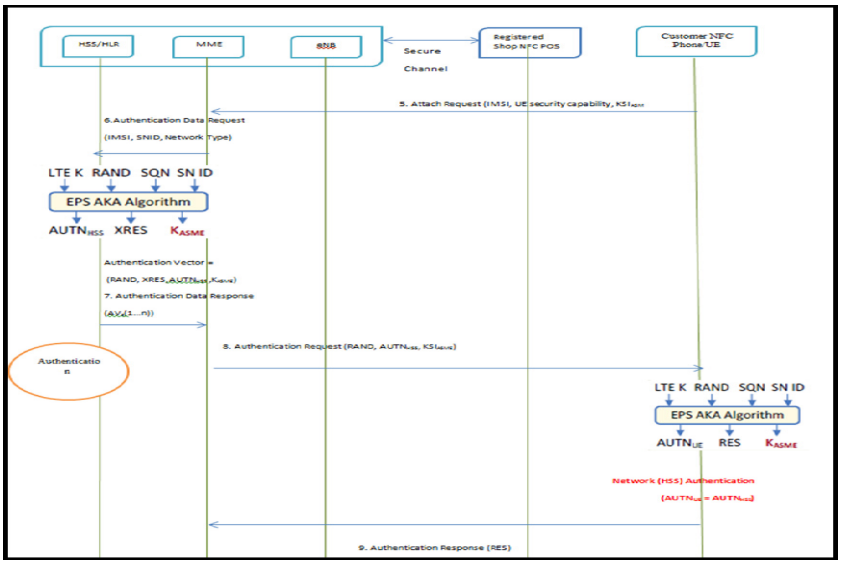


Fig. 4. Identification & Authentication stage

Stage 3 NAS Security

This part will cover the process of Non-Access Stratum (NAS) Security Mode Command (SMC) procedure which allows the system to perform its security through the ciphering & protecting the integrity. The Key Derivation Algorithm (KDF) is mainly used to generate long term master secret key from a short shared secret key also to arrange for a secure channel session. There are two keys generated in the NAS security stage by using the KDF algorithm which are K_{NASenc} & K_{NASint}. Both UE amd the MME get the keys according the K_{ASME} which make sure that the level of confidentiality between UE and MME is high in [7, 8]. The NAS SMC will trigger the client terminal to generate the NAS ciphering key (K_{NASenc}) and NAS integrity key (K_{NASint}). Figure 5 explains this stage.

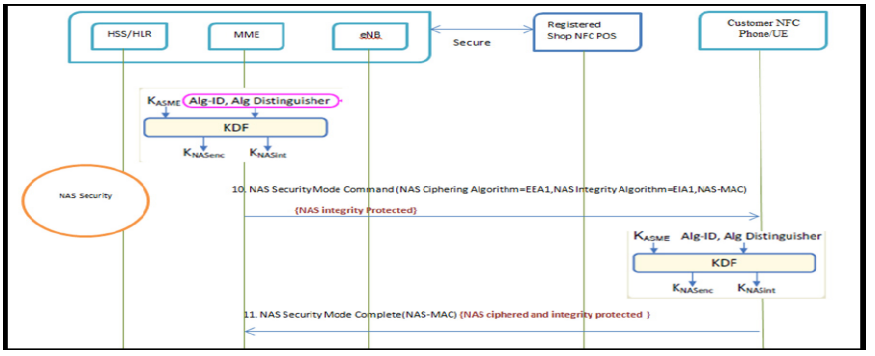


Fig. 5. NAS Security stage

Stage 4 AS Security

In this stage system, one will generate K_{eNB} which will be used in the eNB(enhanced nodeB) to produce the following keys, respectively.

- K_{UPenc}: This key is used to provide confidentiality protection between UE and eNB, Both UE and eNB get the keys according to KeNBS and identity of encryption algorithms.
- K_{RRCint}: This key is essential for protecting the integrity between the UE and eNB
- K_{RRCenc}: The confidentiality of RCC will be protected by this key which deduced according to the K_{eNB} and the encryption algorithm identifier

Figure 6 explains AS security stage.

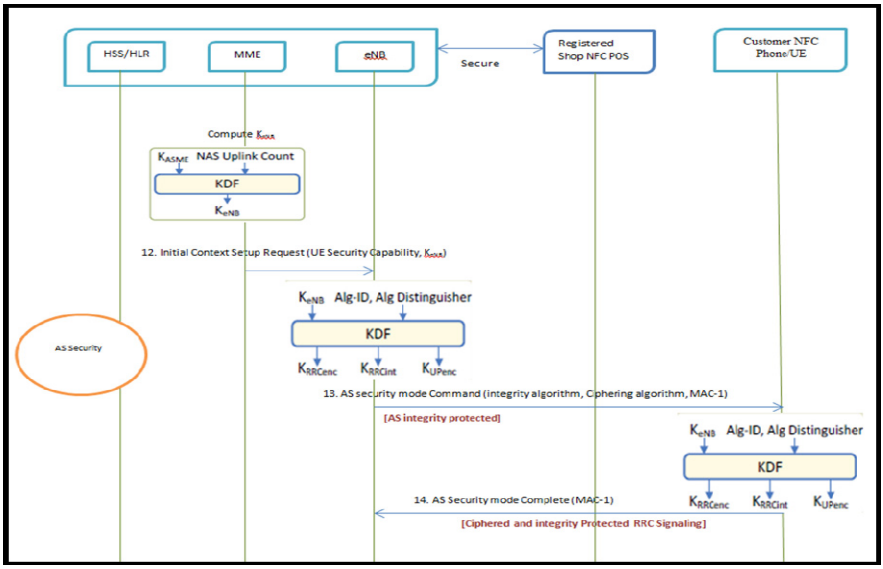


Fig. 6. AS Security stage

Stage 5 Transaction Execution

In this stage, the client will confirm the price by selecting accept button and will confirm the payment process. The PI will be generated by the customer and the process information is prepared to be sent to the POS which includes (Receipt No, Total price, Total count). The backend system will generate time stamp for the transaction to minimize the risk of replay attack. After that, the transaction process will be activated through the HSS which reflect the results to both the POS and the Client in [9, 10]. Figure 7 explains this stage.

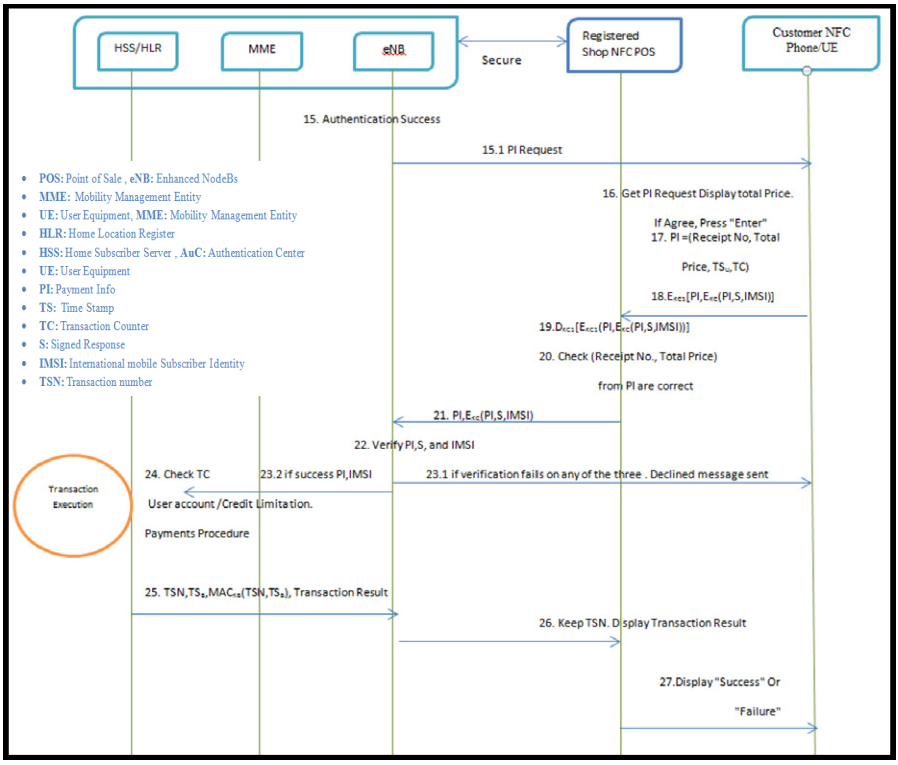


Fig. 7. Transaction execution stage

4 Security Analysis of the Proposed Protocol

There are several attacks that could affect the NFC protocol. The main motivation to use "ProVerif (**Pro**tocol **Verif**ication)" [11] is to analyze and estimate the security capability of a system. ProVerif can test the system from security point of view by focusing on major security requirements and trying all the possible ways to attack it. It is used for verifying security properties for cryptographic protocols using a specification language that is based on an extension of pure Pi-Calculus. Consequently, ProVerif can prove security properties, correspondence and observational equivalence. Note that this proofing capability helps analyze the secrecy and authentication of security protocols that are permitted from the provider.

The mutual authentication is used to authenticate a communication between two parties in which each one is authenticating the other by using nonce, in order to prevent the system from masquerading, spoofing and the man-in-the-middle attacks. By testing the ProVerif on our protocol, we will be able to find the security threats that will affect messages that follow. For example, the attacker can send a response to the processor before receiving the request itself. We could use the value only once to create a mutual authentication between two processors and protect the system. The

ProVerif tests all the probable attempts of the intruder to break the mutual authentication definition in the NFC protocol. Also, the ProVerif will try all assumptions to identify any gap in the confidently of the NFC protocol and will work to introduce an intruder to the protocol and try all the possibility to attack the system messages through the wireless channel. The ProVerif tests the key exchange methodology in the NFC protocol and determines the possible attacks on the system via the public channel. The main challenge is to make sure that the key can be exchanged safely across the system. In our proposal, the integrity is based on the message transaction between two processors. We encrypt the transaction to make sure that the message has not been modified by any outside attacker. That means only the legitimate sender and receiver can read the message and the integrity of the original message is assured. Therefore, providing the security of the session keys and mutual authentication of the protocol can detect any modification attacks.

Also, the system provides the mutual authorization process by identifying the client through the IMSI (International Mobile Subscriber Identity) of the phone, which makes sure that only authorized user is eligible to proceed with the transaction. Also the SNID (Serving Network Identity) generated by the MME will work to identify the specific client to proceed with the authentication process through the EPS AKA algorithm. The IMSI will be used by the HSS to check the credit limit for particular client in order to proceed with the deduction process.

5 Conclusion

This paper proposed a protocol to enhance the security and privacy of NFC transactions based on the LTE network. The proposed protocol basically secures the interactions between all the five parties, which are Client, POS, eNB, MME, and HSS(home subscriber server). Our approach is new since it is based on LTE network, which is different from other old networks like GSM & 3G studied extensively in the previous literature. We provided design specification for implementing our secure protocol over LTE network. Finally, we presented security analysis of our protocol based on the formal method, ProVerif.

References

1. ISO/IEC 14443, Identification cards - Contactless integrated circuit cards Proximity cards
2. Han, C.-K., Choi, H.-K.: Evaluation of Authentication Signaling Loads in 3GPP LTE/SAE Networks. In: 2009 IEEE 34th Conference on Local Computer Networks (October 2009)
3. Massoth, M., Bingel, T.: Performance of Different Mobile Payment Service Concepts Compared with a NFC-Based Solution. In: 2009 Fourth International Conference on Internet and Web Applications and Services, ICIW 2009, pp. 205–210 (2009)
4. NFC Forum, White paper on ‘smart posters’. Tech. Rep. (April 2011)
5. Chen, W.D., Hancke, G.P., Mayes, K.E., Lien, Y., Chiu, J.H.: NFC Mobile Transactions and Authentication Based on GSM Network. In: NFC, 2nd International Workshop on Near Field Communication, pp. 83–89 (2010)

6. Chen, W.D., Hancke, G.P., Mayes, K.E., Lien, Y., Chiu, J.H.: Using 3G Network Components to Enable NFC Mobile Transactions and Authentication. In: Progress in Informatics and Computing, PIC 2010 (2010)
7. NFC Forum, White paper on 'essentials for successful NFC mobile ecosystem'. Tech. Rep. (October 2008)
8. NFC Forum, White paper on 'the keys to truly interoperable communications'. Tech. Rep. (October 2007)
9. ISO/IEC 18092 (ECMA-340), Information technology - Telecommunications and information exchange between systems - Near Field Communication - Interface and Protocol (NFCIP-1)
10. 3GPP. TS 36.331 V9.1.0, 3rd Generation Partnership Project; Technical Specification Group Radio Access Netword; Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC) Protocol Specification
11. Blanchet, B., Smyth, B.: ProVerif 1.85: Automatic Cryptographic Protocol Verifier, User Manual and Tutorial, `http://www.proverif.ens.fr/manual.pdf` (July 11, 2011)

De-Noising Model for Weberface-Based and Max-Filter-Based Illumination Invariant Face Recognition

Hoang-Nam Bui, In-Seop Na, and Soo-Hyung Kim

School of Electronics and Computer Engineering,
Chonnam National University,
Gwangju, South Korea
buihoangnam1988@gmail.com, ypencil@daum.net, shkim@jnu.ac.kr

Abstract. In the topic of illumination invariant face recognition (IIFR), although the state-of-the-art Multi-scale Weberface (MSW) and Multi-scale Quotient Image (MQI) give best results against other illumination insensitive feature extraction methods, they are computationally heavy and easy affected by noises hiding in face shadow. In this paper, we propose a lightweight de-noising model to boost the IIFR system based on max-filter and Weberface called GMAX and GWEB respectively. In this model, we try to eliminate the influence of quantum noise and quantization noise on ill-illuminated images by average smoothing and Gaussian smoothing. After that, linear discriminant analysis (LDA) is adopted to improve verification rate. Never before, a comparative study on popular approaches in the literature fully implemented on the challenging data set Extended Yale B is also provided. The proposed method gives excellent results in term of both computational time and accuracy.

Keywords: illumination, face recognition, max filter, Weberface, de-noising model, principle component analysis, linear discriminant analysis.

1 Introduction and Related Work

Despite the satisfactory face recognition rate in ideal and controlled environment, the recognition performance in uncontrolled condition still remains many challenging problems: the variations in pose, expression, appearance, aging and illumination. Though these problems of image-based face recognition have been raised since early 1960s [16], the exploding researches on illumination issues have actually begun just since last decade. Illumination is an interesting problem of face recognition. Light orientation together with 3D structure of the human face results in the shade and shadows area on the human face [1], which results in a serious problem. For this reason, the variation between two images of the same person taken under two different illumination conditions is much larger than the variation between that of two persons under the same illumination (Fig. 1.). Hence, this poses serious challenges to any classifier taking directly the intensity images as the input.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_47, © Springer-Verlag Berlin Heidelberg 2014

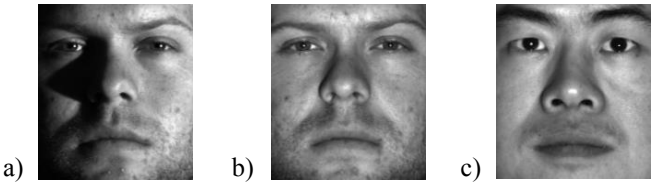


Fig. 1. Variation in lighting angle can result in larger variation between two images of the same identity (a and b) than that of 2 identities under the same lighting condition (b and c).

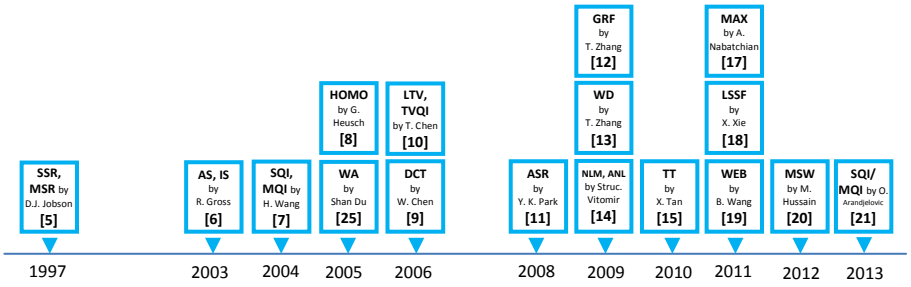


Fig. 2. State-of-the-art methods

Until now, there have been a large number of researches which target at solving the IIFR problem. According to a systematic survey by Xuan Zou [1], there are two main categories: passive approaches and active approaches. Active approaches utilize additional active imaging devices to capture directly face intrinsic information (3D face capturing devices) or face from those spectra other than visible spectra (thermal infrared or near-infrared sensors). However, there are some minor problems such as, 2D face representation is still necessary to improve 3D face recognition rate; or the dependence of near-infrared images upon environmental illumination.

In contrast to active approaches, passive approaches aim at using 2D images only. Compared with low performance template matching, statistical illumination variation modeling methods like PCA with eigenfaces [2], LDA with Fisherfaces [3] did improve the recognition rate, but no truly satisfactory rate was obtained. Other approaches are called physical modeling, which tries to recover face albedo (face texture) or face normal (3D shape) from several input images. The third category includes photometric normalization methods, such as histogram equalization [25], gamma correction [25], DCT normalization [9], wavelet normalization [22], and local normalization [25]. Some of above methods give significant improvement, however the accuracy is insufficient or no thorough evaluation was reported.

Figure 2 shows several photometric normalization and illumination insensitive feature extraction (IIFE) methods at my best knowledge. And, this paper makes a contribution to the exploding research on IIFE, the last category of the passive approaches. Our work is twofold. First, we try to eliminate the affection of noises in the face shadow areas. Second, we apply LDA to increase the verification capability

of the system. The remaining of this paper is divided into 3 sections. Section 2 discusses in detail about our proposed de-noising model and recognition system. Experimental results supporting our proposed method are reported in section 3. Finally, conclusions are given in section 4.

2 Proposed Method

Our proposed method utilizes the idea of using MAX [17], the state-of-the-art method based on Lambertian theory [6], and WEB [19], the state-of-the-art method based on Weber's law, for their friendly implementation and remarkable accuracy. Fig. 3 illustrates our proposed model. From the input image, the face is detected, aligned and cropped to size 192x168 for data collecting. For the purpose of improving recognition performance, a de-noising model is applied before illumination invariant feature extraction. In our de-noising model to be applied before IIFE, each face image goes through a processing pipe-line, in which the image is filtered with a Gaussian kernel after being half-scaled. After input data acquisition, we use PCA and LDA for feature extraction. In testing phase, the simple classifier 1-nearest-neighbor (1-NN) is used for recognition.

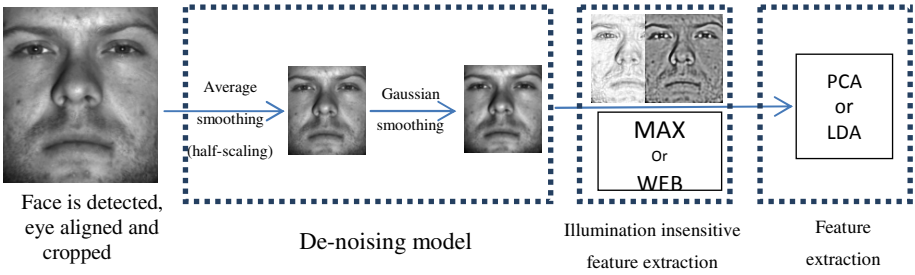


Fig. 3. The proposed illumination invariant face recognition system based on a de-noising model, MAX/WEB, PCA/LDA

2.1 De-Noising Model

Quantum Noise and Quantization Noise

Although many state-of-the-art illumination invariant feature extraction methods give excellent reflectance estimation, they are still influenced by information loss in dark areas of the face. The reason for this degradation is the quantum noise and quantization noise, which was once mentioned in [21]. After MAX and WEB, the noise in the poorly lit regions has been amplified; this is originally imperceptible boundary of the shadow caused by weak ambient illumination. Quantum noise is considered to be isotropic and follows Gaussian distribution, while quantization noise is non-Gaussian and is signal dependent. Fig. 4 shows examples of quantization noise and quantum noise. Here, quantum noise produces unreliable information and quantization noise causes white artifact on MAX output, and gray artifact on WEB output.

Two Levels of De-Noising

There are two levels of de-noising: average smoothing and Gaussian smoothing. In the first level, the input image is half-scaled, causing detail loss and eliminating noise at the same time. We need to make sure that the remaining information is sufficient to discriminate between face classes, and that the reduced noise has less influence on the wrong classification. The output image, if up-scaled to the original size, is similar to that one which is applied with an averaging filter. Though average smoothing eliminates quantum noise, it does not compensate quantization noise.

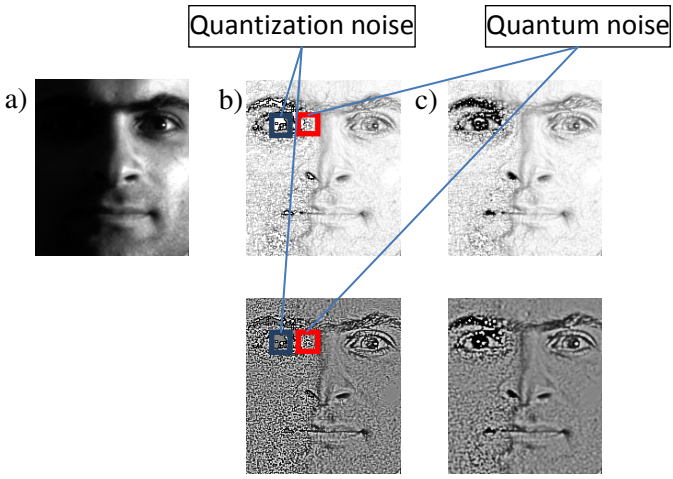


Fig. 4. Quantization noise and Quantum noise for MAX and WEB before de-noising (b) and after de-noising (c). (a) is the original image.

In the second level, we consider quantum noise as sparse signal and isotropic, a Gaussian filter (1) is a proper selection to eliminate quantum noise, but it should have small size to maintain face feature. Gaussian smoothing especially works well with quantum noise around quantization noise boundary, hence it enhances the details around this area. We can see the significant recovery of face detail after using the de-noising model in Fig. 4.

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \cdot e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (1)$$

In term of performance, there are two advantages of applying both levels. The first advantage is low computational cost when only a quarter of original feature vector dimension. Second, the Gaussian kernel $G(x, y, \sigma_0)$ needed on the half-scaled sample is the corresponding half size of that on the original sample ($\sigma_0 = \sigma/2$) and hence lowers more the execution time.

2.2 Face Recognition

As aforementioned, both principal component analysis (PCA) and linear discriminant analysis (LDA) are adopted for feature extraction. The difference from previous work

is that we employ LDA for its strong classification characteristic, which outperforms plain representation characteristic of PCA. The advantage of using LDA is discussed in Experimental results section. In training phase, M face samples of N classes are used as the gallery. First, PCA is used to extract valuable dimensions from the high-dimensional input data. We choose first M-N eigenfaces (in case $M \geq 2N$, or N eigenfaces in case $M < 2N$) to obtain initial face space representation. Second, in the LDA step, the previous face space is projected on to $N - 1$ best discriminating dimensions. $N - 1$ Corresponding Fisherfaces are extracted to build the new face space. In testing phase, 1-nearest-neighbor method is used for classification.

3 Experimental Results

This paper mainly focuses on illumination variation in face recognition. The cropped version of Extended Yale B database [22] was used to evaluate the system. The database contains 2432 images of 38 individuals, each of which was taken under 64 lighting conditions. A lighting condition is represented by 2 parameters identifying the lighting angle: azimuth and elevation. In the cropped version, all faces are eye-aligned, then cropped and resize to 192x168. Images are grouped into 5 subsets based on the angle θ of the light source direction. They are subset 1 $(0 \leq \theta \leq 12)$, 2 $(13 \leq \theta \leq 25)$, 3 $(26 \leq \theta \leq 50)$, 4 $(51 \leq \theta \leq 77)$, 5 $(78 \leq \theta)$. Before the experiment, we need to find coefficient α for GWEB method mentioned in [19]. Here we choose $\sigma_0 = 1$ and try different integer value of α from 1 to 10. The test shows that $\alpha = 2$ give best recognition rate, and this value will be used in our next experiments.

Table 1. Error rate comparison with other methods

Method	Subset 2 456	Subset 3 455	Subset 4 524	Subset 5 705	Error Rate
PCA	45	268	493	684	69.63%
LDA	0	7	300	661	45.23%
GRF [12] + PCA	0	1	97	103	9.39%
ANL [14] + LDA	0	4	82	113	9.30%
NLM [14] + LDA	0	2	59	130	8.93%
DCT [9] + LDA	0	0	42	138	8.41%
GRF [12] + LDA	0	6	79	75	7.48%
ASR [11] + LDA	0	1	26	27	2.52%
WEB [19] + LDA	0	0	28	24	2.85%
MAX [17] + LDA	0	0	23	25	2.24%
MQI [7] + LDA	0	0	16	24	1.86%
*MAX + PCA + DVS [17]	0	0	15	24	1.82%
*MAX + PCA [17]	0	0	14	20	1.59%
MAX [17] + PCA	0	0	11	18	1.36%
WD [13]+ LDA	2	0	13	14	1.35%
*MAX + PCA + WVS [21]	0	0	9	16	1.17%
SQI [7] + LDA	0	0	5	12	0.79%
GMAX 0.45 + LDA	**0**	**0**	**11**	**2**	**0.61%**
MSW [20] + LDA	**0**	**0**	**7**	**4**	**0.51%**
GWEB 0.8 + LDA	**0**	**0**	**3**	**1**	**0.19%**

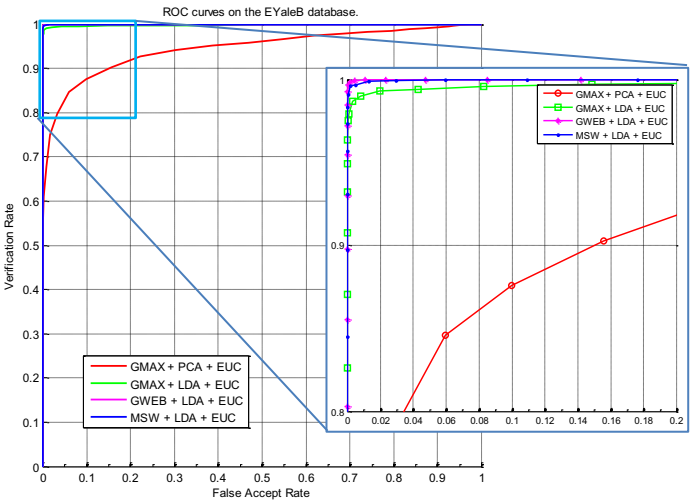


Fig. 5. Evaluation with ROC curves [23] for Extended Yale B database

3.1 Face Recognition Performance

In the experiment, subset 1 is selected to be the training set. Table 1 shows the number of failure cases on each of subset 2 to 5. The results with the star * marked are referred from the corresponding papers. As we can see, this time our proposed method GWEB gives best accuracy with 4 failure images, compared with the state of the art method MSW with 11 failure images. The proposed method gives excellent results on the most severely illuminated faces (subset 5).

3.2 Face Verification Performance

Figure 5 provides 4 ROC curves: GMAX + PCA, GMAX + LDA, GWEB + LDA, MSW + LDA. We can see that (1) the discriminating capability of LDA significantly improves the verification rate of the original system using PCA in. (2) Compared with the perfect ROC curve of MSW, although our proposed GMAX is worse, but GWEB is better. (3) In term of the verification rate in an authentication system, MSW and GWEB are current best choices. (4) However, in a system where the execution time is a concern, our proposed method GWEB and GMAX is still preferable (Table 2.).

Table 2. Average processing time for an face sample of size 96x84

Method	MAX	GMAX	GWEB	MSW	GRF	SQI
Time (milliseconds)	1.8	3.7	3.6	7.7	4.3	11.7

4 Conclusions

In this paper, we proposed a lightweight de-noising model to boost the IIFR system based on max-filter and Weberface called GMAX and GWEB respectively. Our experiment shows competitive results in term of both computational time and accuracy compared with popular approaches in the literature when conducted on the challenging data set Extended Yale B. We found that image half-scaling is not as effective as using Gaussian filter in improving the recognition rate. But, the combination of both methods can not only improve the recognition rate, but also reduce the execution time. Although our method is a promising selection in term of both verification rate and computational expense, the automatic estimation of σ and α_0 is still open and will be considered in our future work.

Acknowledgement. This research was supported by the MSIP (Ministry of Science, ICT & Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (NIPA-2013-H0301-13-3005) supervised by the NIPA (National IT Industry Promotion Agency). This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2013-022495).

References

1. Xuan, Z., Kittler, J., Messer, K.: Illumination Invariant Face Recognition: A Survey. In: First IEEE International Conference on Biometrics: Theory, Applications, and Systems, BTAS 2007, September 27-29, pp. 1–8 (2007)
2. Turk, M.A., Pentland, A.P.: Face Recognition Using Eigenfaces. In: IEEE Conference on CVPR, pp. 586–591 (1991)
3. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Trans. PAMI 19, 711–720 (1997)
4. Horn, B.K.P.: Robot Vision. MIT Press, Cambridge (1986)
5. Jobson, D.J., Rahman, Z., Woodell, G.A.: Properties and Performance of A Center/Surround Retinex. IEEE Transactions on Image Processing 6(3), 451–462 (1997)
6. Gross, R., Brajovic, V.: An Image Preprocessing Algorithm for Illumination Invariant Face Recognition. In: Proc. of the 4th International Conference on Audio and Video—Based Biometric Personal Authentication, Guildford, UK, June 9-11, pp. 10–18 (2003)
7. Wang, H., Li, S.Z., Wang, Y., Zhang, J.: Self Quotient Image for Face Recognition. In: Proceedings of the International Conference on Pattern Recognition (2004)
8. Heusch, G., Cardinaux, F., Marcel, S.: Lighting Normalization Algorithms for Face Verification. In: IDIAP (March 2005)
9. Chen, W., Er, M.-J., Wu, S.: Illumination Compensation and Normalization for Robust Face Recognition Using Discrete Cosine Transform in Logarithm Domain. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 36(2), 458–466 (2006)
10. Chen, T., Yin, W., Sean, Z.X., Comaniciu, D., Huang, T.S.: Total Variation Models for Variable Lighting Face Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(9), 1519–1524 (2006)

11. Young Kyung, P., Seok Lai, P., JoongKyu, K.: Retinex Method Based on Adaptive Smoothing for Illumination Invariant Face Recognition. Signal Processing 88(8), 1929–1945 (2008)
12. Taiping, Z., Yuan-Yan, T., Bin, F., Zhaowei, S., Xiaoyu, L.: Face Recognition Under Varying Illumination Using Gradientfaces. IEEE Transactions on Image Processing 18(11), 2599–2606 (2009)
13. Taiping, Z., Bin, F., Yuan, Y., Yuan Yan, T., Zhaowei, S., Donghui, L., Fangnian, L.: Multiscale Facial Structure Representation for Face Recognition Under Varying Illumination. Pattern Recognition 42(2), 251–258 (2009)
14. Štruc, V., Pavešic, N.: Illumination Invariant Face Recognition by Non-Local Smoothing. In: Proceedings of Biometric ID Management and Multimodal (BIOID) Communication (September 2009)
15. Xiaoyang, T., Triggs, B.: Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions. IEEE Transactions on Image Processing 19(6), 1635–1650 (2010)
16. http://en.wikipedia.org/wiki/Facial_recognition_system
17. Nabatchian, A., Abdel-Raheem, E., Ahmadi, M.: Illumination Invariant Feature Extraction and Mutual-Information-Based Local Matching for Face Recognition under Illumination Variation and Occlusion. Pattern Recognition 44(10-11), 2576–2587 (2011)
18. Xiaohua, X., Wei-Shi, Z., Jianhuang, L., Yuen, P.C., Suen, C.Y.: Normalization of Face Illumination Based on Large-and Small-Scale Features. IEEE Transactions on Image Processing 20(7), 1807–1821 (2011)
19. Biao, W., Weifeng, L., Wenming, Y., Qingmin, L.: Illumination Normalization Based on Weber's Law With Application to Face Recognition. IEEE Signal Processing Letters 18(8), 462–465 (2011)
20. Hussain, M., Muhammad, G., Bebis, G.: Face Recognition Using Multiscale and Spatially Enhanced Weber Law Descriptor. In: 2012 Eighth International Conference on Signal Image Technology and Internet Based Systems (SITIS), November 25-29, pp. 85–89 (2012)
21. Ognjen, A.: Making the most of the Self-Quotient Image in Face Recognition. In: To be published, Proc. IEEE Conference on Automatic Face and Gesture Recognition (FG 2013), Shanghai, China (April 2013)
22. Lee, J.C., Ho, J., Kriegman, D.: Nine Points of Light: Acquiring Subspaces for Face Recognition under Variable Lighting. In: Proceedings of the IEEE Conference on CVPR, vol. 1, pp. 519–526 (2001)
23. Štruc, V., Pavešic, N.: The Complete Gabor-Fisher Classifier for Robust Face Recognition. EURASIP Advances in Signal Processing (2010)
24. Shan, D., Ward, R.: Wavelet-Based Illumination Normalization for Face Recognition. In: IEEE International Conference on Image Processing, ICIP 2005, September 11-14, vol. 2, pp. II-954-7 (2005)
25. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 2nd edn. Prentice Hall

Virtual Reality Based Assessment Tool for Measuring Human Perceptual Sensitivity to Erroneous Motion

Min-Hyung Choi[1], Mohammed Bahni Alquzi[1], and Min Hong[2]

[1] Dept. of Computer Science and Engineering, The University of Colorado Denver
[2] Dept. of Computer Software Engineering, Soonchunhyang University
{min.choi,Mohammed.Alquzi}@ucdenver.edu,
mhong@sch.ac.kr

Abstract. Recent advances in neuroscience have shown that the neuropathological disorders are closely related with diseases such as Alzheimer's. Those damages are particularly associated with the intermediate visual perceptual processing which can cause the motion perception defects and abnormal visuo-spatial functions in daily living of patients. In this paper, we propose a virtual reality based assessment tools for measuring human perceptual sensitivity to dynamic erroneous motions, particularly designed to assess possible early stage of brain damages and its associated visual dysfunctions. The proposed method contains multiple assessment layers to check the awareness of erroneous motion in natural scenes at various severities. Our VR based game type environment provides an effective test bed for various dynamic motion based perceptual sensitivity experiments. Our initial human subject tests show that game based test environment produces more coherent and consistent data, preferable to survey based methods.

Keywords: Perceptual sensitivity, Physically-based modeling, 3D simulation.

1 Introduction

The human visual system perceives the changes of motion which are essential information to survive in our daily life. The motion perception can be achieved from fundamental and critical functions such as depth perception, timing of events, distinction of entities from their environment, posture control, and so on. The malfunction of motion perception can cause serious impairment, such as the inability to hold a coherent reaction or respond properly to simple environmental changes. Previous research projects have reported various successful assessment tools in visual and static image based methods, but those were not sufficient to address real-world scenario. A simple and easy measurement system for human perceptual sensitivity to physically conforming and non-conforming motion is essential for controlled in-depth studies in functional analysis of early stage of potential brain damages.

Virtual Reality provides a computer-simulated environment which supports realistic physical presence in the real world and in imaginary world. A VR system's believability can be greatly improved by adopting real-time physically based

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_48,

simulation in the scene [1, 2, 3]. The perceptual acuity of healthy normal human subjects has been studied extensively in computer graphics and psychology. The resulting consensus shows that there is a surprisingly large degree of error tolerance in our perceptual sensitivity. For instance, we are very sensitive to erroneous human motion. However, when an animation of an artificial object is viewed alone, one can be convinced of its plausibility due to the human brain's inability to accurately disseminate the object's precise and localized behavior. But the tricky part is that perceptual sensitivity fluctuates based on various factors including age groups and the degree of cognitive impairment. Inspired by this condition, we propose a new technique to measure the level of brain damages through virtual reality (VR) based visual perceptual sensitivity analysis. For instance, a subtle and early stage of brain damage could cause inability to discern physically conforming behaviors from physically improbable movements. Our research's focus is to develop an immersive 3D visualization and motion analysis system where we can accurately assess a test subject's vision, ability to recognize presented visual signals, and reaction to the given visual information.

2 Related Work

VR related researches have been performed actively in various fields such as training, game, simulation, and so on. An important issue in this work is that we can inject physically non-conforming behavior at will, and we can control the severity of distortion to our liking. Some earlier researches show that human beings are lacked to accurately distinguish any faults in dynamic motions and O'Sullivan et al. [4] introduced the evaluation system for visual quality of animations from a physically correct motion point of view. Some experiments [5, 6, 7] were also carried out to study the perceptual sensitivity of people for the movement of natural bodies, particularly when the bodies and objects are deformable and complex material properties. According to research results [4, 8], normal people have a level of tolerance in establishing the error from the visual perception and we present the experiments and results in order to more accurately determine these limits, particularly focused on sensitivity to erroneous dynamic motion in natural immersive scene setting. There are very little studies done on the human perceptual sensitivity to erroneous motion of natural objects especially when the target objects are highly deformable and complex material properties are involved. However, based on the previous studies [9, 10], we can postulate that perceptual sensitivity to complex natural phenomena is also limited.

3 Design of Virtual Reality Based Assessment Tool

To test an ability to discern physics awareness in everyday activities, we have developed a natural outdoor scene animator where various causality and eccentricity tests are implemented. Causality refers to the ability to detect whether one event causes another. In a natural outdoor scene, a test subject observes various movements

in the scene and evaluates the displayed animation to extract causes of the motions. Then s/he has to discover the connections between cause and consequential movements in the scene. Among all the movements, there is an embedded non-conforming behavior which would defy consistent force field in the scene, with varied severity. The test subject is asked to identify it within a pre-determined time period. The varied severity of physically non-conforming behavior can be modeled and animated with controllable dynamics simulation system. For example, a correct cause/consequence would be exemplified with strong wind resulting in strong flag motion in right direction. In such an environment any wavering movement of a flag to an opposite direction is easily detected as a non-conforming behavior.

To develop this realistic virtual environment, we adopted a well-reputed Unreal game engine platform. We utilize Speed Tree Modeler [11] to create various forms of plant-life influenced by wind for virtual animations in real-time. Kismet script is used to implement the 2AFC (two alternative forced-choice) severity control mechanism.

All participants are instructed for the overall procedures of experiment and the operator explains the details of controller device for virtual reality test. The objectives of the experiment and observed trails are introduced by an operator. They are also exposed to several practice runs of virtual environment to be familiarized them. All experiments are performed in a series of scenes under 2AFC, method of constant stimuli directed test units, and stirring from the perceptual sensitivity for unnatural motions in natural environments.

4 Experimental Tests and Results

Test subjects are exposed to outdoor natural environments similar to real everyday activities, such as walking on a park. Various natural objects are moving under the influence of natural phenomena, such as wind. A particular object is defying physics and displaying non-conforming behavior. For example, flag is wavering toward incorrect direction of wind at artificially altered speed. We use two test parameters to evaluate the perceptual sensitivity to erroneous motion in natural scene: flag movement direction and flag movement intensity. A flag movement direction is artificially perturbed at various degrees and presented in a scene. Surrounding motions, including movements of trees and gaseous motions in the air, are consistent to provide sufficient visual indications of force field.

For experimental tests, we recruit ten young healthy and normal people as volunteers. The average age is 26.5 years and they didn't have any histories of mental illness. A simple test with moving wind direction is performed at different angles. Test subjects are presented with two scenes with different wind directions and they are asked to pick a visually conforming one, a standard procedure in 2AFC (two alternative forced-choice) severity control mechanism. Typically those two presented scenes are very similar except that they are different in wind direction angle, and consequently the object responses to the different wind directions are clearly presented in physics based animation. One of them (no particular order) has exactly right movement based on the current wind direction while the other one includes a

deviated fixed angle from the exact angle, resulting in physically non-conforming behavior incurred by the intentionally distorted behavior. In this experiment, the severity of the distortion is dictated by the amount of angle of deviation of wind. The varying degrees of the angles are controlled by the severity controller. The severity controller has eight values categorized from narrow angles to wide angles at 20 degree interval. In practice, test subjects reports that it's not easy to discriminate conforming behavior with narrow angle less than 45 degrees, but it's quite obvious to detect cases beyond 120 degrees.

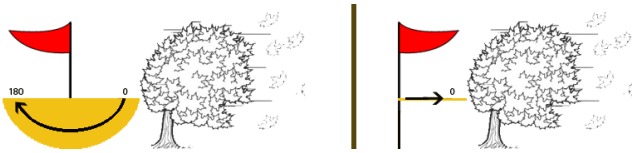


Fig. 1. 2D side-view demonstration of 180 degree severity category

Fig. 1 shows an example of such case in 2D illustration. In the left picture, the motion of trees suggest that the wind direction is from right to left, but the flag shows a completely opposite direction of motion. The flag in the right picture shows conforming behavior. The time interval between stimuli is 5 seconds. After viewing the two stimuli, subjects report their judgments by pressing one of the two designated buttons on the computer keyboard. A new set of trials begins one minute after the subject's response is completed.

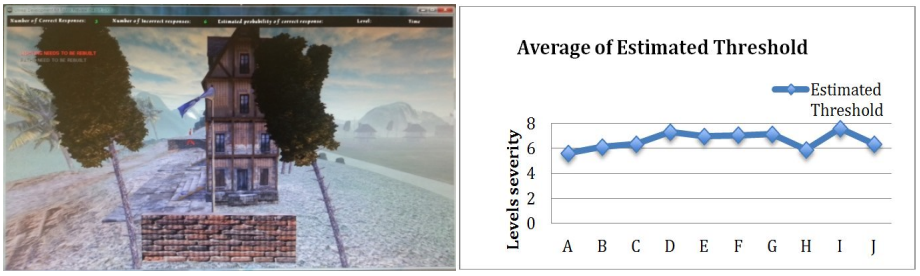


Fig. 2. Snapshot of the proposed human perceptual sensitivity measuring tool in VR on the left and Result of threshold estimation for all participants (0 ~ 180 degree) on the right

After the completion of the entire trial, the level of severity is summarized to a normalized scale from 1 to 8, where 1 being least sensitive to physically conforming behavior and 8 being the strongest sensitivity. Fig. 2 shows a snapshot of our VR based system and a graph of participants' threshold estimation at an average of 6.62. This result presents that the threshold values of participants are between 5.60 ~ 7.58 and it represents the range from level 6 to level 7 of correction. This test result suggests that healthy young people can successfully distinguish physically conforming and non-conforming motions from a natural scene within a reasonable

margin of error bounds. We hypothesize that if this test is applied to patients with mild brain damage, it would give lower percentage of threshold estimation.

5 Conclusion

In this paper we have proposed a VR based assessment system for testing the human perceptual sensitivity from the flawed movement. The proposed system is aimed to measure the test subjects' sensitivity from physically confirming or non-confirming motion in a natural and composite scenario. For this purpose, an outdoor environment is created and loaded with artificial as well as natural motions made with Unreal engine. The results of our experiment show that perceptual sensitivity of physically confirming conduct in a natural setting is in a logically rigid band. Based on our results of experiment, it is considered to be an excellent initiation for the later full-fledged clinical experiments. For the future study, the algorithm can be improved by adding optic flow to do a predictive detection of several scenarios linked with the real natural scenarios.

References

1. Bro-Nielsen, M.: Surgery simulation using fast finite elements. In: Höhne, K.H., Kikinis, R. (eds.) VBC 1996. LNCS, vol. 1131, pp. 529–534. Springer, Heidelberg (1996)
2. Beer, G., Smith, I., Duenser, C.: The Boundary Element Method with Programming: For Engineers and Scientists. Springer (2008)
3. Cotin, S., Delingette, H., Ayache, N.: A hybrid elastic model for real-time cutting, deformations, and force feedback for surgery training and simulation. The Visual Computer 16(7), 437–452 (2000)
4. O'Sullivan, J., Dingliana, T., Giang, M., Kaiser, K.: Evaluating the Visual Fidelity of Physically Based Animations. ACM Transactions on Graphics 22(3) (2003)
5. Riva, G.: Virtual reality as assessment tool in psychology. Virtual Reality in Neuropsycho-physiology, 71–80 (1997)
6. Theeuwes, J., Kramer, A.F., Kingstone, A.: Attentional capture modulates perceptual sensitivity. Psychonomic Bulletin & Review 11(3), 551–554 (2004)
7. Kovács, I., Julesz, B.: Perceptual sensitivity maps within globally defined visual shapes. Nature 370(6491), 644–646 (1994)
8. Clement, J.: "Students' Preconceptions in Introductory Mechanics". The American Journal of Physics 50(1), 66–71 (1982)
9. Abraham, S., Choi, M.-H.: Optimization of Collision Handling based on Differential Thresholds of Human Perception. In: Proceedings of International Conference on Computer Graphics and Virtual Reality (July 2011)
10. Choi, M., Yu, S., Pelak, V.: Assessment of Visual Dysfunction Using Virtual Reality Game Environment. Journal of Future Game Technology (JFGT) 1(2), 82–91 (2011)
11. Sechrest, M., Maher, M., Bordes, J.: NVidia Apex: High definition physics with clothing and vegetation. In: Game Developer Conference (2009)

Assessment of Compatibility between Standard Medical Systems of u-RPMS and HL7

Hyun Nam-Gung, Young-Hyuk Kim, Il-Kown Lim, Jae-Gwang Lee,
Jae-Pil Lee, and Jae-Kwang Lee

Department of Computer Engineering, Hannam University,
Daejeon, Korea
{ghnam,yhkim,ilkim,leejk,jplee,jklee}@netwk.hnu.kr

Abstract. To provide improved environment for medical information service, mobile devices can be used for delivering personal information to HIS (Hospital Information System). MII (Medical Integration Interface) of u-RPMS (ubiquitous Sensor Network Remote Patient Monitoring System) in the previous study plays such role as an interface. However, information may not be delivered due to difference in version of the medical data processing standard HL7(Health Level 7). In this paper, the compatibility according to the HL7 version of HIS was assessed when sending personal information via u-RPMS to HIS. As a result, it was found that it is compatible when the syntax of v2.x is sent to v3.0 by using XML(Extensible Markup Language) Parsing.

Keywords: HL7, HIS, TC215, u-RPMS, XML Parsing.

1 Introduction

Currently, there is active research on standard and switching method of medical information using IT. This technology is commonly called e-healthcare, and can solve the problem with complicated and difficult process of hospital information system by exchanging standardized information between medical institutions and devices. Exchange of medical information and utilization of exchanged information is recognized as the most powerful way to improve efficiency of medical system. For instance, when a patient is moved from a small hospital to a larger one, it costs a considerable amount of money and time to attach copies of medical record, image information, examination, etc. While this is related to profitability, the most important reason perhaps is because hospitals have different information systems [1].

To solve this problem, HL7 was developed based on the need for standard in exchanging information, and has been adopted in developed countries. In South Korea, too, many hospitals use HL7 to establish the hospital information system. In the previous research[2] u-RPMS was studied in order to effectively deliver information by using this medical service.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_49,

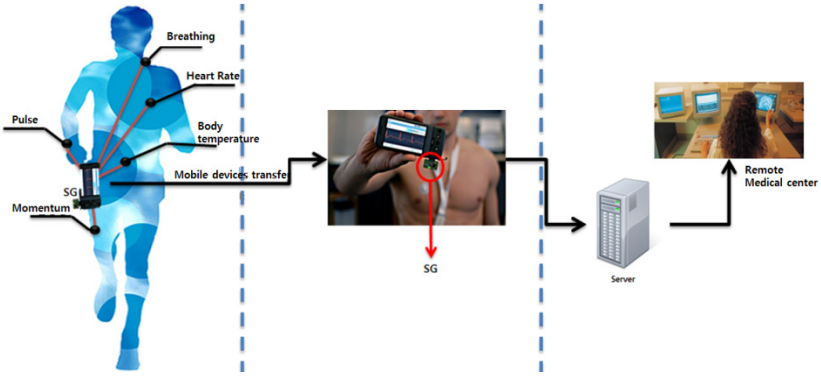


Fig. 1. u-RPMS System

As a result, MII compatible between u-RPMS and HIS targeting MedIntegraWeb that is compatible with HIS using HL7 was designed. It has the advantage of providing healthcare service anywhere based on the improved real time and compatibility.

In this paper, compatibility between MII and HIS is evaluated and a solution is proposed to solve the problem with syntax and segment caused by difference in HL7 versions. In Chapter 2, previous research and HL7 standard technology is examined. In Chapter 3, compatibility between MII and standard technology-based medical system is assessed and, by using XML Parsing, a method to send the syntax of v2x equipment to v3.0 HIS is studied. Lastly, in Chapter 4, the conclusions, problems, and future research problem are suggested.

2 Related Researches

In this chapter, the research of compatibility of HL7-based HIS and MII is divided into MII and HL7.

2.1 MII

Figure 2 below, in addition to the conventional u-RPMS module, MII for conversion to HL7 format is installed in the smartphone. Initially, biometrics is converted to the string type and this data is sent to HL7 Generator of MII, converting it to HL7 format. Later, for security, HIGHT encryption is performed and finally it is sent to HIS via socket communication module of MII. All communication must be notified whether it was successful via ACK (Acknowledge) by sending AA(Response), AD(Error), or AR(Reject).

For implementation, HAPI(HL7 Application Programming Interface) ver.1.2 was used, and Java language was used in normal PC environment to check if there is any problem before implanting in the device.

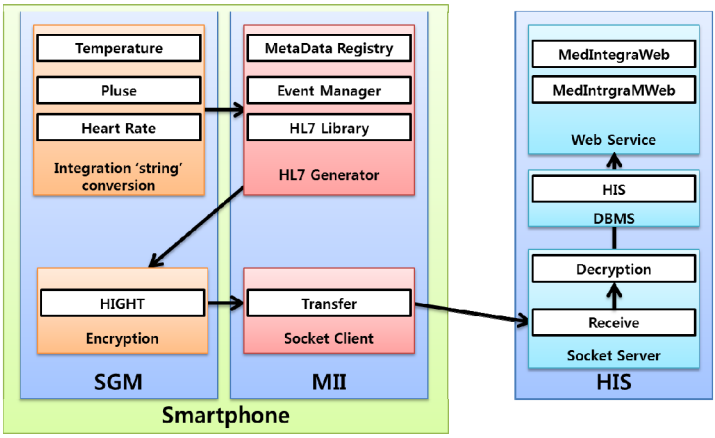


Fig. 2. Entire System Structure

2.2 HL7

HL7 is one of the standards organizations approved by ANSI (American National Standards Institute). It provides standards for data exchange, management, and integration for delivery and assessment of treatment, management, and healthcare service of clinical patients, aiming to ensure interoperability between healthcare information systems. In other words, it is a unified language used for exchanging medical information between different medical information systems.

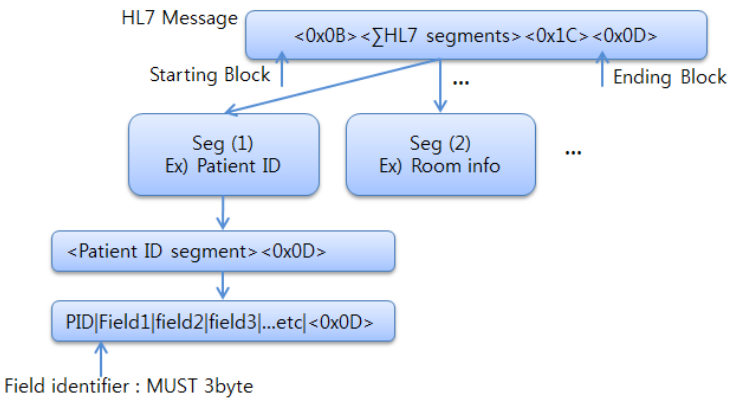


Fig. 3. HL7 Message Syntax Example

In Fig. 3, in HL7 message, <0x0B> is added as ASCII code, a starting block, before a message in order to distinguish different messages, and <0x1C><0x0D> as an ending block at the end. Between the starting block and ending block, HL7 message is placed and each segment signifies information of elements included in

HL7. This segment is made by combination of fields, and each field is divided by |, which is called a field identifier. And, the first field of each segment must be a field identifier made of 3 bytes defined in HL7 standard document [3].

1. HL7 Version 2.X

In general, 2.2, 2.3, 2.3.1, 2.4, 2.5, 2.6, and so forth is used. There is a slight difference between different versions, but the overall structure is similar.

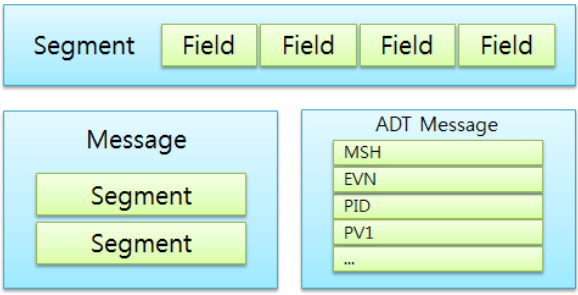


Fig. 4. HL7 Message

As for the structure of HL7 Message Fig. 4, its elementary unit is the data type made of combination of letters and numbers and is called Field. These fields have the meaning be grouped by similar information unit and this is called Segment. Each segment becomes Message which shows an Event as it is combined again.

2. HL7 Version 3.0

HL7 Version 3.0 is the standard which newly appeared to supplement the problems of the existing 2.X. Its biggest problem starts from Optionality. As shown in Abstract Message Definition, Message syntax has a lot of selective elements. Segments are repeated, omitted, or grouped. This Optionality contributed to spread of HL7, but the points that there is not any clear methodology with autonomy got to be shown as the limit of HL7 in the future.

HL7 Version 3 adopted the UML (Unified Modeling Language)-based object-oriented methodology. This tries to escape from the frame of OSI 7 layered model of OSI that 7 of HL7 means being separate from the existing v2.X.

It suggested the standard which can be tested by using the strict analytic techniques and the techniques to generate messages and minimizing selectivity of more Trigger Events and Message Formats and made efforts to provide the techniques which can verify suitability between vendors. Version 3 makes messages by using the object-oriented development methods and RIM(Reference Information Model). RIM is the core part of the methodology to develop HL7 Version 3 which provides expressions to show semantic and lexical connection between information delivered in the fields of HL7 messages [4].

3 Assessment of Compatibility between MII and HIS

3.1 HL7 v2.x and v3.0 Syntax

Among the existing version of HL7, such as 2.1, 2.2, 2.3, 2.3.1, 2.4, 2.5, 2.5.1, 2.6, 2.7, and 3.0, v2.4, v2.6, and v2.7 are the standards that are most widely used. Table 1 lists data to be inputted. Because the syntax of v2.x series is compatible, data were entered according to each syntax and Fig. 5 and Fig. 6 shows patient information as syntax in v2.x and v3.0 syntax respectively.

Table 1. HL7 v2.x Private Data Message Table

Segment	*Data*
Send Location	ELAB-3
Sender	GHH LAB
Receive Location	BLDG4
Receiver	GHH OE
Date	2002-02-15 at 09:30
Patient ID	555-44-4444
Born	1962-03-20
Name	Eve E.
Family	Everywoman
Hospital	Particia Primary
Doctor	Howard Hippocrates
Order	15545^GLUCOSE
Observation	182mg/dl

This syntax is a personal data message syntax of v2.x. It is encoded based on HL7 v2, and also called bar-type syntax. This MSH segment includes message type. This message contains data in the table above, was made at 9:30, February 15, 2002, and MSH segment is a very important segment in the message structure.

MII v2.x syntax fills values in segments like MSH, EVN(Trigger Event), and PID(Patient Identification) before sending. The v3.0 syntax receives data via XML, which causes problems with compatibility between different versions. Therefore, it must be converted according to v3.0 syntax.

```
1 MSH|^~\&|GHH LAB|ELAB-3|GHH OE|BLDG4|200202150930||ORU^R01|
  CNTRL-3456|P|2.4<cr>
2 PID|||555-44-4444||EVERYWOMAN^EVE^E^^^^L|JONES|19620320|F|||
  153 FERNWOOD DR.^^STATESVILLE^OH^35292||(206)3345232|(206)752-121
  ||||AC555444444||67-A4335^OH^20030520<cr>
3 OBR|1|845439^GHH OE|1045813^GHH LAB|15545^GLUCOSE|||200202150730
  |||||||||555-55-5555^PRIMARY^PATRICIA P^^^^MD^^|||||||||F||||||
  444-44-4444^HIPPOCRATES^HOWARD H^^^^MD<cr>
4 OBX|1|SN|1554-5^GLUCOSE^POST 12H CFST:MCNC:PT:SER/PLAS:QN||^182|
  mg/dl|70_105|H|||F<cr>
```

Fig. 5. HL7 v2.x Private Data Message Syntax

```
<observationEvent>
  <id root="2.16.840.1.113883.19.1122.4" extension="1045813"
    assigningAuthorityName="GHH LAB Filler Orders"/>
  <code code="1554-5" codeSystemName="LN" codeSystem="2.16.840.1.113883.6.1"
   displayName="GLUCOSE^POST 12H CFST:MCNC:PT:SER/PLAS:QN"/>
  <statusCode code="completed"/>
  <effectiveTime value="200202150730"/>
  <priorityCode code="R"/>
  <confidentialityCode code="N" codeSystem="2.16.840.1.113883.5.25"/>
  <value xsi:type="PQ" value="182" unit="mg/dL"/>
  <interpretationCode code="H"/>
   <referenceRange>
     <interpretationRange>
       <value xsi:type="IVL_PQ">
         <low value="70" unit="mg/dL"/>
         <high value="105" unit="mg/dL"/>
       </value>
     <interpretationCode code="N"/>
   </interpretationRange>
 </referenceRange>
<author>
 <time value="200202150730"/>
 <modeCode code="WRITTEN"/>
 <signatureCode code="S"/>
    <assignedEntity>
    <id root="2.16.840.1.113883.19.1122.3" extension="444-444-4444"/>
      <assignedPerson>
         <name>
            <given>Harold</given>
            <given>H</given>
            <family>Hippocrates</family>
            <suffix qualifier="AC">MD</suffix>
         </name>
      </assignedPerson>
    </assignedEntity>
 </author>
<recordTarget>
    <patientClinical>
    <id root="2.16.840.1.113883.19.1122.5" extension="444-22-2222"
     assigningAuthorityName="GHH Lab Patient IDs"/>
    <statusCode code="active"/>
      <patientPerson>
        <name use="L">
           <given>Eve</given>
           <given>E</given>
           <family>Everywoman</family>
        </name>
        <asOtherIDs>
           <id extension="AC555444444" assigningAuthorityName="SSN"
             root="2.16.840.1.113883.4.1"/>
        </asOtherIDs>
      </patientPerson>
    </patientClinical>
 </recordTarget>
<inFulfillmentOf>
      <placerOrder>
      <id root="2.16.840.1.113883.19.1122.14" extension="845439"
       assigningAuthorityName="GHH OE Placer orders"/>
      </placerOrder>
   </inFulfillmentOf>
 </observationEvent>
```

Fig. 6. HL7 v3.0 Private Data Message Syntax

Fig. 6 shows conversion of HL7 v2.x syntax in Fig. 5 to v3.0 syntax. However, in MII in the previous study, this conversion process was not included [5].

3.2 HL7 v2.x Syntax Converting to XML Syntax

Using the designed MII, the syntax in Fig. 5 can be XML-converted to improve compatibility as a way to deliver HL7 of v2.x to v3.0 HIS. Fig. 7 shows how the items of each syntax is converted based on XLM to be sent to HIS.

```
<?xml version="1.0" encoding="UTF-8"?>
<ORU_R01 xmlns="urn:hl7-org:v2xml">
    <MSH>
        <MSH.1>|</MSH.1>
        <MSH.2>^~\&amp;amp;</MSH.2>
        <MSH.3>
            <HD.1>GHH LAB</HD.1>
        </MSH.3>
        <MSH.4>
            <HD.1>ELAB-3</HD.1>
        </MSH.4>
        <MSH.5>
            <HD.1>GHH OE</HD.1>
        </MSH.5>
        <MSH.6>
            <HD.1>BLDG4</HD.1>
        </MSH.6>
        <MSH.7>
            <TS.1>200202150930</TS.1>
        </MSH.7>
        <MSH.9>
            <MSG.1>ORU</MSG.1>
            <MSG.2>R01</MSG.2>
        </MSH.9>
        <MSH.10>CNTRL-3456</MSH.10>
        <MSH.11>
            <PT.1>P</PT.1>
        </MSH.11>
        <MSH.12>
            <VID.1>2.4</VID.1>
        </MSH.12>
</MSH>
```

Fig. 7. XML Converted HL7 v2.x Syntax

3.3 XML Converted HL7 v2.x Syntax LOC(Lines of Code)

In cost analysis of the source, one of the important factors is the number of code lines. It also acts as an indicator of processing speed and cost. Table 2 shows the number of lines regarding v2.x, v3.0, and Converted v2.x.

Table 2. HL7 v2.x Private Data Message Table

	v2.x	*v3.0*	*Conerted v2.x*
Lines of Code	4	56	135

Table 2 indicates that v2.x is as about 33 times short as converted v2.s. This much of difference in wireless environment can lead to delayed data processing and increased error rate. In that perspective, as it is used for medical service, more reliable transfer will be possible if MII can be sent to HIS based on v3.0 code.

4 Conclusion

Currently, in fact, there are not many cases in the Korean medical information system market that have adopted HL7. However, HL7 has been established as a standard of medical information system for the purpose of improving the medical service for patients. Remote healthcare service will become possible if user information is

measured in mobile device environment and the data is sent to HS of the hospital. For that reason, in this study, compatibility assessment was performed in order to evaluate whether MII designed in the previous study is suitable for efficient use. It was found that, as MII complies with v2.x HL7, it has a problem with compatibility with v3.0, and the study was performed regarding the solution for it.

As a result, XML Parsing was performed to HIS, which enabled converting it suitable for v3.0. To improve compatibility with v3.0 system, it is important to perform XML Parsing for each syntax and matching it to the appropriate segment before sending it to HIS.

Because this study was based on wireless system instead of general cable system, XML Parsing causes increase of LOC and, thus, the cost, and difficulty with providing accurate real-time medical service. As the problem with data traffic can be substantial, future research will need to enable conversion of v2.x data directly to v3.0data.

Acknowledgement. This research was supported by Basic Science Research Program throuch the national Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (2011-0113029).

References

1. Park, Y.-J., Cho, H.-S.: Exchange Method of the HL7 Message using Open Source on OSGi Environment. Information Science Journal Computing Practice and Letter 18(11), 775–779 (2012)
2. Kim, Y.-H., Lim, I.-K., Lee, J.-W., Lee, J.-G., Lee, J.-K.: Designing Medical Integra Interface with u-RPMS and HIS. In: Park, J.J(J.H.), Leung, V.C.M., Wang, C.-L., Shon, T. (eds.) Future Information Technology Application and Service. LNEE, vol. 164, pp. 339–347. Springer, Heidelberg (2012)
3. Uhm, J., Park, S.-H.: Application of the Modified Real-Time Medical Information Standard for U-Healthcare Systems by Using HL7 and Modified MFER (TS-MFER). Korea Communication Application Journal 37C(8), 680–689 (2012)
4. HL7 Clinical Document Architecture, Release 2.0, HL7 Version 3 Standard (2005)
5. Ringholm: HL7 Message example version2 and version3 (2007)

The Extraction of Spatial Information and Object Location Information from Video*

Kyung-Je Park, Min-Soo Moon, and Ki-Jung Lee

Geurimsoft, Researcher, Seongnam-si, Korea
{park,storymoon,jcm5758}@geurimsoft.com

Abstract. This paper presents a extracting method for 3D spatial information of background and location information of subject from video. First to extract the 3D spatial information, we used rotating angle of frames from video. Second, we detected subjects from frames. Finally, we calculated angles from each frames. To find the relative position of subject in each frame, we mixed the information of surroundings and movement of subject. The image data can be calculated with the user-providing information on the first position of camera.

Therefore the GPS coordinate can be estimated corresponding to the position of protagonist in each frame and many different techniques of enhanced-reality applied services could be developed.

Keywords: Spatial Information, Object Tracking, Structure from Motion, Augmented Reality, Image Registration.

1 Introduction

Nowadays anyone can easily make a film by simply recording video thus the filming locations of movie or drama are good for them. The images produced in the same space at the same angle are giving impressions as if they had become the virtual character of the movie or drama. The pictures taken together with the character are more satisfying.

Thus the filming locations become core information. The spatial information shall be built by the existing Structure from Motion (SFM) technique. SFM analogizes movement of camera occurred when filming from many different angles around an object and restores 3D shape. To digitize the 3D shape and surface color of the objects that exists in the real life [1].

For structuring the spatial information, we used rotating angles between frames and for extracting the location information, we traced the movement of object. Then we could combine the spatial information and object location information in one frame. This paper shows an extracting method of the spatial information and object location information.

* This research is supported by Ministry of Culture, Sports and Tourism(MCST) and Korea Creative Content Agency(KOCCA) in the Culture Technology(CT) Research & Development Program 2013 (R2012030034_00000002).

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_50,

2 SFM(Structure From Motion)

To obtain the movement of camera using images from one camera and estimates the 3D structure. To find the features of images entered and calculate the relationship between images [2, 3]. Figure 1 shows the relationship as below. M is one feature and m_i is corresponding features in respective frames. F represents the geometric transformation of two theatres. The accurate surface cannot be estimated with the surface model of features only.

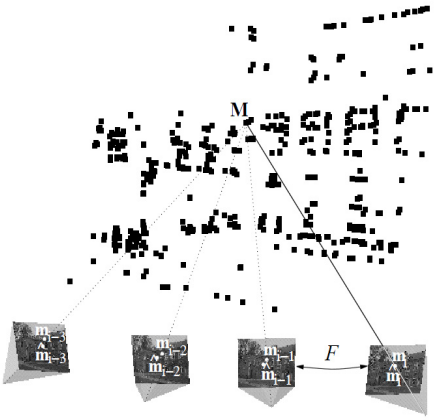


Fig. 1. Feature relationship [3]

Further dense surface estimation is required. The dense surface estimation shall be conducted over several times and the first image pair shall be stationed at the standard stereo composition. And then the depth map shall be calculated thru a stereo matching algorithm. Finally the result of all pairs is combined.

This paper describes the method of extraction and storage of spatial information and object location information using the relationship enabling to coordinate into one frame. The chapter 3 suggests the method of extraction and storage of spatial information and object location information dealing with the extraction of rotating camera angle by matching the features and the calculation of scale, rotation angle, distance between camera and object, latitude and longitude of object by extracting bounding box. The chapter 4 shows the test results and chapter 5 provides the final conclusion.

3 The Proposed Method

This paper has two big processes. The first process is to extract spatial information. To extract image information from the original frame and calculate the rotation matrix by analyzing the information of each frame. The spatial information is extracted by

analyzing the image information and rotation matrix. The second process is about object location information. Using background eliminated original image, objects are extracted. The object extraction is to create bounding box by tracing the outermost line. Using rotation matrix and object information, calculate objects location (latitude, longitude) by calculating rotation and distances. The Figure 2 (a) shows the flow of two processes as below. Figure 2 (b) shows the process how to extract spatial information.

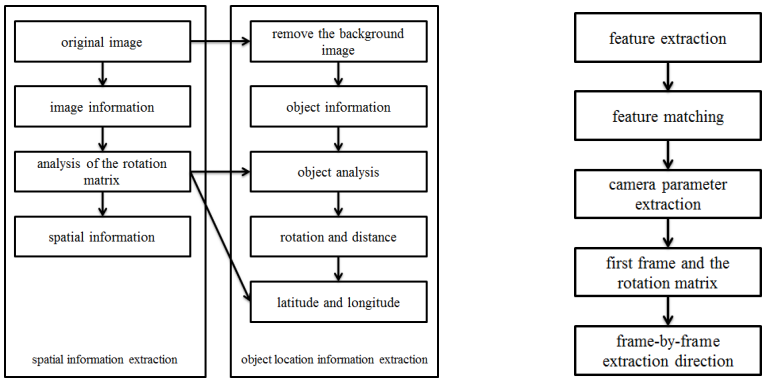


Fig. 2. Framework and Process of extracting Spatial Information

3.1 Feature Extraction and Matching

To use the algorithm of Speeded up Robust Features (SURF) [4] in extracting features. The SURF algorithm is a detection method to extract features using the maximal point of Hessian matrix and the strength of SURF algorithm lies in fast speed. This paper gets a lot of influence of speed due to real-time processing and mobile services. Therefore the fast SURF algorithm is used in detecting the features.

The extracted features reach average 1,800 to 2,000 per frame. As it takes too much time and memory to match all of such features, it is required to decrease the features around 30 per frame using the algorithm of RANSAC (RANdom SAmpling Consenus) [5]. The RANSAC algorithm is a method to find the optimum solution by repeatedly calculating the model parameters while sampling the random data. Once the data is divided into inlier and outlier then the first priority is to reduce the number of unnecessary outlier. To extract the matching points between the respective frames using Brute force matching algorithm[6] and estimate the basic matrix of camera.

The accuracy of basic matrix should be increased by epipolar constraint [7], as the typical frame consists of rotation of the left and right only. The feature found in one frame appears on the same epipolar line of the other frame. The redline represents the epipolar line in Figure 3 as follows.

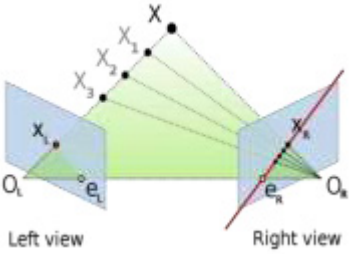


Fig. 3. Epipolar constraint

3.2 Camera Parameter Extraction

The frames have all different camera parameters and the camera parameter should be extracted by unique information for all frames. The camera parameter consists of internal and external parameters. The internal parameter indicates the lens focal length, pixel size and location of the centroid point and the external parameter represents the value of rotation and movement between images.

For the camera to be considered rotating at one position only, the projection change can be expressed as formula 1 assuming that all points are distant from camera.

$$H_{10} = K_1 R_1 R_0^{-1} K_0^{-1} = K_1 R_{10} K_0^{-1} \tag{1}$$

Which can also be simplified as K_k=diag(f_k, f_k, 1) indicating the internal parameter of camera.

The above formula can also be written as follows.

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = \begin{bmatrix} f_1 & & \\ & f_1 & \\ & & 1 \end{bmatrix} R_{10} \begin{bmatrix} f_0^{-1} & & \\ & f_0^{-1} & \\ & & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix} \tag{2}$$

The above formula could be simplified as below.

$$\begin{bmatrix} x_1 \\ y_1 \\ f_1 \end{bmatrix} \sim R_{10} \begin{bmatrix} x_0 \\ y_0 \\ f_0 \end{bmatrix} \tag{3}$$

The parameter can be obtained with smaller factor if the rotation angle is limited to the rotation about the Y axis only. And further, the rotation angle between the two adjacent frames can be obtained.

3.3 Orientation Extraction and Image Stitching

The direction could be extracted by images thru calculation of the direction information of the first frame based on the magnetic north and the rotation angle

between the first frame and the respective frames. The user should enter the direction information of the first frame based on the magnetic north for this research.

It is available to synthesize the respective frames into one frame by integrating the rotation angle by frame obtained from matching of features along with the calculated image scale information. Blending technique is used for color differences and awkwardness which occurred merging the frames with rotation angle by frames and scale information.

Figure 4 shows the input images and the image stitched results.

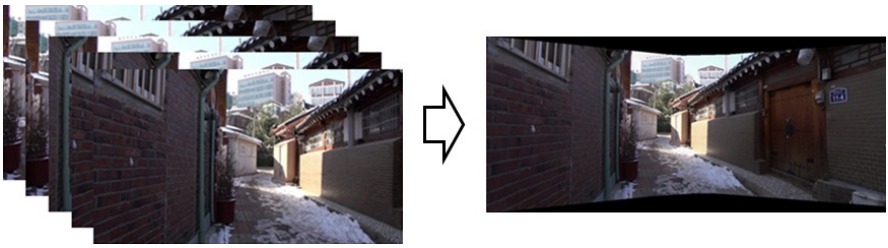


Fig. 4. Result of Image Stitching

3.4 Object Location Information Extraction

The extraction technology of object location information further utilizes the background-eliminated object frames shown in the Figure 5 below.

Fig. 5. The original image and Background removed image

The latitude and longitude of object shall be extracted using the spatial information extraction result of the original frame, bounding box information from object frame, movement information of object and the scale information of object. The process is shown in the Figure 6.

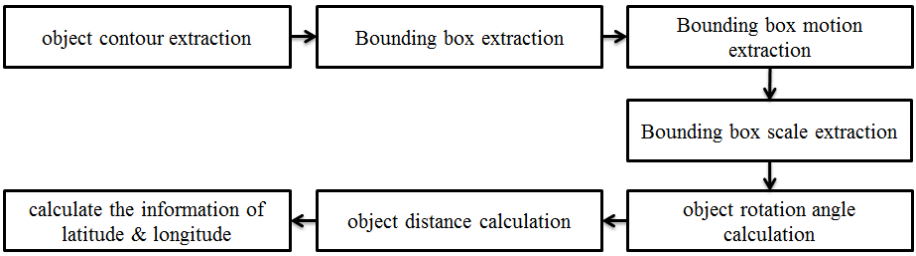


Fig. 6. Process of extracting Object Location

3.4.1 Extracting Bounding Box

To create the object bounding box, extracting the outermost line of object. To calculate the centroid point of bounding box and then extract the movement of the centroid point of bounding box by frame. To calculate the distance on screen by the Euclidean distance, calculate the rotation angle of object movement. The following Figure 7 shows the movement of object.

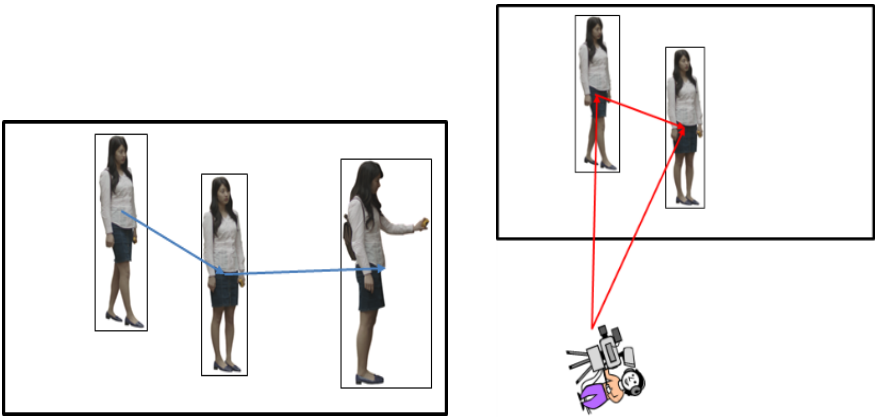


Fig. 7. Object movement and rotation

The object rotation angle represents the object movement angle on the screen and shows the change from previous frame.

3.4.2 Object Distance Calculation

The distance between camera and object shall be extracted by the scale change of bounding box. As the focal length is dependent on the camera and the actual object size is a fixed value thus the focal length and actual object size should be passively and manually entered. The following Figure 8 shows the relationship of object distance calculation.

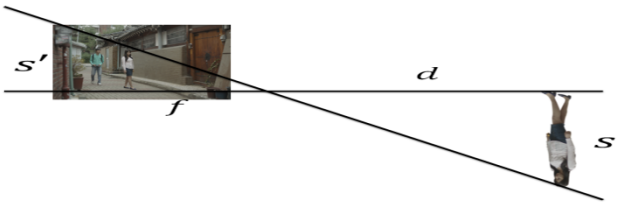


Fig. 8. Calculating Object distance

S is the actual object size, S' is the size on screen, f is the focal length and d is the distance between camera and object. S:S' = d:f is the distance between object and camera, d = (S*f)/S'. The real size of objet S and the size of screen S' are using different units. Therefore the size on screen shall reflect the pixel size information. The unit of pixel size information is dot per inch (dpi).

3.4.3 Calculate the Information of Latitude and Longitude

To calculate the information of latitude and longitude with the information of rotation angle by frame extracted from the original frame, object rotation angle, and distance between camera and object. The rotation angle by frame indicates the rotation of camera and the object rotation angle refers to the object movement angle in the frame. To calculate the initial coordinates of latitude and longitude of the object by the distance between camera direction and object and the user shall provide the initial camera direction.

4 Results

The operating system used to examine this paper has been constructed by Visual Studio 2010 and OpenCV 2.4.1 in Windows7 environment. The frames used for the experiment are some portion of a movie of 'ARCHITECTURE 101`. H.264 Codec is used for all 531 frames.

Table 1. Object latitude and longitude

Frame	Latitude	Longitude
Suji_001.png	37.57974367	126.96812431
Suji_150.png	37.57974628	126.96812696
Suji_300.png	37.57974615	126.96812706
Suji_450.png	37.57975526	126.96813040
Suji_531.png	37.57975820	126.96812951

Table 1 is the information of latitude and longitude extracted from the object by frame. The enhanced-reality technology applied services by extracting the object location by frame are available

The below Figure 9 is a frame showing the experiment results. The entire filming could be verified by saving the spatial information of filming location and by stitching images and the object is synthesized on the stitched frame. The object is located on the latitude and longitude of the object, the scale is dependent on the filming distance and shows fine result.

Fig. 9. Service Results display

5 Conclusion

The object is verified to be synthesized on the very accurate location of the image stitched frame. The very effectively synthesized frames other than just simply combined frames could be obtained by this approach with the spatial information of filming location and object location information. The objects are synthesized in the very accurate location according to the movement of camera. The spatial information and object location information could be extracted from one location only for the time being. Further research is required to be able to extract the same spatial information and object location information from the other locations afterwards.

References

1. Chu, C.W., Park, J.Y., Kim, H.W., Park, J.C., Lim, S.J., Koo, B.K.: Recent Trends of 3D Reconstruction Technology. Electronics and Telecommunications Trends 22(4), 1–11 (2007)
2. Beardsley, P., Zisserman, A., Murray, D.: Sequential Updating of Projective and Affine Structure from Motion. International Journal of Computer Vision 23(3), 235–259 (1997)
3. Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., Koch, R.: Visual Modeling with a Hand-held Camera. International Journal of Computer Vision 59(3), 207–232 (2004)
4. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded up Robust Features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
5. Fischler, M., Bolles, R.: Random Sample Consensus: a Paradigm for Model Fitting with Application to Image Analysis and Automated Cartography. Communications of the ACM 24(6), 381–395 (1981)
6. Levitin, A.V.: Introduction to the Design and Analysis of Algorithms. Addison-Wesley Professional (2002)
7. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press (2000)

Computer Assisted English Learning System with Gestures for Young Children

Seng Il Jung, Joon Yeon Choeh, Sung-Wook Baik,
Soonil Kwon, and Jong-Weon Lee*

Department of Digital Contents, Sejong University, 98 Gunja-dong, Gwangjin-gu, Seoul, Korea
{zoon,sbaik,sikwon,jwlee}@sejong.ac.kr,
reiuzi.u@gmail.com

Abstract. Kids use the computer assisted language learning systems to learn English. The contents of the system are well designed and kids enjoy them. From Cognitive Psychology we found gestures played useful role in learning so we developed the language learning system utilizing gestures. The system provides similar contents as the existing system but tries to enforce users to follow gestures related to given words. We compared the proposed system with an existing one in terms of memorizing test scores. The average improvement achieved using the proposed system was little better than one achieved using the existing system.

Keywords: Computer Assisted English Learning, Gesture Recognition, Depth Sensor.

1 Introduction

There have been many researches in Computer Assisted Language Learning (CALL) and some of them have been used in real life. Kids use the CALL systems to learn English in Korea. Kids watch contents and repeat words shown on a screen as instructed by the system. The contents of the CALL system are well designed and kids enjoy the system. However, there are still opportunities for improving the system. We developed a new Computer Assisted English Learning (CAEL) system that could help kids to remember English vocabularies easily. We combined researches in Cognitive Psychology and in Human-Computer Interaction (HCI).

From Cognitive Psychology we found gestures played useful role in learning. Goldind-Meadow et al. asked kids and adults to memorize a list of items while they explained their solution of the math problem [1]. They found both groups do better in memorizing items when they used gestures during the explanations. Cook et al. asked kids to solve math problems, finding the numbers that equal both sides of the equations [2]. They grouped kids into three groups depending on the methods used to instruct kids. They found kids in all three groups improved their performance after the learning period. One difference was found from the retest after four weeks. Kids used gestures had higher scores for the retest than kids used only speeches.

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_51,

Microsoft introduced the Kinect few years ago. The Kinect can capture the depth of a captured image and extract the skeleton of a user in real time. Because of its cheap price and reasonable accuracies, researchers developed many gesture-based interaction methods and their applications using Kinect [3]. We used the Kinect to develop the gesture-based HCI and applied it to the new CAEL system. We hope the gesture interface included in the developed system help kids remember vocabularies better than the system without using gestures as cognitive psychologists suggested and give kids more playing time.

2 Methods

We developed the CAEL system on a PC with a Kinect. The system showed a kid one word and the gesture related to the word (Figure 1) and played the recorded pronunciation of the word. The kid repeated the pronunciation and followed the displayed gesture. If the kid followed the gesture correctly, the next word appeared on the screen.

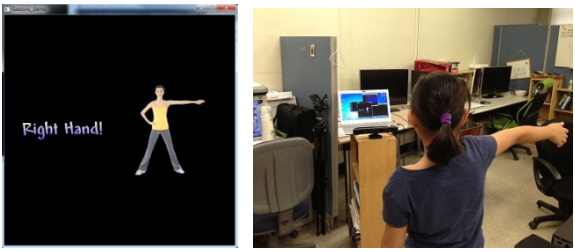


Fig. 1. The proposed system (Left) An example screen shot (Right) A kid using the system

The system is divided into three modules, a database, a gesture recognition and an output. The database contained lists of words and short sentences. These lists were grouped according to the target users. Every word and short sentence in the database cannot be described with a gesture. Generally action verbs are easily represented with gestures. We only trained gestures for parts of words and short sentences in the database and stored them in the database, which were linked with corresponding words or short sentences. These words and short sentences are called gesture words and sentences.

We developed the gesture recognizing module using the Dynamic Time Warping (DTW) algorithm [4] and the Kinect sensor from the Microsoft. The DTW measures the similarity between two sequences varying in time or speed. Each gesture was recorded as the sequence of 2D points of the skeleton detected by the Kinect. These recorded sequences were stored in the database and used to recognize users' gestures. These sequences could be varied by the viewing direction of the Kinect, the orientation of the user, the distance between the user and the Kinect and the height of the user. We applied two normalization steps to overcome these variations so only

one gesture was enough for recognizing the same gestures of users with varying characteristics.

In the first normalization step, we rotated the skeleton detected by the Kinect along the center of the skeleton so the skeleton stood orthogonal to the z direction. By rotating the skeleton we could eliminate the effect caused by the varying viewing direction of the Kinect. The z values were eliminated to get 2D points of the skeleton. This simplified the matching procedure, which detected whether two skeletons were similar or not. We also normalized the 2D coordinates of the detected skeletons using the distances between an elbow and a wrist to reduce effects caused by the heights of users and the distance between the user and the Kinect. These normalization steps were not the general ones but they were working well for the proposed system.

The proposed gesture recognition module had different requirements than other recognition module used for interacting with computers. When gestures are used as an interface, every gesture should be distinguishable from other gestures so every gesture could have a unique meaning. For the presented system, it is not necessary every gesture word to have a unique gesture. We only need to check whether a user follows a gesture representing the word or the sentence shown in a monitor. Some words could be represented with similar gestures or some words could be represented with several varying gestures. Even if we showed the image shown in Figure 2 to represent the word “swing”, kids could do right-handed swing, left-handed swing and swings at various height. We have to link several gestures to the word that could have more than one gestures like the word “swing”.

Fig. 2. Swing gesture

The output module contained images, sounds and texts (Figure 1). Words or sentences were written on one side of a display as texts with corresponding pronunciations. Images representing gestures or images representing words or sentences were shown in the opposite side of the display. More than one image were used for some gesture words if gestures of words were not easily guessed by users with only one image.

3 Evaluation

We designed a user experiment to understand the usability of the presented system by comparing it with the traditional system. We measured a testing score after finishing

every learning period. Using these data, we wanted to find whether kids memorize words easier with the system using gestures than the system not using gestures.

Five ten-year-old kids (three girls and two boys) with normal or corrected-to-normal vision took part in the experiment. We collected words that were suitable for them. However, one of girls already knew seventeen words among eighteen words, so she was excluded from the experiment. Two sets of words used for the experiment are shown in Table 1.

Table 1. Two word lists for kids used in the experiment

Set 1		Set 2	
Doll	Guard	Glue	Listen
Candle	Neck	Airplane	Knee
Catch	Punch	Dig	Raise
Bottle	Parent	Captain	Mirror
Clapping	Storm	Flapping	Swing
Dictionary	Poor	Chalk	Quick
Fan	Salute	Dribble	Soap
Fold	Think	Waist	Unknown
Empty	Fix	Kitchen	Plan

The experiment used a two-factor mixed design. The between-subject factor was the learning system type, the system using gestures (SG) and the system not using gestures (SNG). The other within-subject factor was the learning period with three levels (two tests during the learning period and the final test). During the learning period, word tests were given to participants after learning the words and their memorizing test scores were recorded. We asked kids to learn the same set of word three times.

Each participant learned one set of words using the presented system and the other set of words using the traditional system. It is not possible to represent all English words using gestures. The set of words used for the presented system included words that was described using gestures, called gesture words, and words that was not described using gestures. Ten gesture words and ten non-gesture words were used for the experiment. Twenty words non-gesture words were used for the traditional system. We just presented words and corresponding images with pronunciations to users for the traditional system. No gestures were required for the traditional system.

The improvement of each participant was measured as the difference between the memorizing test scores after using each system and one before using the system. We could not use the memorizing test scores directly because participants had different knowledge about English words.

4 Evaluation Results and Discussion

Average improvement of the system using gestures (SG) and the system not using gestures (SNG) are summarized in Figure 3. The data shown in Figure 3 are the result of the experiment with kids. These data were measured as the difference between the initial and the each memorizing scores. The result exhibited SNG was better than SG for the first trial and SG was better than SNG after the third trial. Kids got almost perfect scores after the third trial.

Kids were shy when they tried SG first time. Kids only followed few required gestures for the first trial. As kids were familiar with SG, they followed more required gestures. This might be a reason for the result shown in Figure 3.

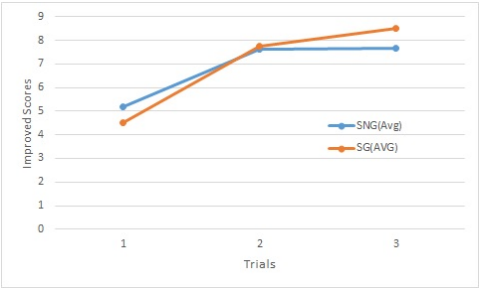


Fig. 3. Average improvement after using SG and SNG

5 Conclusions

The proposed system has been tested with five kids to understand whether adding gestures in the Computer Assisted English Learning (CAEL) System is beneficial to users. We have empirically compared the system using gestures and the system not using gestures in terms of testing scores. We found that it is not clear yet whether adding gestures to CAEL is beneficial to users or not. However kids using SG could achieve better improvement than kids using SNG after fourth or fifth trial since kids learned better when they were familiar with the system with gestures.

We also found that showing gesture images were not enough to guide gestures to kids. Kids were not sure when they had to do gestures even though the system asked kids to do gestures using written sentences. As a future work we will upgrade the proposed system so kids can follow gestures easily. Using the improved system we will perform user tests with many children with various word sets.

Acknowledgements. This work was supported by the Industrial Strategic technology development program, 10041772, (The Development of an Adaptive Mixed-Reality Space based on Interactive Architecture) funded by the Ministry of Science, ICT & Future Planning (MSIP, Korea).

References

1. Goldin-Meadow, S., Nusbaum, H., Kelly, S.D., Wagner, S.: Explaining math: gesturing lightens the load. Psychological Science 12(6), 516–522 (2001)
2. Cook, S.W., Goldin-Meadow, S.: The role of gesture in learning: Do children use their hands to change their minds? Journal of Cognition and Development 7(2), 211–232 (2006)
3. http://research.microsoft.com/en-us/projects/touchless/
4. https://en.wikipedia.org/wiki/Dynamic_time_warping

A Workflow Scheduling Technique for Task Distribution in Spot Instance-Based Cloud

Daeyong Jung[1], JongBeom Lim[1], Heonchang Yu[1],
JoonMin Gil[2], and EunYoung Lee[3,*]

[1] Dept. of Computer Science Education, Korea University, Seoul, Korea
[2] School of Information Technology Engineering, Catholic University of Daegu, Daegu, Korea
[3] Dept. of Computer Science, Dongduk Women's University, Seoul, Korea
{karat,jblim,yuhc}@korea.ac.kr, jmgil@cu.ac.kr,
elee@dongduk.ac.kr

Abstract. The cloud computing is a computing paradigm that users can rent computing resources from service providers as much as they require. A spot instance in cloud computing helps a user to utilize resources with less expensive cost, even if it is unreliable. In this paper, we propose the workflow scheduling scheme that reduces the task waiting time when an instance occurs the out-of-bid situation. And, our scheme executes user's job within selected instances and expands the suggested user budget. The simulation results reveal that, compared to various instance types, our scheme achieves performance improvements in terms of an average execution time of 66.86% over shortest execution time in each task time interval. And, the cost in our scheme is higher than an instance with low performance and is lower than an instance with high performance. Therefore, our scheme is difficult to optimize cost for task execution.

Keywords: Cloud computing, Spot instances, Workflow, Price history, Fault tolerance.

1 Introduction

Recently, due to increased interests for cloud computing many cloud projects and commercial systems such as Amazon EC2 [1], GoGrid [2], FlexiScale [3], have been implemented. Cloud computing provides high utilization and high flexibility for managing computing resources. And, cloud computing services provide a high level of scalability of computing resources combined with Internet technology to many customers [4]. In the most of these cloud services, the concept of an instance unit is used to provide users with resources in cost-efficient way. Generally instances are classified into two types: on-demand instances and spot instances. On-demand instances allow the user to pay for computing capacity by the hour with no long-term commitments. This frees users from the costs and complexities of planning,

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_52,

purchasing, and maintaining hardware and transforms what are commonly large fixed costs into much smaller variable costs [1]. On the other hand, spot instances allow customers to bid on unused Amazon EC2 capacity and run those instances for as long as their bid exceeds the current spot price. The price for spot instance changes periodically based on supply and demand, and customers whose bids meet or exceed it gain access to the available spot instances. If you have time flexibility for executing applications, spot instances can significantly decrease your Amazon EC2 costs [5]. For task completion, therefore, spot instances may incur lower cost than on-demand instances.

Spot market-based cloud environment configures the spot instance. This environment changes spot prices depending on the user's supply and demand. The environment affects the successful completion or failure of tasks in accordance with the changing of spot prices. Spot price has market structure and law of demand and supply. Therefore, cloud service (Amazon EC2) can provide a spot instance when a user's bid is higher than current spot price. And, a running instance stops when a user's bid becomes less than or equal to the current spot price. After a running instance stops, the running instance restarts when a user's bid is greater than the current spot price [6].

In particular, the scientific application makes the current common of workflow. However, the spot instance-based cloud computing takes various performance. In spot instance, an available execution time depends on spot price. The spot price is changing periodically based on user's supply and demand. The completion time for the same amount of a task varies according to the performance of an instance. In particular, the failure time of each instance differs according to the user's bid and the performance in an instance. Therefore, we solve the problem that a completion time of a task in an instance increases when a failure occurs. For an efficient execution of a task, the task is divided into sub-tasks on various types of available instances. We analyze information of the task and the instance from price history. We estimate the size of task and the information of an available instance from the analyzed data. We create workflow using each available instance and the size of task. As a consequence, we propose the scheduling scheme using workflow to solve job execution problem. And we execute user's job within selected instances and expand the suggested user budget.

2 Related Work

A workflow is a model that represents complex problems with structures such as directed Acyclic Graphs (DAG). Workflow scheduling is a kind of global task scheduling as it focuses on mapping and managing the execution of interdependent tasks on shared resources. However, the existing workflow scheduling methods have the limited scalability and are based on centralized scheduling algorithm. Consequently, this method is not suitable for spot instance-based cloud computing. In spot instance, the job execution has to consider available time and cost of an instance. Fully decentralized workflow scheduling system determines the instance to use the chemistry-inspired model in community cloud platform [7]. Throughput

maximization strategy is designed for transaction-intensive workflow scheduling that does not support multiple workflows [8]. Our proposed scheduling guarantees an equal task distribution to available instances in spot instance-based cloud computing.

3 Proposed Workflow Scheme

Our proposed scheme is expanded from our previous work [9] and includes a workflow scheduling algorithm. Fig. 1 presents the relation of workflows and instances and illustrates the roles of the instance information manager, the workflow manager, and the resource scheduler. The instance information manager obtains information for the job allocation and resource management. The information includes VM specifications in each instance and the execution-related information such as the execution costs, execution completion time, and failure time. The execution-related information is calculated by using the selected VM based on spot history. The workflow manager and resource scheduler extract the needed execution-related information from the instance information manager. Frist, the workflow manager generates the workflow for the request job. The generated workflow determines the task size according to the VM performance, the execution time and costs, and the failure time in the selected instance. Secondly, the resource scheduler manages the resource and allocates the task to handle the job. Resource and task managements are needed in order to reallocate when the resource cannot get the information for the task and when the task has a fault during execution.

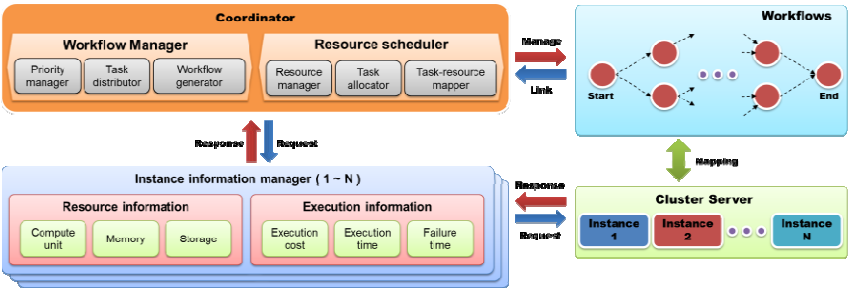


Fig. 1. The mapping relation of workflows and instances

To design the above model, our proposed scheme uses the workflow in spot instance and its purpose is to minimize job processing time within the suggested cost of user. The task size is determined by considering the availability and performance of each instance in order to minimize the job processing time. The available time is estimated by the execution time and cost using the price history of spot instances to improve the performance and stability of task processing. The estimated data is determined to assign the amount of task to each instance. Our proposed scheme reduces the out-of-bid situation and improves the job execution time. However, total cost is higher than when not using workflow.

Our task distribution method determines the size of the task to allocate a task to a selected instance. Based on a compute-unit and an available state, the task size of an instance $|T_i|$ is calculated by

$$|T_i| = \left(\frac{U_i \times A_i}{\sum_{i=1}^{N} (U_i \times A_i)} \right) \times \frac{1}{U_i} \times |T_{request}| \times U_{baseline} \quad (1)$$

where $|T_{request}|$ represents total size of tasks for executing user request. In an instance i, U_i and A_i represent the compute-unit and the available state, respectively. An available state A_i selects 0 (unavailable) or 1 (available). Fig. 2 shows the workflow scheduling scheme in our environment.

```
1:  Boolean S_flag = false    // a flag representing occurrence of a task execution
2:  while(search user's job) do
3:    if   (require job execution by the user) then
4:         take the cost and total execution time by the user;
5:         S_flag = true;
6:    end if
7:    if (S_flag) then
8:       invoke workflow ( );   // thread function
9:    end if
10: end while
11: Thread_Function workflow ( ) begin
12:   While (task execution does not finish) do
13:     forall   instance i ∈ Ins do
14:           retrieve an instance information to meet the user's requirement in an instance i;
15:       analyze an available execution time and cost in an instance i ;
16:       store the analyzed available instance to a queue_instance;
17:      end forall
18:     calculate on priority list for the priority job allocation;
19:     forall instance i ∈ queue_instance do
20:          allocate tasks to the instance i ;
21:     end forall
22:   end while
23: end   Thread_Function
```

Fig. 2. Workflow scheduling algorithm

4 Performance Evaluation

Our simulations were conducted using the history data obtained from the Amazon EC2's spot instances [10]. The history data before 10-01-2010 were used to extract the expected execution time and failure occurrence probability for our checkpointing scheme. The applicability of our scheme was tested using the history data after 12-20-2010.

In the simulations, one type of spot instance was applied to show the effect of an analysis – task time – on the performance. Table 1 shows various resource types used in Amazon EC2. In this table, resource types comprise a number of different instance

types. First, standard instances offer a basic resource type. Second, high-CPU instances offer more compute units than other resources, and can be used for compute-intensive applications. Finally, high-memory instances offer more memory capacity than other resources and can be used for high-throughput applications, including database and memory caching applications. Under the simulation environments, we compare the performance of our proposed scheme with that of the existing schemes without distributions of tasks in terms of various analyses according to the task time.

Table 1. Information of resource types

Instance type name	Compute unit	Virtual cores	Spot price min	Spot price average	Spot price max
m1.small (Standard)	1 EC2	1 core (1 EC2)	$0.038	$0.040	$0.053
m1.large (Standard)	4 EC2	2 cores (2 EC2)	$0.152	$0.160	$0.168
m1,xlarge (Standard)	8 EC2	4 cores (2 EC2)	$0.076	$0.080	$0.084
c1.medium (High-CPU)	5 EC2	2 cores (2.5 EC2)	$0.304	$0.323	$1.52
c1.xlarge (High-CPU)	20 EC2	8 cores (2.5 EC2)	$0.532	$0.561	$0.588
m2.xlarge (High-Memory)	6.5 EC2	2 cores (3.25 EC2)	$0.532	$0.561	$0.588
m2.2xlarge (High-Memory)	13 EC2	4 cores (3.25 EC2)	$0.532	$0.561	$0.588
m2.4xlarge (High-Memory)	26 EC2	8 cores (3.25 EC2)	$1.064	$1.22	$1.176

Table 2. Parameters and values for simulation

Simulation parameter	Task time interval	Baseline	Distribution time	Merge time	Checkpoint time	Recovery time
Value	43,200(s)	m1.xlarge	300(s)	300(s)	300(s)	300(s)

Table 1 shows various information of resource type in each instance and table 2 shows the parameters and values for simulation. The information of spot price is extracted from 11-30-2009 to 01-23-2011 in spot history. The user's bid is taken by the spot price average from information of spot price. The task size is decided by compute-unit rate based on baseline. Initially the baseline denotes an instance m1.xlarge. For example, the task size of an instance m1.small is calculated by the following.

$$| T_{m1.small} | = U_{m1.xlarge} / U_{m1.small} \times | T_{original\ task} |$$

Fig. 2 shows the simulation results about each instance. We consider performance condition of each instance. Each instance sets user's bid to take the spot price average

in table 2. Fig. 2 presents the execution time and costs according to various instance type. The instance with high performance reduces the execution time, but spends higher cost than the instance with low performance. As, in fig. 2(a), the total execution time increases, fig. 2(c) shows that the failure time increases. Fig. 2(d) shows the absolute time in each instance. Absolute time is the time interval between a failure occurrence time to the last checkpoint time.

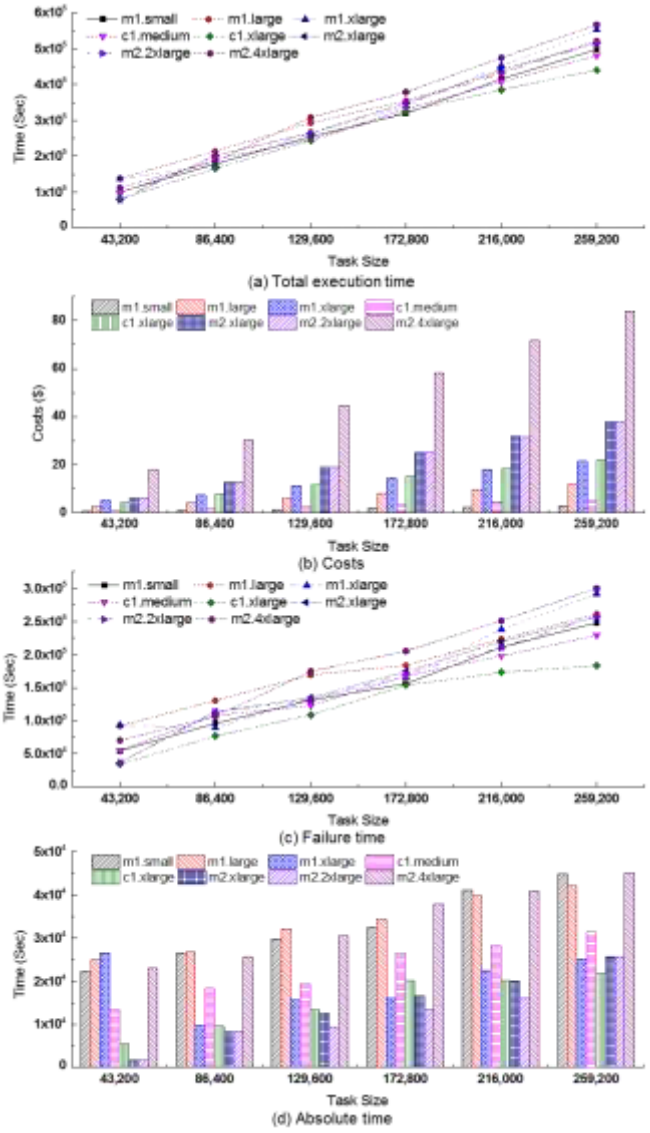


Fig. 3. Simulation result in each instance

Fig. 3 shows the simulation results about our proposed scheme using task distribution. Fig 3(a) shows the total execution time for each instance and their merge. The merge is a summation of the longest execution time in instance, task distribution time, and task merge time. The total execution time of the merge achieves performance improvements in terms of an average execution time of 66.86% over shortest execution time in each task time interval. In fig 3(b), the cost in our scheme increases an average of $15.337 than an instance m1.small and reduces an average of $ 34.214 than an instance m2.4xlarge. A failure time of fig 3(c) and an absolute time of fig 3(d) are smaller than those of fig 2(c) and fig 2(d).

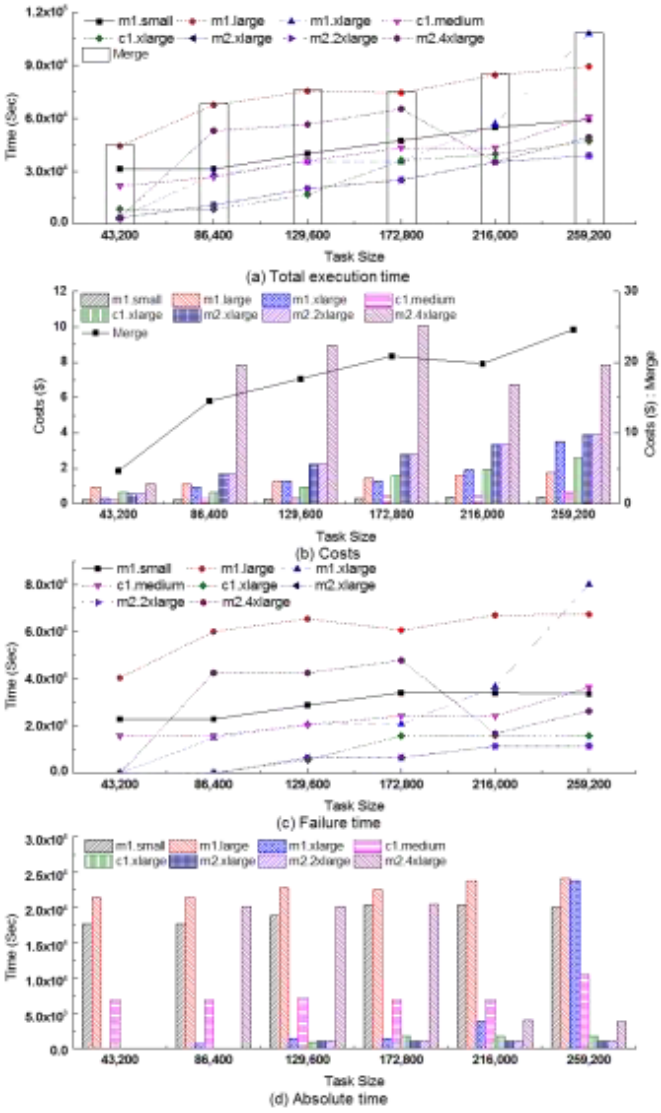


Fig. 4. Simulation result in task distribution

5 Conclusion

In this paper, we propose a workflow scheduling technique for task distribution in unreliable cloud computing environment. Our previous proposed checkpoint scheme takes a checkpointing based on two kinds of thresholds: price and time. Our proposed scheme reduces a failure time and an absolute time. The absolute time of our scheme can be lesser than that of the existing scheme without workflow because our scheme adaptively performs task distribution operation according to available instances. The simulation results show that our scheme achieves performance improvements in terms of an average execution time of 66.86% over shortest execution time in each task time interval. However, our scheme cannot optimize cost for task execution. In the future, we plan to expand our environment with an efficient task distribution operation and more efficient workflow method.

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 & Future Planning (NRF-2013R1A1A3007940).

References

1. Elastic Compute Cloud, EC2 (2013), http://aws.amazon.com/ec2
2. GoGrid (2013), http://www.gogrid.com
3. FlexiScale (2013), http://www.flexiscale.com
4. Van, H.N., Tran, F.D., Menaud, J.-M.: SLA-Aware Virtual Resource Management for Cloud Infrastructures. In: Proceedings of the 2009 Ninth IEEE International Conference on Computer and Information Technology, vol. 2, pp. 357–362. IEEE Computer Society (2009)
5. Amazon EC2 spot Instances (2013), http://aws.amazon.com/ec2/spot-instances/
6. Yi, S., Kondo, D., Andrzejak, A.: Reducing Costs of Spot Instances via Checkpointing in the Amazon Elastic Compute Cloud. In: Proceedings of the 2010 IEEE 3rd International Conference on Cloud Computing, pp. 236–243. IEEE Computer Society (2010)
7. Fernandez, H., Obrovac, M., Tedeschi, C.: Decentralised Multiple Workflow Scheduling via a Chemically-coordinated Shared Space. Research Report, RR-7925, pp. 1–14. Inria (2012)
8. Liu, K., Chen, J., Yang, Y., Jin, H.: A throughput maximization strategy for scheduling transaction-intensive workflows on SwinDeW-G. Concurrency and Computation: Practice and Experience 20, 1807–1820 (2008)
9. Jung, D., Chin, S., Chung, K., Yu, H., Gil, J.: An Efficient Checkpointing Scheme Using Price History of Spot Instances in Cloud Computing Environment. In: Proceeding NPC 2011, pp. 185–200 (2011)
10. Cloud exchange (2011), http://cloudexchange.org

Cost-Effective Content Delivery Networks Using Clouds and Nano Data Centers

Ikhsan Putra Kurniawan[1], Hidayat Febiansyah[2], and Jin Baek Kwon[2]

[1] PT Smartfren Telecom Tbk, Menteng, Central Jakarta, 10340, Indonesia
[2] Sun Moon University, Asan, Chungnam, 336708, Republic of Korea
{putra.ikhsan,havban}@gmail.com, jbkwon@sunmoon.ac.kr

Abstract. Videos-on-demand(VoD) services have already prevalent nowadays, because it is easier, cheaper, faster, and more comfortable way to watch movies at home. Also, the video quality has been increasing to satisfy the customers. The way that the video data is delivered to customers affects the total cost of the VoD service. Content delivery network(CDN) has been used as efficient content delivery, which consists of a number of replication servers around the world to minimize the latency, to avoid a single point of failure, and to distribute server load over the replicas. Recently, there has been some work on cloud-assisted VoD service that adopts a hybrid VoD delivery of centralized one and cloud CDN. This model can save the total cost for the VoD services significantly, compared to the centralized model. A novel concept called nano data center has been proposed, which is a network of gateways for internet services at home, which equipped with some storage. This nano data center can used for some large scale applications with much less energy, by exploiting the available resources of the home gateways that are on line almost all the time.

In this paper, to further reduce the total cost for delivery video contents to VoD subscribers, we propose a hybrid delivery model adopting both of cloud CDN and nano data centers, called *nano cloud CDN*. We formulated the total cost for our proposed model, and we have simulated the model under various parameters and scenarios to prove the effectiveness.

1 Introduction

Videos-on-demand(VoD) services have already prevalent nowadays, because it is easier, cheaper, faster, and more comfortable way to watch movies at home. Also, the video quality has been increasing to satisfy the customers. The way that the video data is delivered to customers affects the total cost of the VoD service. Content delivery network(CDN) has been used as efficient content delivery, which consists of a number of replication servers around the world to minimize the latency, to avoid a single point of failure, and to distribute server load over the replicas. Recently, some researchers proposed to implement CDNs with cloud platforms for VoD services for a cost reduction[1–5]. Also there have been studies on introducing peer-to-peer(P2P) architecture into VoD systems[6–8].

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_53,

Lin et al. proposed a cloud based content delivery network, its structure integrates cloud infrastructure and content delivery network system together [3]. Its system uses Hadoop Distributed File System as a cloud infrastructure services, and creates many data clusters around the world. Therefore, content providers could place content to closest node from end users. Wang et al. also presented a CDN system based on cloud storage[5]. Huang [6] provides the general framework for further research on P2P-VoD systems, including how to design a highly scalable P2P-VoD system to support millions of simultaneous users, how to perform dynamic movie replication strategies, and etc. However, P2P solution becomes unreliable for large scale VoD, when users do not want to participate in it, or when there is not enough available upload bandwidth.

Li et al. proposed a cloud-assisted VoD that adopts a hybrid VoD delivery of centralized one and cloud CDN such as Amazon Cloud Front and Azure CDN[4]. In the cloud-assisted VoD, the service operates with both of cloud CDN and central servers in peak time, while only with central servers in off-peak time. This model can save the total cost for the VoD services up to 30 %, compared to the centralized model. However, all of them do not consider another improvement by using other devices, closer to the clients. Valancius et al. proposes a novel concept called *nano data center*, which is a network of gateways for internet services at home, which equipped with some storage[8]. This nano data center can used for some large scale applications with much less energy, by exploiting the available resources of the home gateways that are on line almost all the time. In the work, they presented VoD service as an example application of the nano data centers, where the home gateways cooperate to store and deliver the video contents just like P2P networks.

In this paper, to further reduce the total cost for delivery video contents to VoD subscribers, we propose a hybrid delivery model adopting both of cloud CDN and nano data centers, called *nano cloud CDN*. We formulated the total cost for our proposed model, and we have simulated the model under various parameters and scenarios to prove the effectiveness.

2 Nano Cloud CDN

2.1 Overview

There are a lot of opportunities that Nano Cloud CDN can exploit from the distribution of VOD on internet. These opportunities cover the popularity of Cloud services on internet that offers many services and features which en-abled pay-as-you-go paradigm. Even though the price of cloud bandwidth is more expensive than ISP bandwidth in term of unit price, the total cost of bandwidth could be different since ISP bandwidth is charged by 95 percentile rules. The same goes with the cloud storage, the storage price keeps decreasing nowadays. Another opportunity is in the scalability of P2P delivery model. Both centralized delivery model and Cloud CDN delivery model have a very bad scalability. When the user increases or the movie library increases, VoD providers must pay great cost to handle it, especially for centralized delivery model. P2P delivery model on

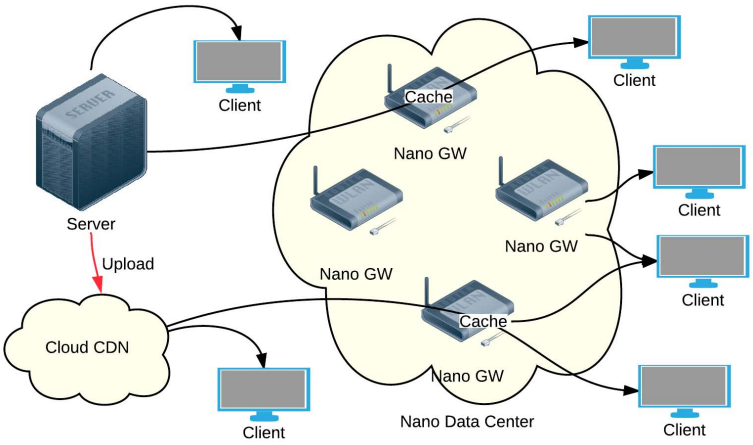


Fig. 1. Nano Cloud CDN

the other hand has a very good scalability. As the number of client increase, the storage capacity and bandwidth capacity increase as well. Another opportunity is in the popularity pattern of VoD on internet related to its hit request. The popularity of video on internet follows Zipf distribution. As mentioned by [4] 50% of requests are going to 10% popular movies. And 70% of total requests are going to 20% of popular movies.

Fig. 1 illustrates the overall model of the nano cloud CDN, which consists of three main components: a server, a cloud CDN, and a nano data center. We assume that some clients have nano gateways in front of them. The server is a source of VoD services that contains all video contents, receives all requests, and decides where the requested contents are delivered from, either the server, the cloud CDN, or the nano data center. The cloud CDN stores a subset of all the movies of the service provider, and delivers them to the requesting clients just like traditional CDNs. The set of movies are uploaded to the cloud in batch, regardless of user requests. And the nano data center caches the movie on the nano gateways through which the streams go. A stream can be established between the server and the client or the cloud, between the cloud to the client, or between the nano data center to the client. This decision depends on the available bandwidth between source and destination. And a client can be streamed from multiple sources like P2P streaming, which is called collaborative streaming. And the client can be served also by its own nano gateway if the requested content are cached in it.

2.2 Total Cost Formulation

First of all, we consider the following components to define a formula for total cost of the VoD service in the nano cloud CDN. Each symbol indicate the cost of each component.

- Server downstream bandwidth (B_{server}): Server bandwidth used to serve clients and to upload movies to cloud CDN.
- Cloud storage size (S_{cloud}): Amount of the movies stored in the cloud CDN.
- Cloud downstream bandwidth ($B_{cloud,down}$): Cloud bandwidth used to serve clients for the nano cloud CDN.
- Cloud upstream bandwidth ($B_{cloud,up}$): Cloud bandwidth used to upload movies from the central server to the cloud CDN.
- Nano gateway storage size (S_{nano}): Storage size in nano gateways to cache movies.
- Nano gateway upstream bandwidth (B_{nano}): Upstream bandwidth provided by nano gateways.

Then, the total cost is defined as follows:

$$C_{total} = B_{server} + S_{cloud} + B_{cloud,down} + B_{cloud,up} + S_{nano} + B_{nano}. \quad (1)$$

This formula is used to compare the total cost of video delivery in a centralized model, a cloud CDN model, and our nano cloud CDN model. For the centralized model, all parameters except B_{server} are zero, and for the cloud CDN model, S_{nano} and B_{nano} are zero.

2.3 Dynamic Strategy

In centralized delivery model, the delivery strategy is quite simple; the server serves all the requests from clients. In Cloud CDN delivery model, that is enhanced by uploading popular movies into the cloud CDN during in idle time and the cloud CDN reduces the server load while serving requests during the peak time. Thus it can remove the spike of server bandwidth usage. By adopting nano data center concept into the cloud CDN, the nano cloud CDN more can enhance the delivery strategy into the next level. However, we have to concern about many issues or parameters before devising the strategy with the minimum cost of delivery. The total cost are affected largely by the number of clients, the number of movies served, and the ratio of nano gateways to the number of clients (We assume that a clients may be behind a nano gateway or not). The output of the strategy is the server bandwidth used for service, the ratio of movies uploaded to the cloud CDN, and the replacement algorithm in nano gateways' caches. The nano cloud CDN finds the minimal total cost of delivery using dynamic strategy which consists of two algorithms:

- Finding minimal total cost
  This algorithm searches the solution that has minimal total cost of delivery by . Then we compare each solution by changing server bandwidth allocation, percentage of movies in cloud, and nano cloud CDN replacement methods. The algorithm is shown in Fig. 2.
- Serving requests to get minimal total cost.
  This algorithm decides the best place to serve requests from clients. There are three possible source to stream, the first is nano gateways, the second is the central server, and the last is the cloud CDN. The algorithm is shown in Fig. 3

```
Initiate variables: Total clients, Nano Cloud Client Percentage, Total movies, ISP
bandwidth price, Cloud bandwidth price, Cloud storage price, Nano Cloud Client
bandwidth price, Nano Cloud Client storage price.
For server allocation bandwidth from 0% to 100%
    For cloud video percentage from 0% to 100%
        For each nano cloud client cache replacement method from Random, LRU,
        GLRU, LFU, GLFU
            Find the total cost of video delivery for this scenario
            If the total cost of this scenario is less than the previous scenario
                This total scenario is the scenario with minimum total cost
            End if
        End loop
    End loop
End loop
```

Fig. 2. Algorithm to finding minimal total cost. The nano cloud CDN runs all possible solutions to search the minimal total cost of delivery.

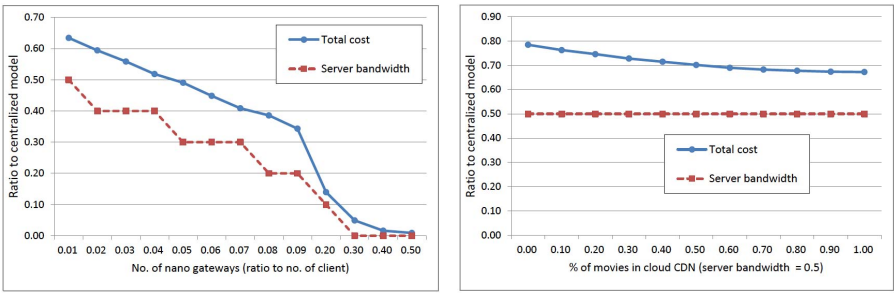
```
If the client is a nano cloud client
    If the client has already cached the movie
        Self stream
    Else if there is another nano cloud client that want to stream the same
    movies at the time
        If another nano cloud client is available to upstream the movie
            Collaborative stream from nano cloud client
        Else if server is available to upstream the movie
            Collaborative stream from server
        Else
            If cloud CDN has the movie
                Collaborative stream from cloud CDN
            Else
                Force server to stream to cloud, forward the stream
                to the client while storing the movie in the cloud
                CDN.
                Collaborative stream from cloud CDN
    Else
        If another nano cloud client is available to upstream the movie
            Stream from nano cloud client
        Else if server is available to upstream the movie
            Stream from server
        Else
            If cloud CDN has the movie
                Stream from cloud CDN
            Else
                Force server to stream to cloud, forward the stream
                to the client while storing the movie in the cloud CDN
                Stream from cloud CDN
Else
    If another nano cloud client is available to upstream the movie
        Stream from nano cloud client
    Else if server is available to upstream the movie
        Stream from server
    Else
        If cloud CDN has the movie
            Stream from cloud CDN
        Else
            Force server to stream to cloud, forward the stream to the
            client while storing the movie in the cloud CDN
            Stream from cloud CDN
End if
```

Fig. 3. Algorithm to serve requests to produce the minimal total cost. A request is served by nano gateways, the server, or Cloud CDN in priority order.

3 Performance Evaluation

For the experiments, we built the simulation program and used a synthetic video popularity model and the trace on VoD user request pattern from other researches. The simulation uses the trace by using a user request pattern and video popularity pattern from previous researches. The main performance metric is total cost and initially the nano gateways has nothing in their cache and the cache replacement algorithm is a random one in default. The total cost depends largely on the price of resources and we set up the price parameters as follows:

- Cloud upstream bandwidth price is taken from Amazon EC2 website[9].
- Cloud storage bandwidth price is taken from Amazon S3 website[10].
- Cloud downstream bandwidth price is free or zero.
- ISP upstream bandwidth price is half of cloud upstream bandwidth price. (This is similar to [4]).
- Nano Cloud Client upstream price is free or zero.
- Nano Cloud Client storage price is taken from Kingston flash drive 16 GB[11].



(a) Nano cloud CDN (b) Cloud-assisted CDN

Fig. 4. Total cost comparison: The server bandwidth is fixed to 0.5 for the cloud CDN model, which is optimal empirically

Fig. 4 shows the total costs of the nano cloud CDN and Li's cloud-assisted CDN[4], where the total cost is the proportion to the base model, i.e., the centralized mode, when we assume 5,000 movies and 5,000 clients. The performance metrics of y-axis are the proportion to the base model. The nano cloud CDN keeps more movies on nano gateways as the ratio of nano gateways to the number of clients increases. At the same time, the server bandwidth usage also goes down. In the cloud-assisted CDN, both of server bandwidth limit and the ratio of movies stored in the cloud CDN are the system parameters, not metric. We found that the best server bandwidth limit for cloud-assisted VoD is around 0.5 or 50% compare to the base model. That is why we fixed the server bandwidth to 0.5. As shown in the figure, as expected, the nano cloud CDN is much more

cost-effective compared to the cloud assisted model. The cloud-assisted model saves about 30% of total cost when the half of movie library is stored in the cloud CDN. The nano cloud CDN on the other hand has a very good scalability. The more nano gateways join the system, it saves more cost. When the ratio of nano gateway is 0.3, it is enough for us to be served without consuming the server bandwidth.

Since the actual prices are not ranging from one provider to another, we conducted various experiment to show the effect of the prices of the resources. We assume that the nano gateway ratio is 20% in this experiment. First, we compare the effect of the different price ratio of server bandwidth price(B_{server}) charged by ISP[12] to cloud downstream bandwidth price($B_{cloud,down}$ and $B_{cloud,up}$) charged by cloud service provider [13]. Table 1 shows that the cost increase as the cloud bandwidth price goes up, but the effect is slight for the nano cloud CDN when the nano gateway ratio is sufficient.

Table 1. Total cost vs. $B_{server}/B_{cloud,down}$

$B_{server}/B_{cloud,down}$	Cloud-assised CDN	Nano cloud CDN
0.25	0.8386	0.4684
0.5	0.7068	0.3878
0.75	0.5936	0.3258
1	0.5017	0.2720

We do not have sufficient information on the pricing to cloud upstream bandwidth. But, what we found is the uploading price in Amazon S3 is free [10]. However, assuming it is not free, we did the simulation by varying the cloud upstream bandwidth price. Similar to the above results, Table 2 shows that the cloud upstream bandwidth does not affect both Cloud CDN and Nano Cloud CDN at all.

Table 2. Total cost vs. $B_{cloud,up}/B_{cloud,down}$

$B_{cloud,up}/B_{cloud,down}$	Cloud CDN	Nano cloud CDN
0.25	0.6939	0.3884
0.5	0.6939	0.3884
0.75	0.6939	0.3884
1	0.6939	0.3884

We investigate the effect of cache replacement algorithm on the cost in the nano gateways with LRU, GLRU, LFU, GLFU. From the results. we conclude that when nano gateway ratio is below 30%, LRU, GLRU, LFU, and GLFU have better efficiency than random replacement, but the difference of the algorithms is minor. Due to the lack of space, the results are omitted here.

4 Conclusion

Content delivery network(CDN) has been used as efficient content delivery, which consists of a number of replication servers around the world to minimize the latency, to avoid a single point of failure, and to distribute server load over the replicas. Recently, some researchers proposed to implement CDNs with cloud platforms for VoD services for a cost reduction. Also there have been studies on introducing peer-to-peer(P2P) architecture into VoD systems.

In this paper, we propose a novel video delivery model by combining the cloud CDN and the nano datacenter concept, and formulate the cost function. Also we suggest the dynamic strategy to achieve the minimal cost. Throughout the simulation, we show that our nano cloud CDN model save around 50% of total cost with only 7% of nano clients with a 16 GB flash memory.

References

1. Nezhad, H.M., Stephenson, B., Singhal, S.: Outsourcing business to cloud computing services: Opportunities and challenges. Technical report, HP Laboratories/HPL-2009-23 (2009)
2. Hajjat, M., Sun, X., Sung, Y., Maltz, D., Rao, S., Sripanidkulchai, K., Tawarmalani, M.: Cloudward bound: Planning for beneficial migration of enterprise applications to the cloud. In: Proc. of ACM SIGCOMM (2010)
3. Lin, C.F., Leu, M.C., Chang, C.W., Yuan, S.M.: The Study and Methods for Cloud based CDN. In: Proc. of IEEE CyberC (2011)
4. Li, H., Zhong, L., Liu, J., Li, B., Xu, K.: Cost-effective partial migration of vod services to content clouds. In: Proc. of IEEE CLOUD (2011)
5. Wang, Y., Wen, X., Sun, Y., Zhao, Z., Yang, T.: The content delivery network system based on cloud storage. In: Proc. of IEEE NCIS (2011)
6. Huang, Y., Fu, T.Z., Chiu, D.M., Lui, J.C., Huang, C.: Challenges, design and analysis of a large-scale p2p-vod system. In: Proc. of ACM SIGCOMM (2008)
7. Kumar, R.A., Ganeshan, K.: A novel dynamic pricing scheme for contributing peers in the vod system. Multimedia Tools and Applications 58(3), 613–632 (2012)
8. Valancius, V., Laoutaris, N., Massouli, L., Diot, C., Rodriguez, P.: Greening the internet with nano data centers. In: Proc. of ACM CoNEXT (2009)
9. Amazon: Amazon ec2 pricing, `http://aws.amazon.com/ec2/pricing/`
10. Amazon: Amazon s3 pricing, `http://aws.amazon.com/s3/pricing/`
11. Kingstone: Kingstone flash drive 16gb, `http://www.kingstone.com/us/`
12. Sapporo, I.I.: Isp comparison, `http://www.ispcompared.com/broadband.htm`
13. Amazon: Amazon cloudfront, `http://aws.amazon.com/cloudfront/`

Adaptive Transformation for a Scalable User Interface Framework Supporting Multi-screen Services

Yuseok Bae, Bongjin Oh, and Jongyoul Park

Electronics and Telecommunications Research Institute,
218 Gajeong-ro, Yuseong-gu, Daejeon, 305-700, Korea
{baeys,bjoh,jongyoul}@etri.re.kr

Abstract. Nowadays many smart devices getting introduced to the market have different performance and capabilities in terms of CPU, memory, screen size, screen resolution, and etc. In this paper, we propose an adaptive transformation for a scalable user interface (SUI) framework supporting multi-screen services which is capable of providing a uniform user experience irrespective of devices of varying size and capabilities. The proposed adaptive transformation dynamically adapts user interfaces to make it suitable for smart devices using transformation policy including layout, content, and appearance of user interfaces as well as device profiles regarding device capabilities.

Keywords: Adaptive Transformation, SUI Framework, Multi-Screen Services, Smart Devices.

1 Introduction

A rapid growth of smart devices such as smart TVs, smartphones, and smart pads is likely to drive a new user interface (UI) framework capable of providing optimal user interfaces suitable for these smart devices considering device performance and capabilities. In addition, it is crucial to support multi-screen services using contents of one source to improve productivity without creating separate contents about a variety of smart devices.

Currently, most of smart devices enable a full-screen browsing based on the HTML5 [1], but it is a possibility of occurring a blurriness or a distortion in terms of screen scaling regarding various screen resolutions. Hence, many service providers maintain UI pages for mobile devices as well as UI pages for desktop PCs in their separate ways.

Moreover, the responsive web design (RWD) [2] is recently attractive because it is capable of adapting layout, content, and appearance according to screen resolution by using CSS3 media queries [3], but still has problems about performance issues of media queries and some older versions of browsers do not support it. Furthermore, the Scalable Vector Graphics (SVG) [4] supports a seamless scaling but is not appropriate for smart devices because it still requires high performance to parse XML-based SVG files and render them.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_54,

Meanwhile, acceptance-diffusion strategies [5] describe the acceptance model for tablet-PCs revealed that playability, cost level, functionality, and complexity significantly affect user acceptance. Especially, in the diffusion model, it showed that playability and user interface have a significant influence on satisfaction, trust and positive behavioral intention. Besides, a new framework for context-aware service composition [6] describes mechanisms for context-aware service discovery, composition, and provisioning in an adaptive manner in order to easily accommodate new services. Moreover, the user adoption model [7] explains the adoption of a new technology from a continuous usage intention of an existing technology.

Therefore, a new scalable user interface (SUI) framework is crucially needed to improve productivity by providing functionality to adaptively transform UI contents of one source appropriate to size and capabilities of smart devices in an effective way.

Consequently, in this paper, we propose an adaptive transformation for a SUI framework to efficiently support multi-screen services considering size and capabilities of smart devices. The proposed adaptive transformation dynamically adapts user interfaces to make suitable for smart devices based on the SUI framework using not only device profiles about size and capabilities of smart devices, but also transformation policies to adapt layout, content, and appearance of user interfaces. In addition, the SUI framework supports efficient collaboration for automatic collection of device profiles and message exchange among smart devices in home networks.

The remainder of this paper is organized as follows. Section 2 briefly describes the SUI framework including working model and system architecture to support the adaptive transformation. Section 3 presents the adaptive transformation for the SUI framework supporting multi-screen services including sequence diagram and properties of transformation policy in order to adapt user interfaces dynamically. The implementation and result is then described in Section 4. Finally, we summarize and conclude this paper in Section 5.

2 SUI Framework

In this section, the working model and system architecture of the SUI framework supporting the adaptive transformation in terms of multi-screen services are described.

2.1 Working Model

Fig. 1 shows a brief working model of the SUI framework to support the adaptive transformation for multi-screen services in smart devices.

The SUI framework works through three kinds of steps: authoring of original UI contents, adaptive transformation of UI contents, and rendering of adapted UI contents.

The SUI Authoring Tool (SAT) enables users to generate original UI contents using UI templates and to emulate them about each device.

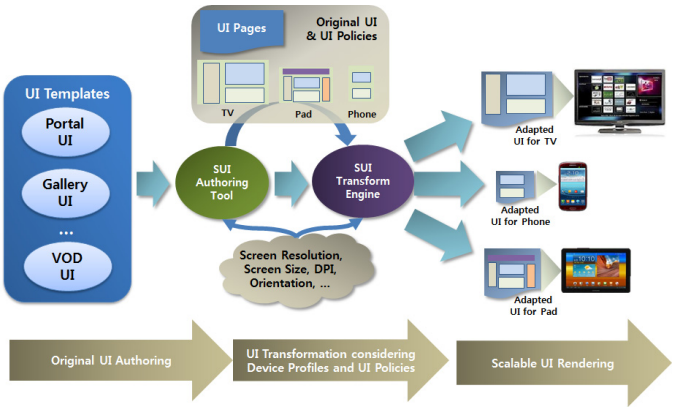


Fig. 1. Working model of the SUI framework

A UI template is a skeleton composed of web-based UI pages to support convenient UI authoring in the SAT. In addition, the SAT helps users to define UI transformation policies about the original UI contents considering size and capabilities of smart devices.

The SUI Transformation Engine (STE) dynamically creates adapted UI contents by performing transformation about original UI contents considering not only smart device's profile such as screen resolution, screen size, orientation, and dots per inch (DPI), but also transformation policies to adapt layout, content, and appearances. Finally, each device renders adapted UI contents suitable for its capabilities.

2.2 System Architecture

Fig. 2 shows system architecture of the SUI framework to support the adaptive transformation for multi-screen services in smart devices.

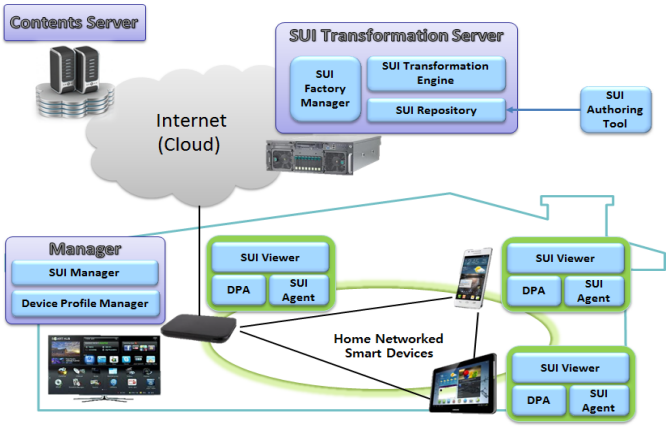


Fig. 2. System architecture of the SUI framework

The SUI Transformation Server has three sub-components for the adaptive transformation. First, the SUI Factory Manager (SFM) intermediates UI request and response messages between the STE and the SUI Manager (SM). Second, the STE generates adapted UI contents by transforming original UI contents considering device profiles and transformation policies. Third, the SUI Repository (SR) stores original UI contents and transformation policies edited using the SAT.

On the other hand, two kinds of manager components are designed for the adaptive transformation and can be installed on a device such as a smart set-top box (STB) within home networks. The former, the SM manages SUI Agents (SAs) and transfers UI request messages to the SFM and receives adapted UI contents as a response message. The latter, the Device Profile Manager (DPM) collects device profiles from DPAs through automatic discovery of the UPnP protocol.

In addition, each smart device has three sub-components for the adaptive transformation. The Device Profile Agent (DPA) communicates with the DPM via the UPnP protocol and provides its own device profile information such as device identifier, device model, screen size, screen resolution, orientation, DPI, and so on. The SA communicates with the SM in order to request a list of available UIs and an adapted UI contents appropriate to its device profile and to receive them as response messages. The SUI Viewer (SV) renders adapted UI contents suitable for its device capabilities delivered from the SA.

3 Adaptive Transformation for the SUI Framework

This section includes a sequence diagram and properties of UI transformation policy to describe the proposed adaptive transformation which adapts user interfaces dynamically using transformation policies as well as device profiles.

3.1 Sequence Diagram

Fig. 3 shows a sequence diagram of adaptive transformation for the SUI framework between servers and smart devices.

The DPM automatically collects device profiles from DPAs via the UPnP protocol. An SA requests a service UI list to the SM and the SM returns a service UI list received from the SR. When the SA selects a service UI, it transfers a pair of identifiers (device_id, ui_id) to the SM. The SM receives a device_profile mapped to the device_id from the DPM and requests a service UI to the SFM using parameters such as device_profile and ui_id. Next, the SFM requests transformed UI contents to STE and the STE creates adapted UI contents by dynamically transforming original UI contents considering device profile and transformation policies after receiving original UI contents and transformation policies from the SR. The STE returns adapted UI contents to the SFM and the adapted UI contents are returned to the SA. Finally, the SV renders the adapted contents delivered from the SA.

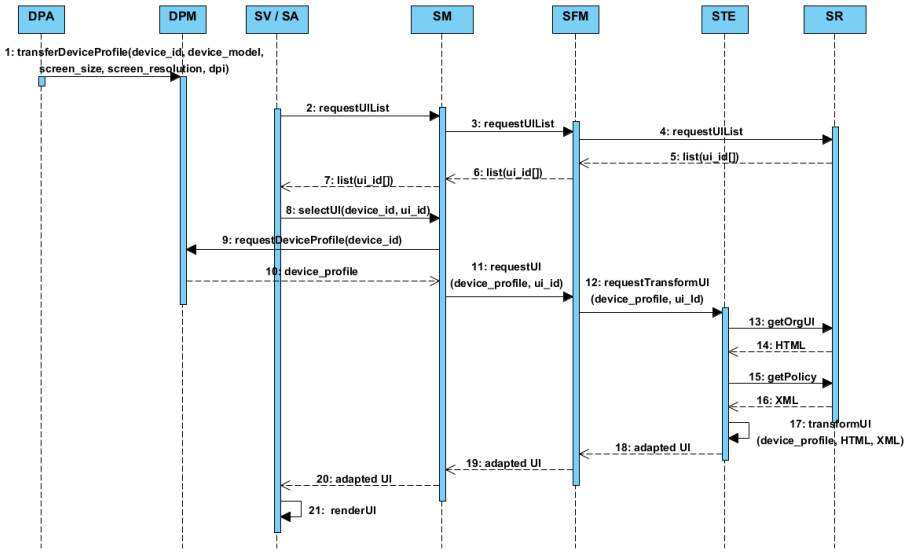


Fig. 3. Sequence diagram of adaptive transformation

3.2 Properties of UI Transformation Policy

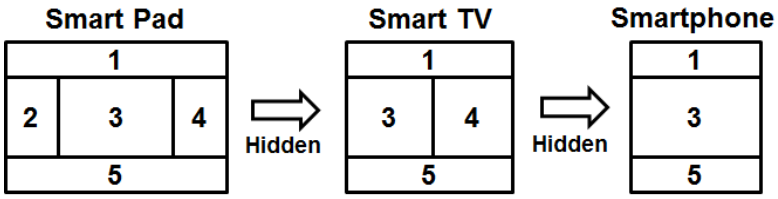
Table 1 shows properties of UI transformation policy used for the adaptive transformation and these transformation policies are stored as XML documents.

Table 1. UI Transformation Policy Properties

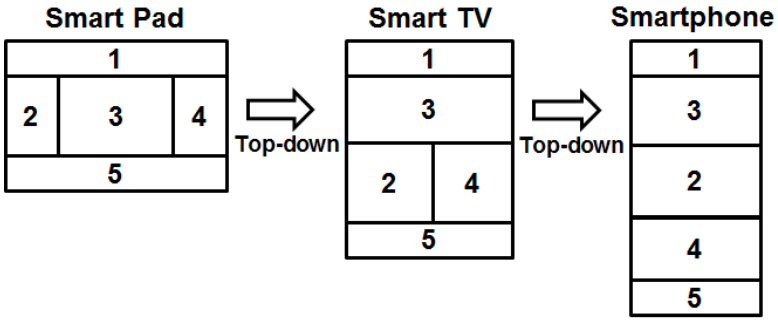
Property	Description	Example
class	UI class name	UI3
priority	Priority of the class	3
scaleWidth	Scale ratio of width	50%
scaleHeight	Scale ratio of height	50%
scaleCountTag	Number of contents to scale	2
inputCSS	CSS file defined by user	Phone.css
scaleRule	Transformation rule (Hidden, Page, Top-down)	Page

A class is the name of UI class to apply an adaptive transformation and is defined as a HTML div element. A priority means the priority order of the class in the UI page. The scaleWidth and scaleHeight is the properties for scale ratio about width and height. The scaleCountTag is the number of contents to scale. The inputCSS is the name of the user-defined CSS. The adaptive transformation defines three kinds of transformation rules: Hidden, Page, and Top-down. The Hidden rule is used for hiding UI contents and the Top-down rule is used for displaying UI contents using top-down sequencing according to the priority order. Finally, the Page rule is used for displaying UI contents through separate pages according to the priority order.

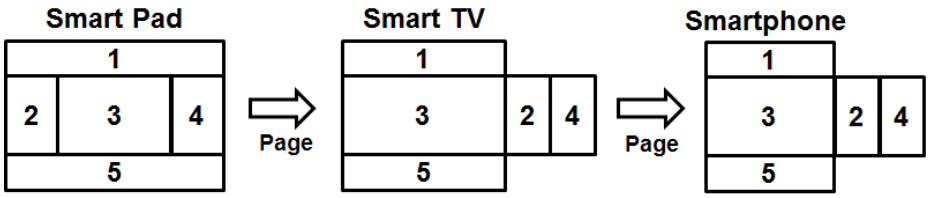
Fig. 4 (a), Fig. 4 (b), and Fig. 4 (c) respectively show working scenarios of three kinds of transformation rules in smart devices.



(a) Hidden



(b) Top-down



(c) Page

Fig. 4. Working scenarios of transformation rules

Especially, Fig. 4 (b) and Fig. 4 (c) show examples of UI transformation according to the priority orders of the UI classes.

4 Implementation and Result

In order to demonstrate the feasibility of the adaptive transformation for the SUI framework supporting multi-screen services in smart devices, we implemented manager components and agents based on the Android platform. In addition, we implemented server components for the adaptive transformation on the Linux operating system. Moreover, we developed a web portal service as an example using the SAT and defined transformation policies about each smart device.

Fig. 5 shows a screenshot including a series of actions such as content authoring using Eclipse, transformation policies about smart devices, and emulation about smartphone in the SAT.

Fig. 5. Screenshot of the SAT

Fig. 6 illustrates a screenshot of adapted UI contents shown in each device through the adaptive transformation for the SUI framework supporting multi-screen services in home networks.

Fig. 6. Screenshot of adapted UI contents

Each device requires adapted UI contents appropriate to its profile. The STE dynamically transforms original UI contents using the profile and transformation policies. The STE generates adapted UI contents and transfers them to the SA of each device through the SM. Finally, the SV renders adapted UI suitable for its profile delivered from the SA.

A smart STB shows UI layout applying Hidden rules to some of UI contents.

A smart pad shows UI layout using Top-down rules and a smartphone shows UI layout combining both Top-down and Page rules considering a limited screen size.

The proposed framework performs the adaptive transformation via the STE considering transformation policies as well as device profiles, whereas the RWD [2] self-adapts UI layout within each device. Thus, the proposed framework's message overhead is partially increased but device overhead is less than the RWD to parse and render UI contents. Moreover, the proposed framework provides more flexible architecture than the RWD to support multi-screen services because both device profiles and transformation policies can be easily extended and applied to the SUI framework according to capabilities of smart devices.

5 Conclusions

In this paper, we have presented an adaptive transformation for the SUI framework supporting multi-screen services in home networks. Also, the proposed SUI framework provides a flexible architecture to dynamically adapt layout, content, and appearance according to transformation policies as well as device profiles.

In future work, we will continue further research to reduce message overhead and improve performance. In addition, we will develop more efficient architecture to accommodate heterogeneous platforms as well as the Android platform.

Acknowledgments. This work was supported by the IT R&D program of MKE/KEIT, [10039202, Development of SmartTV Device Collaborated Open Middleware and Remote User Interface Technology for N-Screen Service].

References

1. Berjon, R., Leithead, T., Navara, D., O'Connor, E., Pfeiffer, S., Hickson, I.: HTML5: A vocabulary and associated APIs for HTML and XHTML. W3C Candidate Recommendation (2011)
2. Gardner, B.S.: Responsive Web Design: Enriching the User Experience. Noblis Sigma: Inside the Digital Ecosystem – Capitalizing on Connectivity 11(1), 13–19 (2011)
3. Rivoal, F.: Media Queries. W3C Recommendations (2012)
4. Dahlstrom, E., Dengler, P., Grasso, A., Lilley, C., McCormack, C., Schepers, D., Watt, J., Ferraiolo, J., Fujisawa, J., Jackson, D.: Scalable Vector Graphics (SVG) 1.1. In: W3C Recommendation, 2nd edn. (2011)
5. Kim, Y.J., Sim, J.B.: Acceptance-Diffusion Strategies for Tablet-PCs: Focused on Acceptance Factors of Non-Users and Satisfaction Factors of Users. ETRI J. 34(2), 245–255 (2012)
6. Gonzalez, A.J., Martin de Pozuelo, R., German, M., Alcober, J., Pinyol, F.: New Framework and Mechanisms of Context-Aware Service Composition in the Future Internet. ETRI J. 35(1), 7–17 (2013)
7. Jeon, H., Shin, Y., Choi, M., Rho, J.J., Kim, M.S.: User Adoption Model under Service Competitive Market Structure for Next-Generation Media Services. ETRI J. 33(1), 110–120 (2011)

Towards Nearest Collection Search on Spatial Databases

Hong-Jun Jang[1], Woo-Sung Choi[1], Kyeong-Seok Hyun[1], Kyoung-Ho Jung[1], Soon-Young Jung[1,*], Young-Sik Jeong[2], and Jaehwa Chung[3]

[1] Dept. of Computer Science Education, Korea University, Seoul, Korea
[2] Dept. of Multimedia Engineering, Dongguk University, Seoul, Korea
[3] Dept. of Computer Science, Korea National Open University, Seoul, Korea
{hongjunjang,ws_choi,ks_hyun,jungkh,jsy}@korea.ac.kr,
ysjeong@dongguk.edu, jaehwachung@knou.ac.kr

Abstract. In this paper, for the first time, we present the concept of nearest collection (NC) search. Given a set of spatial data points D and a query point q, a nearest collection search retrieves a certain subset c ($|c| = k$), called collection from D. We formally define a *collection* as clustered k objects and the *nearest collection search* problem. Since the brute-force approach of this problem requires large computational cost, we propose two approaches using database techniques to reduce search space. The first approach is the multiple query method which uses existing method (i.e. k-nearest neighbor query) based on normal R-tree. The second approach is the effective NC query processing based on the branch and bound method using an aggregate R-tree (simply aR-tree). Our experimental results show that the efficiency and effectiveness of our proposed approach.

Keywords: Spatial database, nearest collection query, k-nearest neighbor query.

1 Introduction

Spatial objects have been used in a large number of applications from people's daily life (e.g., location-based service, etc.) to scientific topics [1]. Particularly, technological advancement such as GPS and wireless communication leads the proliferation of location-based services on mobile devices. For example, people may depend on online map services to plan their trips. Driven by increasing number of these applications, efficient processing of spatial query has always been very important.

The researches in spatial databases have focused on retrieving nearest neighbor [5] and its variants [6], [7] (e.g., reverse nearest neighbor, group nearest neighbor, etc.). The k-nearest neighbor search in spatial databases retrieves the k objects from a datasets D that are closest to a query point q. The reverse nearest neighbor search retrieves the spatial points whose nearest neighbor is a given query point. The group

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_55,

nearest neighbor search retrieves data points that minimize their sum aggregate distances to given set of query points.

In this paper, we propose a new type of search, called nearest collection (NC) search, which has many applications in LBSs and decision support systems. Unlike nearest neighbor and its variants searches, NC search find the nearest subset (called collection) that satisfies given condition (where sum aggregate distances between all pairs of the elements in a subset, is less than threshold δ) from a given query point.

In Fig. 1, it show the result of k-NN search and NC search where k is 3 and collection threshold δ is 30. The result of k-NN search $\{p_1, p_2, p_3\}$ shows the k nearest object from the query point. However, the result of NC search $c_1 = \{p_1, p_4, p_5\}$ shows a nearest set of points among the sets, where the sum of distances between all pairs of the elements in a set c_1 is less than 30. As a result, the NC search can be applied to following examples and many applications in spatial domain.

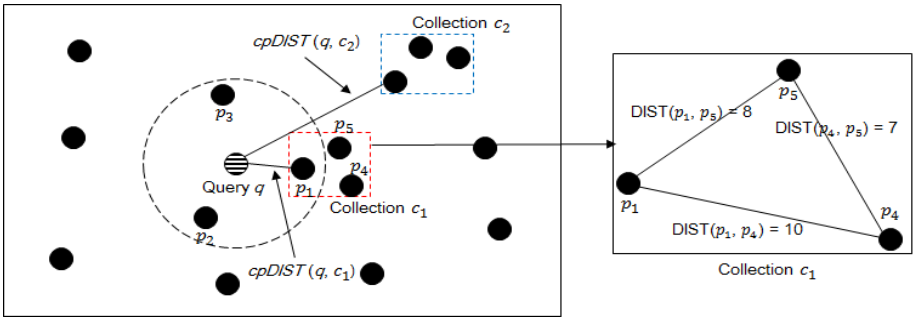


Fig. 1. Example of k-NN and NC searches

Example 1 (Location-based services): In a LBS region with a large set of points (such as restaurants or clothing stores etc.), tourists want to visit places that are located within limited distance to have a variety of experiences at a time. In this case, a NC search would return the nearest set of spatial points with constraints.

Example 2 (Decision supporting systems): Consider a person whose decision is to open a new pharmacy in a region. Given several choices (i.e. sites) to meet condition (e.g. rental cost) for its new location, the person want to choose the place which has a cluster of k hospitals at a short distance. In this case, NC searches can support the decision of site selection.

To the best of our knowledge, nearest collection problem have not been studied in the literature. The brute-force search algorithms of NC problem are obviously too expensive to be of any practical use. Hence, we need to design new algorithms to process the NC search more efficiently by taking the new database perspectives into account.

The rest of this paper is organized as follows. Section 2 reviews work related to our research. We formulate the problem of nearest collection in Section 3. We propose the algorithms for processing NC queries using spatial indexes (i.e. R-tree and aR-tree)

in Section 4. Section 5 describes the implementation and presents the results of experiments. In Section 6, we conclude our work and present some direction for future research and acknowledgments.

2 Related Work

R-tree [3] and its variants are the most widely deployed indexing structure for spatial databases. A leaf node of R-trees contains a data entry referring to a spatial point. The data points close to each other are grouped into minimum bounding rectangles (MBRs). An internal node contains a MBR which includes all MBRs associated with the child entries. The aggregate R-tree (aR-tree) [2] [4] stores, in each intermediate entry, pre-aggregated sums of the objects in the sub-tree.

In previous works, to process a nearest neighbor query on spatial databases, R-tree should be traversed in the depth first or the best first fashion. The depth first search [8] starts from the root of the tree and visits the child whose minimum distance from the query point is the smallest. If it reaches a leaf node, it retrieves the candidate of NN and traces back. The best first search [9] maintains a min-heap for entries being visited. The heap elements are sorted in the increasing order of their minimum distance from query point. In the search, the entry with the minimal distance from query point is visited first. In this paper, these two tree traversal manners are used in the proposed search algorithms.

3 Problem Formulation

For NC search problem formulation, we give formal definition of collection and the nearest collection search problem..

Definition 1 (*Collection*): Given a set of spatial objects (i.e. points) O, and a user-specified k and collection threshold δ, any collection c of O is a set composed of k objects which have the sum of Euclidean distance between objects in c is less than or equal to δ. It is formally defined as follows:

$$c(\delta) = \{o \mid o \in O, \sum_{1 \le i < k} \sum_{i < j \le k} |o_i - o_j| \le \delta\} \quad (1)$$

Definition 2 (*cpDIST*): Given a collection c including $\{o_1, o_2, \ldots, o_k\}$ and a point p, the *cpDIST* is the Euclidean distance between c and p. It is formally defined as follows:

$$cpDIST(c, p) = min_{1 \le i \le k}\{|o_i - p|\} \quad (2)$$

Definition 3 (*Nearest Collection Search*): Given a set of collections C and a query point q, the *nearest collection search* of q finds the collection c with the smallest *cpDIST*(c, q) in C.

4 Nearest Collection Search

Given a data points set D and an integer k, the naïve implementation of nearest collection search requires computing the number of combinations of selecting k elements from $|D|$, easily leading to a complexity of $O(\binom{|D|}{k})$. For a large D, such computation becomes costly.

In this section, to reduce the processing costs, we present two approaches for processing nearest collection queries using R-tree and aR-tree. First, we will give a simple algorithm for finding NC of a given query point q, using existing nearest neighbor method. Then, we will present an efficient query processing method using spatial pruning.

4.1 Multiple Query Approach

The basic idea of multiple query approach (MQA) is to perform incremental k-nearest neighbor (k-NN) queries which use the depth first traversal and combine their results. This MQA algorithm is shown in Algorithm 1.

Algorithm 1 Multiple Query NC Search
Input: a R-tree T, a query q, an integer $k \geq 2$, a threshold value δ
Output: a set $R \subset D$

```
1.  R ← {};   i ← 1;   λ ← i * k;   create a min-heap H
2.  while (R is empty) do
3.      delete every elements in H;
4.      C ← {};   C ← k-NN (T, q, λ);
5.      for each (S in k-combination of C) do
6.          insert S into H; // S is a subset of k-distinct elements of C.
7.      while (H is not empty) do
8.          remove top subset S from H;
9.          if (the sum of distances between all pairs of elements in S ≤ δ) then
10.             R ← S; // S is the nearest collection satisfying the given condition
11.             break;
12.         else // S does not satisfy the given condition
13.             continue;
14.     if (every objects in R were retrieved for k-NN ( T, q, λ )) then break;
15.     i ← i + 1;   λ ← i * k;
16. return R;
```

To avoid confusion, we will use the notation λ-NN instead of k-NN. At the i-th step, λ-NN with $\lambda = i * k$ is invoked and the result is stored in C (line 4). Then every subset S in k-combination of C is inserted into the min-heap H (line 5-6). Note that the root of H is a subset S' which has the minimal *cpDIST*. Subset S removed from H is checked if it satisfies a given condition, until there is no subset in H. If some S satisfies a given condition, then S is the nearest collection (line 10-11). If there is no subset satisfying a given condition, then i is increased, and $(i + 1)$-th step begins.

4.2 Branch and Bound Approach

MQA may incur multiple accesses to the same page while a NC query processing is performed. To avoid this problem, the database method (i.e. spatial pruning) is explored to boost the nearest collection query processing on aR-tree. We introduce the branch and bound algorithm which computes the nearest collection of a query point q by expanding the entries of the min-heap H according to their distance from q.

Lemma 1: Given a MBR m including data points P, aggregated value i of m and user-specified k, the length of diagonal line of $m * \binom{k}{2}$ is more than the sum of Euclidean distance between points in P.

The Proof of Lemma 1 is simple because the maximum distance between two points in m is the length of diagonal line of m. Accordingly, diagonal line of m provides an upper bound ub on sum aggregate distances between k objects in m. Therefore, if $ub \leq \delta$ and $i \geq$ k where δ is a collection threshold, then m includes one or more collections. By Lemma 1, we can get a candidate of NC result, and prune another nodes with it. The pruning rule is defined as follows:

$$minDIST(q, N) \geq maxDIST(q, m),$$ where q is a query point, N is a node, m is a candidate of NC result.

If a node N satisfies above pruning rule, then N can be pruned.

Algorithm 2 Branch and Bound NC Search
Input: an aR-tree T, a query q, an integer k ≥ 2, a threshold value δ
Output: a set $R \subset D$

```
1.  R, C ← { }; τ ← + ∞;     insert all entries of the root of T into the min-heap H;
2.  while (heap H is not empty) do
3.      remove top entry e from H;
4.      if (the minimum distance from q to e  ≥ τ) then discard e; // pruning
5.      if (aggregated value of e  ≥ k) then
6.          if ((length of diagonal line of MBR of e * (k 2)) ≤ δ ) then
7.              τ ← the maximum distance from q to e;
8.          if (e is a leaf node including data points) then
9.              insert all data points in e into C;
10.             if (a collection c exists in C) then
11.                 τ ← the maximum distance from q to c;
12.         else // e is an internal node
13.             for each (child entry e_i of e) do
14.                 insert e_i into heap H;
15.     else
16.         add e to C;
17. run Incremental_Refinement(C) and compute R;
18. return R;
```

Branch and bound algorithm is shown in Algorithm 2. This algorithm retrieves the qualified NC objects by traversing the aR-tree index T in a best first manner. Specifically, we maintain a minimum heap H with entries. First we insert root of aR-tree into H (line 1). Each time we pop out an entry from H (line 3), and check whether or not the entry can be pruned (line 4). If the entry is not pruned, it traverses aR-tree.

5 Experimental Results

In this section, we perform an experimental evaluation to prove the efficiency of the proposed methods. All the following experiments are running on a PC with Intel Core2 Quad Q8200 2.33GHz, 4GB memory.

We have used both synthetic and real datasets in the experiments. For the synthetic datasets, we generated the datasets which have clustered distribution (i.e., Gaussian distribution) with the Spatial Data Generator [10]. All data objects are located in the space of size 1K x 1K. For the real datasets, we also used the MBRs of census blocks of Iowa, Kansas, Missouri and Nebraska based on Tiger/Line [11]. This real dataset is highly skewed. For our algorithms, we measured their performance using two metrics, namely the search space reduction ratio and the number of I/Os. Table 1 shows the parameters and the default values in our experiments.

Table 1. Defualt Experiment Setting

Parameters	Default values
Size of a Collection (i.e. k)	4
Collection Threshold	250
Distribution of Data	Gaussian
Distribution of Queries	Uniform
# of Data	40,000 objects

Fig. 2 illustrates the search space reduction (SSR) ratio of MQA and BBA search. Let $|D|$ be the cardinality of data points D and $|C|$ be the size of candidates of NC result, then the search space reduction ratio is defined as:

$$SSR = \frac{|C|}{|D|} \times 100 \tag{3}$$

From Fig. 2, we can see that our two algorithms (based on R-tree and aR-tree) have very small search space (i.e. less than 0.5%), compared with full search space of the linear scan method (i.e. the brute-force approach). As the size of datasets increase, the search spaces decrease for both Gaussian distribution (MQA-G, BBA-G) and real (MQA-R, BBA-R) datasets. When the size of dataset increases, the density of objects increase too. Consequently, incremental k-NN queries are invoked less, so the MQA has smaller search spaces than BBA.

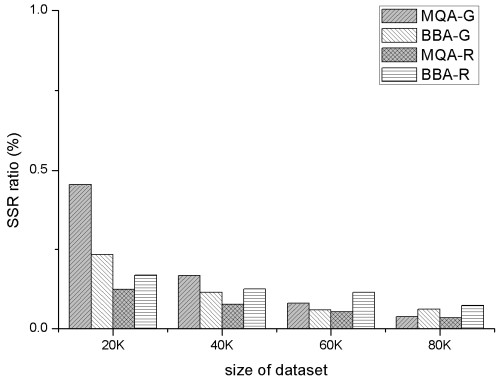


Fig. 2. Search space reduction ratio of MQA and BBA

Fig. 3 illustrates the comparison of I/O costs between MQA and BBA search. Clearly, BBA algorithm shows much less I/O costs than MQA algorithm for all datasets. This is because, BBA uses pruning method based on a candidate collection. However, MQA search uses incremental algorithm for retrieving nearest collection, which results in multiple access to the visited nodes.

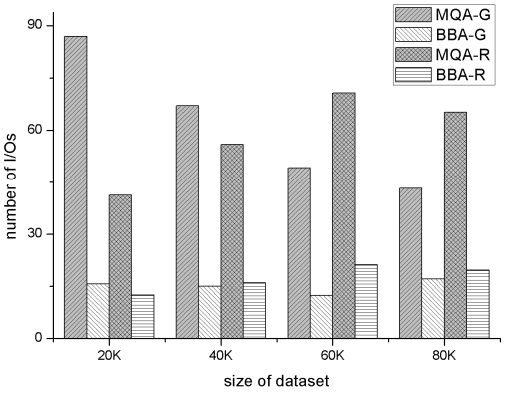


Fig. 3. Number of I/Os of MQA and BBA

6 Conclusion

This paper studies the nearest collection search that has many real life applications. We introduce multiple query approach which processes incremental k-nearest neighbor queries to find a nearest collection. To efficiently process the query, we have proposed branch and bound search algorithms for solving nearest collection search problem based on aggregate R-tree. We have conducted thorough experiments

on both synthetic, and real datasets. As a result, both search algorithm showed effectiveness in reducing search spaces (while naïve approach can cause full access to datasets). According to our experiments, the branch and bound algorithm were able to prune more search spaces than incremental k-nearest neighbor algorithm. Consequently, the I/O efficiency of branch and bound search algorithm is better than multiple query approach.

In future, we aim to develop more efficient algorithms for the NC search, and extend our work for moving object databases.

Acknowledgement. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2013-R1A1A2010616).

References

1. Aji, A.: High performance spatial query processing for large scale scientific data. In: SIGMOD PhD Symposium (2012)
2. Lazaridis, I., Mehrotra, S.: Progressive approximate aggregate queries with a multiresolution tree structure. In: Proc. ACM SIGMOD (2001)
3. Guttman, A.: R-trees: a dynamic index structure for spatial searching. In: Proc. ACM SIGMOD (1984)
4. Papadias, D., Kalnis, P., Zhang, J., Tao, Y.: Efficient OLAP operations in spatial data warehouse. In: SSTD (2001)
5. Roussopoulos, N., Kelly, S., Vincent, F.: Nearest neighbor queries. In: Proc. ACM SIGMOD (1995)
6. Korn, F., Muthukrishnan, S.: Influence sets based on reverse nearest neighbor queries. In: Proc. ACM SIGMOD (2000)
7. Papadias, D., Shen, Q., Tao, Y., Mouratidis, K.: Group nearest neighbor queries. Proc. ICDE (2004)
8. Cheung, K., Fu, A.W.C.: Enhanced nearest neighbor search on the R-tree. ACM SIGMOD Record 27(3), 16–21 (1998)
9. Hjaltason, G., Samet, H.: Distance browsing in spatial databases. ACM Trans. Database Systems 24(2), 265–318 (1999)
10. Spatial Data Generator by Yannis Theodoridis, `http://www.rtreeportal.org/software/SpatialDataGenerator.zip`
11. U.S. Cencus Bureau. Tiger/Line Shapefiles, `http://www.census.gov/geo/www/tiger/shp.html`

Application of Environment Constraints in Improving Localization Accuracy

Dinh-Van Nguyen, Eric Castelli, Trung-Kien Dao,
Duc-Tho Le, Lan-Huong Nguyen, and Salim Attig

Dept. of Pervasive Spaces and Interaction, MICA Institute (HUST-CNRS/UMI 2954-INP Grenoble), Hanoi University of Science and Technology, Hanoi, Vietnam
{dinh-van.nguyen,eric.castelli,trung-kien.dao,
duc-tho.le,lan-huong.nguyen}@mica.edu.vn,
attigs@e.ujf-grenoble.fr

Abstract. In this paper, a technique to improve indoor localization results on the basis of environment information is introduced and validated. A generic model of environment allows data transforming and speeding up modelling process is proposed, consisting in building two components: a description of environment data using XML and a set of automatic modules for converting XML into other types of database. In this study, the environment information is used in improving the accuracy of a WiFi-based indoor localization system by applying constraints.

Keywords: environment modelling, pervasive environment, indoor localization, WiFi-based localization, environment constraints.

1 Introduction

Previous research and development for indoor localization include cellular networks, infrared, ultrasonic, computer vision, RFID, etc. [1-3] However, these technologies suffer either from the limited accuracy, range, a lack in the infrastructure, or high deployment price. Due to the widespread availability of local wireless networks, the interest in localization based on WiFi signal strengths are increasing because no additional infrastructural costs beyond the existing WiFi infrastructure is required, making it become a promising indoor localization scheme. Much research has been devoted to methods for the WiFi-based localization [4]. In this study, the use of environment information in improving the accuracy of a localization system based on WiFi is introduced.

During past decades, many methods are proposed in order to model environments which can be used for automatic processing using computer. Depending on the scale of the environment and the specific problems, techniques of modelling are also diverged such as: modular modelling system (MMS) [5], open modelling engine (OME) [6], UbikSim [7], etc. One of the most successful techniques is using Geographic Information System (GIS) [8] which provides both data management, analysis and visualization tools in a single package. However, these complex tools

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_56,

require a high level of understanding the underlying concepts as well as programming skills.

Having these issues stated, a tool for modelling environment that is at the same time accurate, easy to use, extensible and reusable for improving localization services is developed in this study. To achieve these requirements, we approach the modelling problem from an object oriented and modular view embedded in XML databases to assure easy-to-use features, flexibility and meaningfulness. Furthermore, the modelling efforts are reduced by including predefined objects from other sources into the environment.

2 Environment Modelling and Management

In this study, the environment model is used in multiple goals: providing data for localization, optimal path finding, and visualizing the results in 3D. These tasks are different in storing and using database but need description of the same environment. Thus, a generic environment model is required, which can easily be derived to the other databases. Besides, the generic model must not lose data semantic as well as data correctness. Therefore, a structured XML model must be carefully designed to preserve these information and generic characteristic.

Since the main target of an environment model is to describe human interaction with environment, each entity in environment is managed and has its characteristics defined. By modelling the environment in object-oriented databases, we can manage to describe interactions between objects, human and other factors such as time, conditions. At the same time, defining all objects in details from scratch takes a lot of time. Thus, the model should only have generic objects which can be redefined later according to real situations. The challenge here is how to add such features to an XML document. One way to solve this problem is to define all imported objects as external XML documents. All these objects will be linked together to form the final model structure through an automatic merging module. Also, inside these generic modules, the process can be repeated in a same manner to define sub-objects as components. This recursive structure will make it possible for easily modifying parts of the model. Instances of object will be mapped to tags in the main XML structure to create a final detailed XML description of the environment.

3 Indoor WiFi-Based Localization and Improvement Using Environment Constraints

3.1 Indoor Localization Using WiFi

With conventional receivers, distance to the access points (APs) can only be estimated from the measured RSSI (received signal strength index) with help of a RF (radio-frequency) propagation model constructed based on the fact that a radio wave traveling through a certain environment will undergo specific types of signal attenuation. To start off, the empirical model widely used in previous works [10] is considered:

$$P = P_0 - 10n\log(r/r_0), \tag{1}$$

where P_0 is the known signal power at a reference distance r_0; P, the signal power at an unknown distance r; n, the path-loss exponent. Once the parameters P_0, r_0 and n are determined from experiments, the distance can be estimated given the RSSI.

Equation (1) is a propagation model in an environment without obstacle between the AP and the receiver. When walls and floors are considered in calculation, attenuation due to these factors must be included, and the propagation equation becomes

$$P = P_0 - 10n\log(r/r_0) - k_d \sum_{i=1}^{n_w} d_i/\cos\beta_i, \tag{2}$$

where n_w is the number of walls and floors in the middle of the AP and the receiver, d_i is the thickness of the i^{th} wall/floor, β_i is the angle of arrival corresponding to the i^{th} wall/floor, and k_d is the attenuation factor per wall/floor thickness unit.

Equation (2) is a deterministic model, i.e., the uncertainty of RSSI at a distance is not taken into account. To address this limitation, a probabilistic model is used. In reality, given the RSSI P, the distance r might not be exactly the value calculated from Eq. (2), but is within a range around this value, which is denoted by $\bar{r}$. To be more precise, $\bar{r}$ will be the nominate value of the distance r with highest probability. Given a RSSI P, the distribution of the distance is assumed to follow the normal (or Gaussian) distribution with median $\bar{r}$:

$$\rho(r,P) = \Pr(r|P) = \left(1/\sigma\sqrt{2\pi}\right)\exp\left(-(r-\bar{r})^2/2\sigma^2\right), \tag{3}$$

where σ is the standard deviation, which is also a function of P. For simplicity, σ is assumed to be related to $\bar{r}$ by a linear relation $\sigma = k_\sigma \bar{r}$, where k_σ is a constant.

To localize a user in environment, a hybrid approach using WiFi RSSI is used. From a tuple of RSSI information received from the user's smart device, the system needs to determine the user location. With the probabilistic propagation model, the user localization method can be established. The main idea is to find the location with maximal summation of probabilities corresponding to the visible APs, i.e., maximizing $\rho_\Sigma(x,y,z) = \sum_{i=1}^{n_{AP}} \rho(r_i(x,y,z), P_i)$, where n_{AP} is the number of visible APs, $\rho_\Sigma(x,y,z)$ is the probability that the user is located at position (x,y,z), and $\rho(r_i(x,y,z), P_i)$ is the probability component based on the i^{th} visible AP.

The search process can be achieved by gridifying the space surrounding the environment into a number of points and calculate ρ_Σ for each of them to find the point that maximizes ρ_Σ, which will be the estimated user position. The wall and floor information involved in these calculations is extracted (using XQuery) from the environment map. With a large environment where the number of points is big, it is possible to reduce the searching time by first gridifying the space with fewer points to find a rough position, then repeating the same process once or twice with the subspace around this position for fine tuning.

The user location is determined by the user's smart device. As the user moves in the environment, the app installed on his smart device regularly scans the WiFi signals from neighbour APs and send the collected RSSI to the central server through the WLAN. Based on the RSSI information, the server will estimate the user location by using the proposed method. The localization result consists of a 3D position with (x, y, z) coordinates, together with the floor and the zone that the user belongs to.

3.2 Indoor WiFi Localization Enhancement Using Environment Constraints

In this section, we will focus on improving localization results using contextual data (data which we can derive from environment characteristic or relationship between objects in environments). There are three cases for improving localization using environment physical and contextual constrains: (1) the altitude correction, (2) the invalid area of localization, and (3) the area which is unreachable for user.

The first case is where, due to precision of calculation, the returned localization result has "invalid altitude" for user. It is trivial that user should be right on the floor for each position (except for lifts and stairs). Thus "invalid altitude" implies the case where returned altitude is in the middle of two floors. To correct this, the location brought to the closest floor.

The second case is where the returned user location is in an invalid area such as inside walls or objects. In this case, the location is brought back to the valid location closest to the previous result. A map of valid areas is built by considering walls, objects as 2D polygon and using polygon Boolean operations.

The last case is where the previous user location is taken into account to determine whether it is possible for user to reach new location in the given time. To decide whether it is possible for user to reach new location, a visibility graph is built using line-sweeping algorithm. Using this visibility graph, the shortest path between two locations can be determined. Given the max speed of user and time difference between two localization results, it is possible to check if the user can reach the new location or not.

4 Experiment Results

An environment consists of a 10-floor building equipped with smart-devices is targeted as a pervasive environment. Specifically, top 3 floors (see Fig. 1a) of the building are modeled in detail with 14 WiFi APs and the localization system is deployed on these floors to evaluate the effect of environment constraints on localization results. In the evaluation process, a user uses his/her smartphone and makes a round trip with solid pathway demonstrated in Fig. 1b. The user is tracked in real time, firstly using WiFi-based localization algorithm alone (case 1) then secondly using the same algorithm with improvement using environment constraints (case 2).

There are totally 86 measurements of WiFi signals recorded for evaluation. In Fig. 2a, the distribution of localization error is shown in two cases: with and without using environment constraints, while Fig. 2b represents the reliability of the

localization result as a function of the localization error. The results are summarized in Table 1. It can be seen that, with environment constraints in used, the localization results are improved significantly.

Table 1. Localization results without and with environment constraints

	Average error (m)	Maximal error (m)	Error at reliability of 90% (m)
Without constraints	6.81	28.94	12.47
With constraints	2.47	6.65	4.67

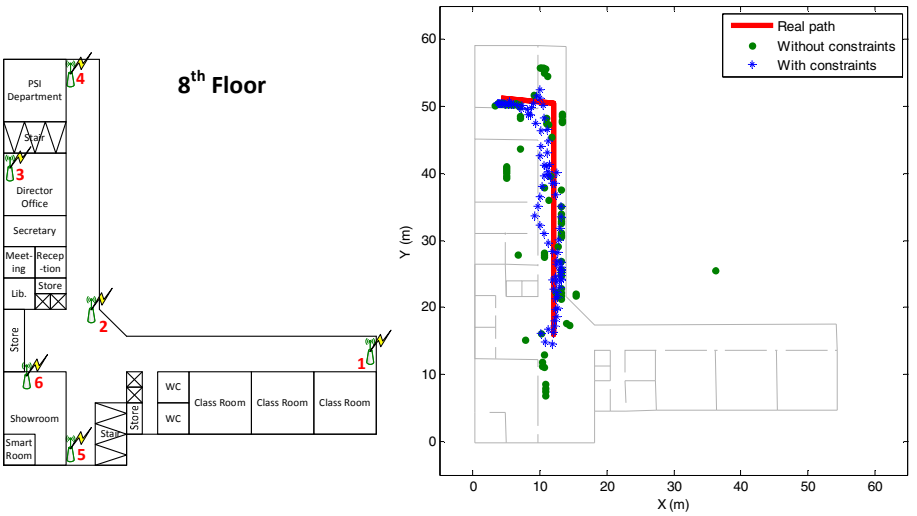


Fig. 1. (a) Testing environment with WiFi APs, and (b) localization results

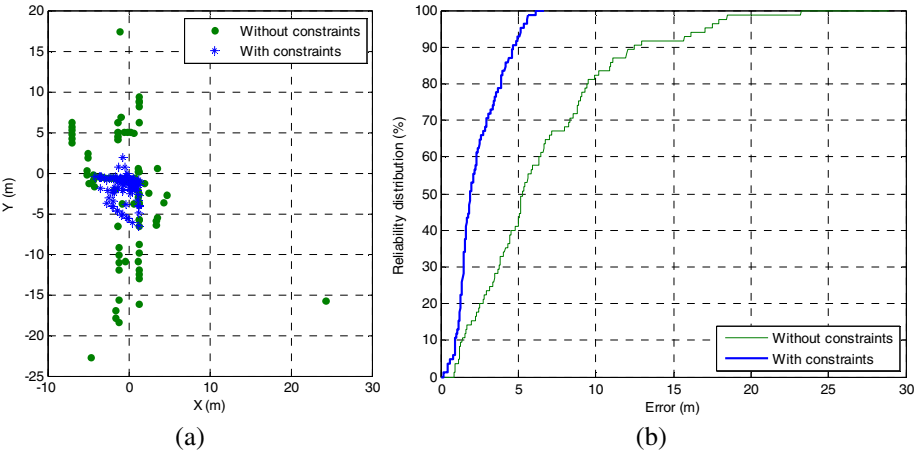


Fig. 2. (a) Distribution of localization error, and (b) localization reliability

5 Conclusion

In this paper, a method to model environments for pervasive systems, and its application in improving localization accuracy, is introduced. The method takes advantages of XML format such as ease of use and user-friendly interface to make a generic form of data. From these XML description documents of environments, several automatic modules for conversion of data types such as SQL, 3D model, are built. The method not only significantly reduced the modelling process efforts but also improved localization results using contextual data. Experimental results from localization service show that the database helped to increase the accuracy of localization since not only the program has more detailed and accurate data of environment but also the ability to adapt to environment changes.

Acknowledgment. The authors of this paper would like to thank the Ministry of Education and Training and the VLIR's Own Initiatives Program for their financial supports of the research projects under grant references B2011-01-052 and VLIR-UOS ZEIN2012RIP19.

References

1. Hakan, K., Shuang, H.Y.: A survey of indoor positioning and object locating systems. International Journal of Computer Science and Network Security 10(5) (2010)
2. Liu, H., Darabi, H., Banerjee, P., Liu, J.: Survey of wireless indoor positioning techniques and systems. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 37(6), 1067–1080 (2007)
3. Medina, A.V., Gómez, J.A., Ribeiro, J.A., Dorronzoro, E.: Indoor position system based on a zigbee network. In: Chessa, S., Knauth, S. (eds.) EvAAL 2012. CCIS, vol. 362, pp. 6–16. Springer, Heidelberg (2013)
4. Munoz, D., Lara, F.B., Vargas, C., Enriquez-Caldera, R.: Position Location Techniques and Applications. Academic Press (2009) ISBN: 978-0123743534
5. Leavesley, G.H., Markstrom, S.L., Viger, R.J., Hay, L.E.: The Modular Modelling System MMS. In: Int. Conference on Integrating Geographic Information Systems and Environmental Modelling, Breckenridge (1993)
6. Reed, M., Cuddy, S.M., Rizzoli, A.E.: A Framework for Modelling Multiple Resource Management Issues - an Open Modelling Approach. In: McDonald, A.D., McAleer, M. (eds.) MODSIM 1997 - International Congress on Modelling and Simulation, vol. 2, pp. 681–686. Modelling and Simulation Society of Australia Inc., Australia (1997)
7. Campuzano, F., Garcia-Valverde, T., Garcia-Sola, A., Botia, J.A.: Flexible Simulation of Ubiquitous Computing Environments. In: Novais, P., Preuveneers, D., Corchado, J.M. (eds.) ISAmI 2011. AISC, vol. 92, pp. 189–196. Springer, Heidelberg (2011)
8. Abel, D.J., Kilby, P.J., Davis, P.R.: The system integration problem. International Journal of Geographic Information System, 1R–12R (1994)

9. Vaskeviciusa, N., Birkb, A., Pathakc, K., Schwertfegerd, S.: Efficient Representation in Three-Dimensional Environment Modelling for Planetary Robotic Exploration. Advanced Robotic Special Issue: Section Focused on Advanced Space Robotics 24(8-9) (2010)
10. Figueiras, J., Frattasi, S.: Mobile Positioning and Tracking: From Conventional to Cooperative Techniques. John Wiley & Sons (2010) ISBN: 978-0470694510

Inferring Probability of Guessing from Item Response Data Using Bayes' Theorem

Byoung Wook Kim, Ja Mee Kim, and Won Gyu Lee*

Dept. of Computer Science Education, Korea University, Seoul, Korea
{byoungwook.kim,jamee.kim,lee}@inc.korea.ac.kr

Abstract. Outlier detection is a primary step in many data mining applications. Outlier means a marked response data correctly by guessing in item response data. Guessing an answer or a judgment about something without being sure of all the facts act as a noise in data mining. It is important to clean noise data for producing good results in data mining. In order to clean noise data, it is needed to detect correct answers marked by guessing among item response data. In this paper, we present a Bayesian approach to infer a probability of guessing for items.

Keywords: Item Guessing, Bayes' Theorem, Item Response Data.

1 Introduction

Using a multiple-choice response format on achievement tests is convenient for many reasons. However, a disadvantage of a multiple-choice format in contrast to a free-response format is the psychological phenomenon of guessing: It is likely that an examinee will choose any of the given response options at random if examinee has no idea which of them is the correct one, given that refusing to respond does not seem to be a fair option [1], [2]. As a consequence, a multiple-choice response format is troublesome because an item is often scored as solved even though the examinee may not have the ability level to solve that item [3].

In this case, although an examinee does not have the ability to solve questions, they are classified as a group with the capacity by guessing. Such item response data can be said that the outline. In many data analysis tasks a large number of variables are being recorded or sampled. One of the first steps towards obtaining a coherent analysis is the detection of outlaying observations [4]. It is needed to detect outliers which are response data correctly by guessing for data mining [5]. For example, if examinees seriously do not response and mark the answers for items, response data of the examinees will interfere with the correct analysis. Therefore, measurement method of guessing is needed for analyzing data correctly.

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_57,

Traditionally, item difficulty, item discrimination and item guessing are calculated by analyzing item response data with Item Response Theory (IRT) in the field of education [6]. However, few studies have been conducted to find out a probability of guessing for items (examinee guessing) from item response data. In this paper, we present a Bayesian approach to infer a probability of guessing for items.

In this paper, we adopted Bayes' Theorem for inferring probability of guessing for items. Bayesian statistics is a subset of the field of statistics in which the evidence about the true state of the world is expressed in terms of degrees of belief or, more specifically, Bayesian probabilities [7]. Such an interpretation is only one of a number of interpretations of probability and there are many other statistical techniques that are not based on degrees of belief. Bayes' Theorem derives the posterior probability as a consequence of two antecedents, a prior probability and a likelihood function derived from a probability model for the data to be observed. Bayes' Theorem computes the posterior probability with total score of examinees.

2 Preliminary

2.1 Bayesian Probability

Bayesian probability is to measure of state of knowledge in the probability theory. It is a different interpretation that the frequency of probability occurrence or physical attributes of system.

Bayesian probability is discussed in two point of aspect, one of them is that Bayesian statistical law is to prove the reasonable and universal [8]. This can be explained as an expansion of the logic. On the other hand, in aspect of Subjectivist Theory of Probability, the knowledge state can be measure using degree of belief. Most machine learning methods were made on the basis of the principle of Bayesian theorem. Bayesian probability is most popular ones in probability theorems in theoretical psychology, sociology and economics [9, 10]. In order to evaluate the probability of hypotheses, to clarify the prior probabilities, and to change the new probability values by the new observed data [10]. Bayesian theorem is presented the necessary step and standard of formula in calculation of probabilities [11].

Bayesian theorem is shows the relationship of the prior probability and posterior probability. In aspect of Bayesian probability, it is explained how to be updated posterior probability when presented new evidence. The posterior probability is proportional to product of prior probability and likelihood.

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)} \tag{1}$$

H is hypothesis and *D* is observed data. *P(H)* is prior probability of *H*. namely, *H* is probability of true until observed *D*. *P(D|H)* is likelihood of conditional probability that is true with observed *D*. *P(D)* is prior probability of *D*. *P(H|D)* is posterior probability. This is probability that hypothesis is true given data and hypothesis of prior belief.

2.2 Item Guessing

From true-false items or multiple items, it is possible that one of examinee is make a guess right answer. If do not provide punish the false answer, it is possible to occur in the future inspection. Accordingly, item guessing is important of element in item analysis.

For item guessing, first, we take a number of guess who answered examinee and correct answers among them. If effect of test is higher than examinee will be not answer that make a guess right answer. Therefore, we estimate a number of guessed examinees and estimate a number of correct answers among guessed examinees using probability theorem. When number of guessed answer is G, number of right answer among guessed is as follows:

$$G_r = \frac{G}{Q} \tag{2}$$

where *Gr* is the number of examinee who guessed right answer, *G* is the number of guessed examinee and *Q* is the number of correct answer.

Guessed examinees don't know a correct answer. It is assumed that examinee is equality answered by the test. Therefore, correct answer in guessed answer is *G*/*Q*. For example, 60 examinees in 100 examinees are answered 4-multiple choice items guessing, the number of right answerer is 15. In contrast, number of wrong answerer is as follow:

$$G_W = G \times \frac{Q-1}{Q} = W \tag{3}$$

where G_w and *W* are the number of guessed wrong answerer and *Q* is the number of answer.

It is not possible to know the number of guessed answerer. We don't know the number of guessed right answerer. However, we know the number of wrong answerer. Therefore, we can estimate the number of guessed answerer using number of wrong answer examinee. Wrong answerers are guessed answer because it is equal to wrong answerer and guessed wrong answerer. Then we get equation as follow:

$$G = \frac{WQ}{Q-1} \tag{4}$$

In equation 4, a numerator is product number of wrong answers (*W*) and the number of answer (*Q*), and a denominator is the number of answer (*Q*) minus 1. For example, 45 examinee in 100 examinee are answered 4-multiple choice items guessing, number of guessed wrong answerer is 60 (45*4/3=60). The number of guessed right answerer is compute using equation 1:

$$G_R = \frac{G}{Q} = \frac{WQ}{Q-1} \times \frac{1}{Q} = \frac{W}{Q-1} \tag{5}$$

For equation 4, 45 is the number of wrong answer. So 15 (45/4-1) is the number of correct answer with guessing. Item guessing is the number of guessed right answerer in total answerer:

$$P(C_G) = \frac{G_R}{M} \tag{6}$$

where G_R is the number of correct answer examinees with guessing and M is the number of total examinee.

A notation in result of equation 6, average of right answerer is proportional to average of wrong answers. In Fig. 1, probability of a correct answer, $P(C)$, is 0.55(=55/100), probability of guessed correct answer, $P(C_G)$, is 0.15 (=15/100), and given know the correct answer then probability of guessed right answer, $P(C_G|C)$, is = .27.

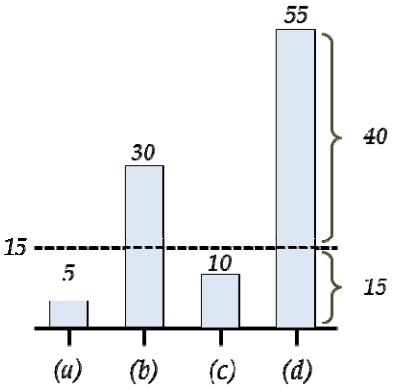


Fig. 1. A calculation of item guessing in classical test theory

3 Experiment

3.1 Data Collection

This study used item response data obtained from the Information and Communication Technology (ICT) literacy tests (test scores of 12,531 elementary students) as training data. This examination is comprised of 36 multiple-choice items. In general, item response data shown in Table 1 can be obtained as outcome generated through a test. In the item response data, each transaction (i.e., student) has its own score and is considered differently, and '1' means a correct answer while '0' means wrong. One row has one variable which is the total score of each student. A wide variety of data mining techniques utilize this item response data as raw data to extract useful information. Research that has analyzed item response data are as follows.

Table 1. The structure of item response data

Student ID	I_1	I_2	…	I_n	Total Score
$\boldsymbol{S_1}$	0	1	…	1	$\boldsymbol{TS_1}$
$\boldsymbol{S_2}$	1	1	…	1	$\boldsymbol{TS_2}$
…	…	…	…	…	…
$\boldsymbol{S_m}$	0	1	…	1	$\boldsymbol{TS_m}$

In order to support the validity of our approach, we generated item response data marked randomly based on Item Response Theory. It is difficulty to collect data that real correct answer with guessing. Therefore, we simulated item response data randomly. We can know correct answer with guessing in simulation data. In this paper, we generated item response data of 100 examinees as test data. In order to investigate validity of results, real item response data of 100 examinees were used as control data. Finally, we compared the correct answer with guess between test data and control data.

3.2 Calculation of Posterior Probability

In this paper, we adopted that the similar ability groups will response to same items as prior distribution. Experiment procedure is follows:

1. Calculating the total score of one examinee in simulation data.
2. Grouping same score examinees of training data with the total score of the examinee in simulation data.
3. Calculating a conditional probability of correct answer with guessing under total score of the examinee in simulation data.

3.3 Results

In this section, we present a prior probability and post probability of correct answer with guessing and a comparison of correct answer with guessing between test data and control data using post probability.

Table 2 presents an example of prior probability and post probability of correct answer with guessing for examinees got 10 scores. $P(C_G)$ is a prior probability of correct answers without knowing total score of examinees. $P(C_G|C)$ is a post probability of correct answers with knowing total score of examinees. As shown in the table 2, a post probability is high than prior probability to all items. That is, the information about total score of examinees provides to support a post probability. If total score is high than 10 scores then the difference of $P(C_G)$ and $P(C_G|C)$ is reduced. This means that guessing probability of higher ability examinees is lower than lower ability examinees.

Table 2. An example of prior probability and post probability of correct answer with guessing for examinees got 10 score

Item	$P(C_G)$	$P(C_G\|C)$	Item	$P(C_G)$	$P(C_G\|C)$
1	0.1229	0.1858	19	0.1289	0.2165
2	0.0762	0.1370	20	0.1365	0.1849
3	0.0760	0.1245	21	0.1036	0.1954
4	0.0987	0.1820	22	0.0595	0.1753
5	0.1420	0.1858	23	0.0576	0.1648
6	0.0895	0.1533	24	0.1398	0.1906
7	0.0468	0.1121	25	0.0556	0.1753
8	0.0813	0.1782	26	0.1295	0.1887
9	0.0798	0.1657	27	0.1050	0.1983
10	0.0786	0.1628	28	0.0592	0.1734
11	0.1912	0.2261	29	0.0773	0.1973
12	0.1022	0.1734	30	0.0496	0.1609
13	0.0931	0.1772	31	0.1961	0.2270
14	0.0805	0.1849	32	0.1010	0.1992
15	0.0495	0.1686	33	0.1065	0.1897
16	0.0832	0.2002	34	0.1121	0.1944
17	0.1476	0.2136	35	0.1529	0.1954
18	0.0557	0.1456	36	0.1296	0.1964

Table 3. An examples of correct answer with guessing between test data and control data using post probability for examinees got 10 score

Item	Test data	Control data	Item	Test data	Control data
1	0.0557	0.0557	19	0.1515	0.0216
2	0.0411	0.0411	20	0.0555	0.0555
3	0.0374	0.0125	21	0.0391	0.0391
4	0.0182	0.0546	22	0.0351	0.0351
5	0.0186	0.0372	23	0.0330	0.0494
6	0.0153	0.0613	24	0.1144	0.0572
7	0.0336	0.0672	25	0.0876	0.0876
8	0.0713	0.0356	26	0.0377	0.0377
9	0.0166	0.0497	27	0.0793	0.0595
10	0.1140	0.0489	28	0.0347	0.0867
11	0.0452	0.0226	29	0.0197	0.0197
12	0.0693	0.0867	30	0.0322	0.0483
13	0.0177	0.0532	31	0.0908	0.0227
14	0.0739	0.0185	32	0.0598	0.0797
15	0.0169	0.0674	33	0.0000	0.0759
16	0.0601	0.0400	34	0.0583	0.0194
17	0.0854	0.0214	35	0.0586	0.0391
18	0.0291	0.0728	36	0.0196	0.0589

Table 3 presents an example of correct answer with guessing between test data and control data using post probability for examinees got 10 score.

In test data, the correct answer is selected randomly. In control data, the correct answer is marked by real examinees. Therefore, there is no relation among the correct answer in test data. As shown in the table 3, correct answer with guessing in test data is high than in control data. The sum of correct answer with guessing is 1.8263 in test data and 1.7395 in control data. The sum of test data is high than control data. In view of the results, post probability using Bayes' Theorem have an effect of inferring a probability of guessing.

4 Conclusion

In this paper, we present a Bayesian approach to infer a probability of guessing for items. Prior probability of guessing was updated by total score of examinees. This post probability has an effect to distinguish between correct answers with guessing and true correct answers. The result of experiment shows our approach is available for measuring a degree of guessing of examinees.

Education data mining is an emerging discipline that focuses on applying computer science techniques to educationally related data in a ubiquitous learning environment. Therefore, it becomes more important that gathering information of learner's features or academic achievements for providing proper personal learning material to learners. However, meaningless raw data by guessing randomly obstruct producing successful results. This method can be used in data cleaning stage to detect answers marked by guessing in data mining.

Acknowledgements. This study was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2013R1A2A2A03016926).

References

1. Espinosa, M.P., Gardeazabal, J.: Optimal correction for guessing in multiple-choice tests. Journal of Mathematical Psychology 54(5), 415–425 (2010)
2. Baradaran, A., Ahanghari, S., Semiari, S.R.: The Impact of Correction for Guessing Formula on MC and Yes/No Vocabulary Tests' Scores. The Journal of Applied Linguistics 2(2(5), 80–98 (2009)
3. Kubinger, K.D., Holocher-Ertl, S., Reif, M., Hohensinn, C., Frebort, M.: On Minimizing Guessing Effects on Multiple-Choice Items: Superiority of a two solutions and three distractors item format to a one solution and five distractors item format. International Journal of Selection and Assessment 18(1), 111–115 (2010)
4. Romero, C., Ventura, S.: Educational Data Mining: A Review of the State of the Art. IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews 40(6), 601–618 (2010)

5. David, J.H., Barbara, S.M.: Anomaly detection in streaming environmental sensor data: A data-driven modeling approach. Environmental Modelling & Software 25(9), 1014–1022 (2010)
6. Lord, F.M.: Formula scoring and number-right scoring. Journal of Educational Measurement 12, 7–12 (1975)
7. Wong, M.L., Lee, S.Y., Leung, K.S.: Data mining of Bayesian networks using cooperative coevolution. Decision Support Systems 38(3), 451–472 (2004)
8. Cooper, G.F., Herskovits, E.: A Bayesian method for the induction of probabilistic networks from data. Machine Learning 9, 309–347 (1992)
9. Earman's, J.: Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory. The MIT Press, Cambridge (1992)
10. Hwang, K.-S., Cho, S.-B.: A Bayesian inference model for landmarks detection on mobile devices. Journal of Korea Information Science Society: Computing Practive 13(1), 35–45 (2007)
11. Korb, K.B., Nicholson, A.E.: Bayesian Artificial Intelligence. Chapman & Hall/CRC (2003)

Implementation of Water Pollution Response Information Systems Based on IP-USN

H.S. Shim, G.Y. Min, and D.H. Jeoung

Korea Institute of Science & Technology Information,
Department of MIS, Dongguk University
{hsshim,duke}@dgu.edu, william1540@nver.com

Abstract. This study is implementation of water pollution response information systems (WPRS) based on Internet Protocol-Ubiquitous Sensor Network (IP-USN). We can perform real time monitoring and management of discharging facilities, enabling prevention of water pollution incidents. WPRS also effectively establishes integrated.

1 Introduction

Water pollution incident in South Korea, and about 50~60 cases a reported annually, and, Four Rivers Restoration Project[1] covers four of the country's five major river basins, as they face many water-related problems including floods, droughts, water quality issues and growing demand for recreational water facilities. But increased risk of water pollution. We must do this to prevent the environmental disaster from getting worse.

This study of the water pollution control and water pollution control information for the information sharing through the system to respond effectively to utilize building was studied.

WPRS aims to supervise discharging facilities in real time with remote monitoring instead of on-site monitoring; and to efficiently and systematically manage the water quality of effluent from discharging facilities through improved methods of assessment. WPRS is intended to effectively establish integrated watershed management systems.

2 Water Pollution Prevention Systems

Water pollution prevention systems is composed of water TMS(Tele-Management system), national water quality automatic monitoring system, and water pollution measurement system based on IP-USN, as shown in Figure 1.

This system is developing location-based information services, item/river water pollution, influencing factors such as meteorological data and water pollution linked

[1] Four Rivers Restoration Project explained by the claimed objectives of securing water resources and preventing floods.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_58,

to statistical analysis. DB server for managing the data collected and WAS, WEB servers, DBMS, and GIS system was constructed. And all the servers in a separate task of unifying the server were configured as a server.

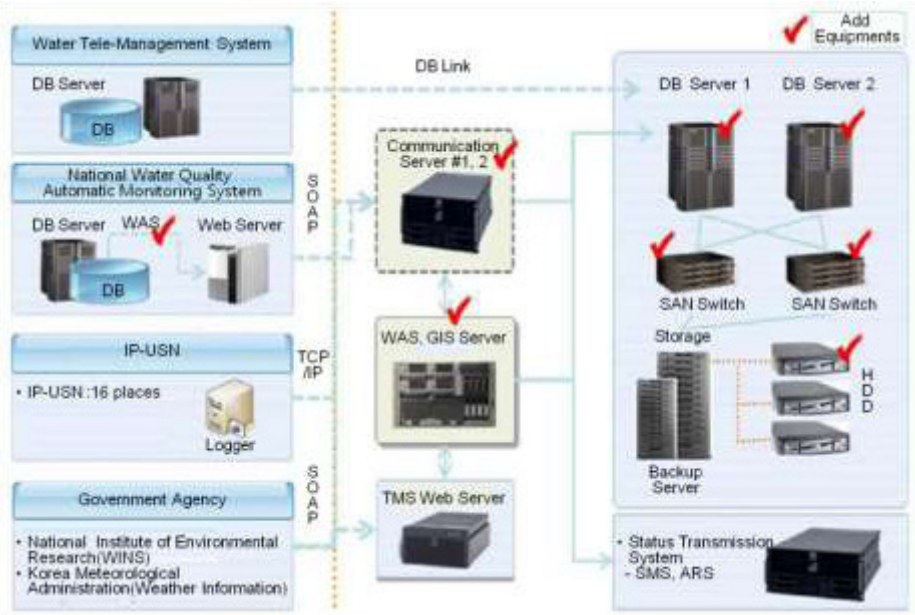


Fig. 1. Water Pollution Response Information System H/W Architecture

3 Implementation of WPRS

3.1 Information As-Is Analysis

First, coverage analysis was analyzed the law and work manuals, than consisting of 16 functions and 49 tasks were defined business functions model, and evaluated the coverage. Second, user evaluated by functional quality satisfaction perspective, system requirement perspective, task function and Component technology utilization perspective. Third, infrastructure analyzed CPU use rate, network use rate.

3.2 WPRS To-Be Model

WPRS's values, capacities and services collaboration to achieve the vision in terms of evolution, the development must be consistent with the long-term direction of development.

WPRS is analyzing the data in the required information were linked as shown in Figure 2. (1)Water pollution incident management process analyzed for SOA-based Software, (2) Applied standard API for agencies interoperability, and (3) IT infrastructure to support business processes to maximize performance virtualization system is operated.

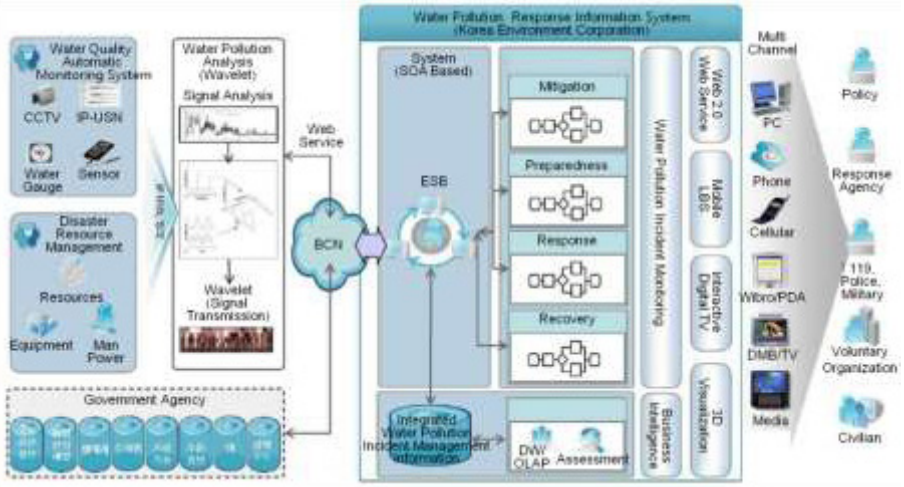


Fig. 2. Concept of Water Pollution Response Information System

(1) WPRS Linked Water Pollution Incident Management Process

WPRS linked water pollution incident management process, also Water pollution incident management process based on the BPM to call in a service request is sent to legacy systems and institutions. Service Repository is logical and physical storage and management of the service, as shown in Figure 3.

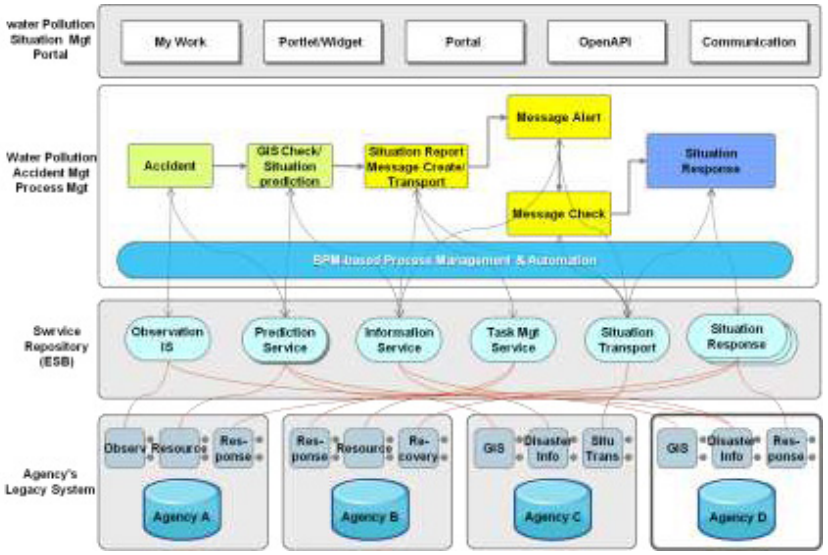


Fig. 3. Water Pollution Accident Management Process based on SOA

In response to the various accident types and circumstances Dynamic response to a combination of services can be provided.

(2) Water Pollution Monitoring System Establish Information Sharing

WPRS Applied standard API for agencies interoperability. And conjunction with the relevant agencies' water pollution information for water quality monitoring system was expanded, as shown in Figure 4. Disaster information is automatically collected in conjunction with the other agencies have implemented of system.

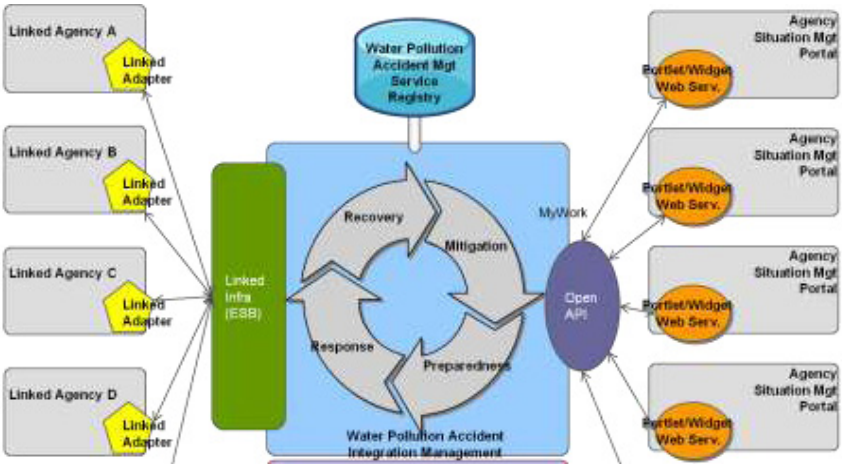


Fig. 4. Water Pollution Information sharing

(3) Server Virtualization Design

IT infrastructure to support business processes in order to maximize the performance of virtual machine operating room to perform enterprise-wide workload management, automated resource allocation (Auto-Provisioning) shall apply, as shown in Figure 5.

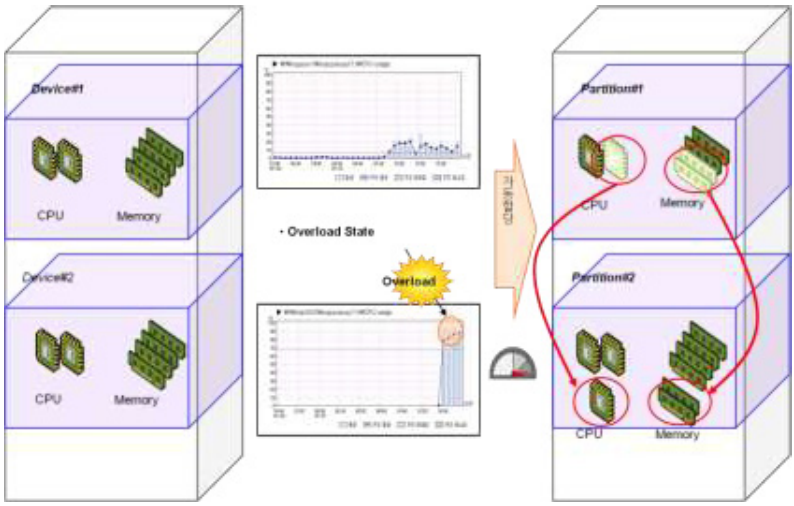


Fig. 5. Server Virtualization

4 Conclusion

Effectively prevention and response can .information sharing through the Water pollution response systems. Than WPRS aims to supervise real time with remote monitoring instead of on-site monitoring; and to efficiently and systematically manage the water quality of effluent from discharging facilities through improved methods of assessment. WPRS is intended to effectively establish integrated watershed management systems.

Acknowledgement. This research was supported by a grant from information strategy plan project(2012), Korea Environment Cooperation, Korea.

References

1. Korea Gov. The Four Rivers Restoration Project (2010)
2. International RiverFoundation. South Korea's Four Rivers Restoration Project (2012)
3. Hankyoreh, The environmental fallout of the Four Major Rivers Project (Special Report) (2013)
4. Korea Environment Institute, Korea Environmental Policy Bulletin, KEPB Issue 4, vol. VIII (2010)
5. Choi, Y.W., Sung, D.G., Jeon, H.S., Cho, G.S.: A Study on the Development and Application of GIS-based Stream Water Quality Management System. GIS Review 10(2), 185–370 (2007)
6. Jang, H.S., Cho, J.C., Kim, D.S., Kim, I.T.: Quality Management system for National Water Quality Monitoring System. Korean Society of water and Waster, 541–542 (2010)
7. Korean Environment Corporation. Water Pollution incident Response Information System ISP (2011)
8. Korean Environment Corporation, Pollution incident Response Manual (2010)
9. National Institute of Environmental Research, Water Pollution incident Response System Training (2010)
10. Korean Environment Corporation, `http://waterkorea.or.kr/index.do`
11. National Institute of Environmental Research, `http://www.nier.go.kr/eric/portal/kor`

Implementation of M2M System for a Racing Car

Min-seop Song*, Seong-Hyun Beak, and Jong-wook Jang

995 Eomgwangno, Busan jin-gu, Busan, 614-714, Computer Engineering,
Dong Eui University, Korea
{seobejj,jwjang}@deu.ac.kr, smartsma@naver.com

Abstract. As mobile network services have been widely used these days, the development technology of WCDMA or LTE and their applications are gradually expanded, and thus there is a trend that a lot of IT fusion industries are emerging. In this paper, M2M system was developed that utilize the OBD-II data and transfer several sensors data to an external server for other external devices to be able to confirm system of a racing cars. A M2M system for a racing car was implemented that reads information from the a variety of sensors inside the vehicle using the OBD-II connector, converts the data for users to see easily, and transfers the data to external data server using a mobile networks module. To conduct efficiency test, an ordinary vehicle and a racing car driving at high-speed on the actual circuit were used for the performance test of the developed system; generated data were transmitted through the OBD-II connector; it was confirmed that data were received without error and loss by the M2M system for racing car. In addition, it was also confirmed that the same data, as were transmitted to an external server using mobile networks, were sent and received normally.

Keywords: M2M, WCDMA, LTE, Vehicle Network.

1 Introduction

In recent years, with increasing interest for auto racing domestically, racing events are held on a regular basis for amateur drivers rather than professional racers[1].

IT sector is responsible to send the data of several sensors to a TCP server for the racing car driver and maintenance team; the transmitted data are stores and analyzed for use to identify driving habits and the failure or the main parts or sensors inside the vehicle.

There are many car racing types such as Formula, Kart, Touring Car, Stock Car, etc.; particularly Touring Car racing is similar to a general passenger vehicle because it basically uses a race car made based on the mass production vehicle.

Such Touring Car Racing reduces the weight of the vehicle by a considerable amount. Bye replacing the external steel body with lighter plastic body. And using a few essential sensors after removing most unimportant sensors.

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_59,

Because all domestic passenger cars sold from January 2005 have been obliged to work together with OBD-II connector. This system is available in most vehicles equipped with OBD-II home and abroad[2].

Implemented in this study will be available in many areas that combined a vehicle and IT.

2 Related Research

2.1 OBD-II

On-Board Diagnostics, or OBD is a term used in the automotive industry. Sensors are mounted in recently produced cars for various instrumentation and control; these devices are controlled by ECU (Electronic Control Unit)

ECU was originally developed for the purpose of precise control of core functionality of the engine such as ignition timing and fuel injection, variable valve timing, idling, and threshold settings. It now control every part of the vehicle systems including automatic transmission, driving system, breaking system and steering system, with the development of the performance vehicle and computer.

These electronic diagnostic systems have continued to evolve, and was recently settled to a standard diagnostic system called OBD-II (On-Board Diagnostic version II).

All cars adopt standardized DTC (Diagnostic Trouble Codes) and connection interfaces (ISO-J1962) according to OBD-II standard. But five different electronic signals exist the historical background; such signal system commissioned developers to a large burden.

In order to resolve these incompatibility issues, all cars were required to use the ISO 15765-4 standard, which are being sold from 2008 in the U.S. market, the world's largest automobile market[3].

If the currently used standard ISO-J1962 connector and an external scanner are connected, communication is possible with ECU using the scanning software installed in PC or PDA using the OBD-II standard.

If a fault occurs in the car, OBD-II indicates the details of the error through a 5-digit diagnosis code.

Fault types and fault code are also standardized; general auto repair shops easily detect the problem to repair it using the OBD-0II standard fault codes.

3 System Design and Key Features

3.1 Structure of M2M System for a Racing Car

This system that is implemented in this paper is composed of a transmitter-receiver to communicate with the OBD-II network inside the car; a transmitter-receiver to transfer data to an external server; a transmitter-receiver to communicate with GPS;

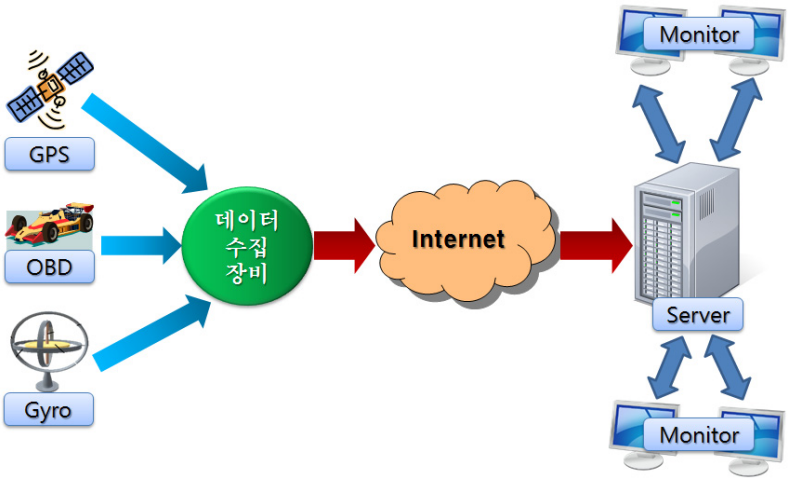


Fig. 1. Structure of M2M System for a racing car

a transmitter-receiver to communicate with Gyro sensor; and a MCU to integrate and control the data from all the modules.

Figure 1 shows the process that: as soon as the information- such as vehicle information from OBD-II network while driving; and time, position and vehicle tilt information from GPS- arrive at Cortex-M3, they immediately are transferred to an external server and stored.

3.2 Design of M2M System for a Racing Car

In the system implemented in this study, the key feature is to synchronize the interior information of OBD-II, and the information of Gyro sensor and GPS.

The communication between each module and the Cortex-M3 was implemented with UART communication, which is largely classified into two types of Polling and Interrupt, among which Interrupt type was used in the synchronization.

Using Interrupt type, it controls RXNE pins, receives and integrates data from each sensor, and transfer data to the server via WCDMA.

Figure 2 shows the flowchart of M2M System for a racing car. With the start of the system, connection is set in the WCDMA for the data linkage with the server. The setting of connection is continued until the interlocking succeeds with the server.

On successful connection with the server, data are received from OBD-II, GPS and Gyro sensor. If the data is not correctly received from each module or errors occurred, it requests again for the data.

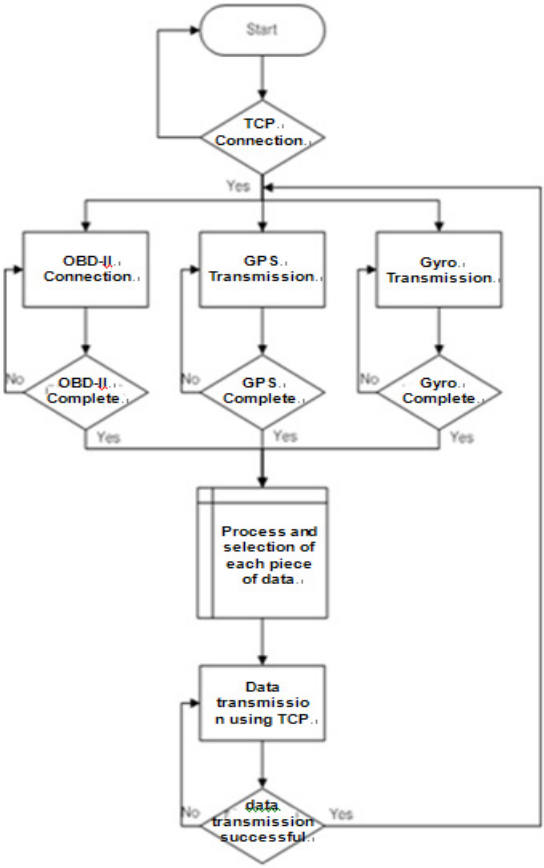


Fig. 2. Flowchart of the M2M System for a racing car

First of all, six pieces of information are received from OBD-II for data processing. Table 1 below shows the six information from OBD-II.

Table 1. PID of OBD-II[5]

Mode	Pid	Returned Data bytes	Description
01	0D	1	Vehicle Speed
	0C	2	Engine RPM
	5C	1	Engine oil temperature
	05	1	Engine coolant temperature
	11	1	throttle position
	2F	1	Fuel level input

Secondly, regarding the data from GPS, time and position data are received for processing from $GPRMC among standard protocol, NMEA-0813.

Lastly, among the data from Gyro sensor, the ROLL, PITCH and YAW data of the earth. The above data received from the three sensors are sent to the external server through WCDMA module.

4 Implementation of M2M System for a Racing Car

In this paper, the system development environments are: MCU of STM32 that uses Cortex-M3 core; GPS that uses MTK3329 chips; 9-axis G-sensor; and WCDMA USIM chip using KT Communication Company.

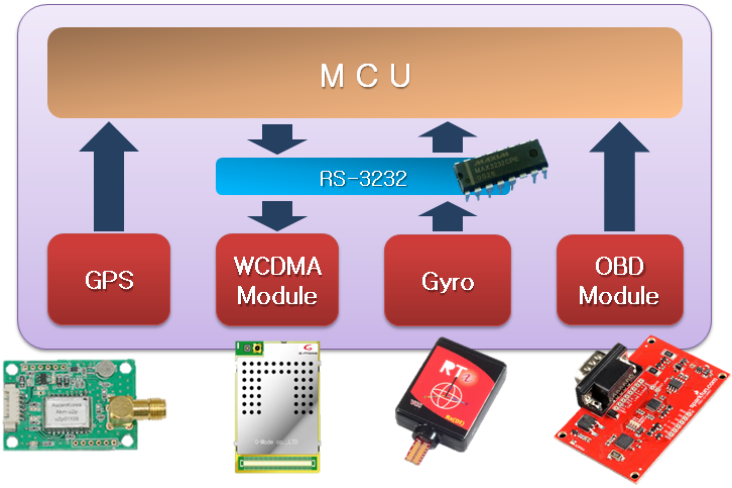


Fig. 3. Hardware diagram of M2M System for a racing car

Figure 3 shows the hardware diagram of the M2M System for a racing car.

The integrated vehicle diagnostic recorder system consists of 4 major areas: OBD-II

transmitter-receiver module; GPS
transmitter-receiver module; Gyro sensor
transmitter-receiver module; and WCDMA
transmitter-receiver module.

Thus configured hardware is integrated in the firmware on the STM32 board.

Fig. 4. Server screen and test of M2M System for a racing car

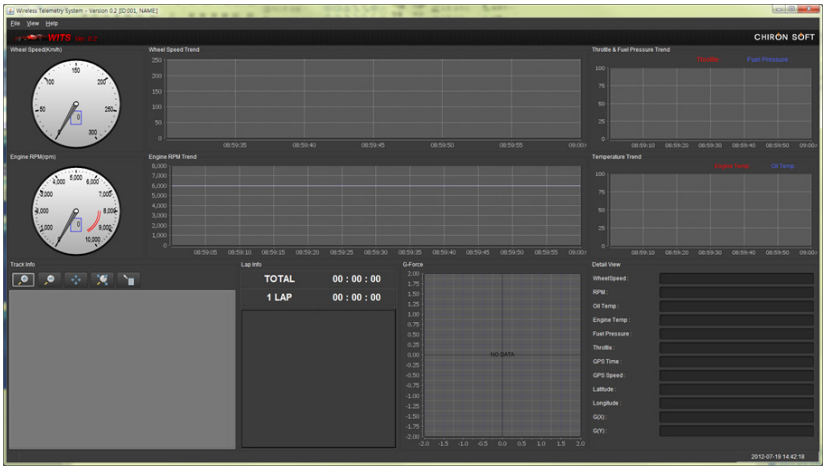


Fig. 5. Server program screen of M2M System for a racing car

Figure 5 shows an implemented screen of M2M System for a racing car, which can be monitored from the server. The top circular shape on the left displays the speed information of the vehicle received from the GPS data; the bottom one displays RPM information received from the OBD-II. They take the form of the instrument panel of the actual vehicle for the users to be able to identify easily.

The left side of the lower part is a map to check where vehicle is actually running using the information from the GPS; the rectangular portion of the middle part is the implementation of the instrument panel to read the gage panel by the time on the left side. The Y-axis of the rectangle represents the vehicle speed; the X-axis shows time when the corresponding speed has reached. The bottom of the rectangular part was

also implemented to make it easier to identify the RPM part of the instrument panel at different times.

The square in the lower part displays the points out of the data from the G-sensor, enabling to identify the shaking and tipping of the vehicle at a glance.

The upper-right corner shows the Throttle Position indicating how hard the driver stepped on the gas pedal. X-axis displays a value from 0 to 100 in percent as the unit.

The middle-right part presents the engine temperature showing the temperature of the engine of the vehicle in real time.

The bottom right displays the information in a concise format that is sent for each sensor, changing the value of each sensor in real time.

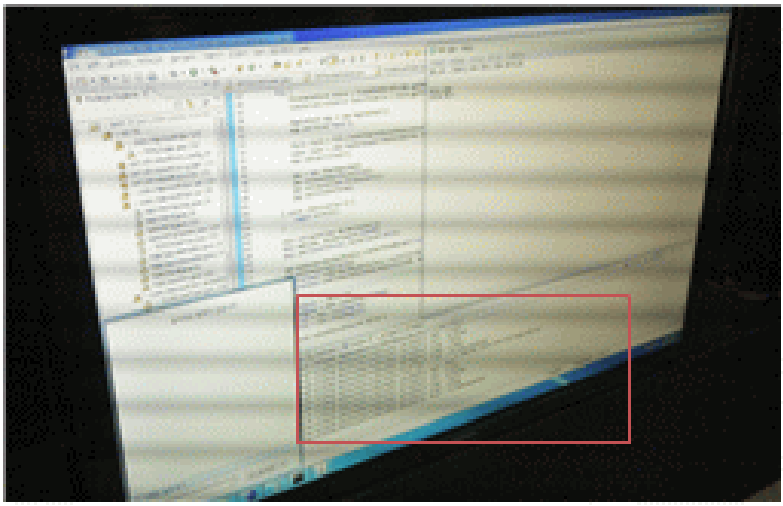


Fig. 6. Logging of M2M System for a racing car

Figure 6 shows the log records of the racing vehicle state monitoring system. It can be verified that the information generated when the server side runs a system is being recorded in the corresponding system. Thus, the remaining logs enable the user to find the status of the transfer of vehicle information, accurately determine the cause of the problem of the system, and identify vehicle information, by separately saving the logging records.

5 Conclusion

In this paper, a system was implemented that transfers accurate internal information inside of the racing vehicle and the location of the vehicle to an external server with WCDMA using the information inside of the vehicle, a GPS and a gyro sensor. This system may be employed in a car EDR system to trace back the process and causes of accidents, or to shorten the driving record by identifying the driving behavior of the racer.

In addition, because the information about the interior of the vehicle is sent to an external server, the data may be stored safe in case the system is lost or damaged.

Further study will be performed in the future to implement a system, which transfers all video information and vehicle information through LTE module after collecting and processing the entire video information and vehicular information.

The system will also be embedded for actual mounting on a vehicle, which will implement a more perfect racing vehicle state monitoring system.

Acknowledgments. This work was supported by the Brain Busan 21 Project in 2013.

References

1. Shin, E.S., Im, B.H.: Development of a Low-Cost Racing Car. Journal of Industrial Science Technology and Institute 20(2), 19–25 (2006)
2. Choi, D., Hong, D., Hong, S.: Embedded Real-Time Software Architectures for Automotive Systems. In: Korean Society of Automotive Engineers 2005 Symposium (Electrical and Electronics, ITS Division), pp. 43–50 (2005)
3. Kim, K., Kim, H., Lee, J.: Study on Vehicle Stability Enhancement for the Pace Formula Vehicle. Transactions of KSAE 19(1), 25–31 (2011)
4. OBD-II PIDs, Wikipedia, http://en.wikipedia.org/wiki/OBD-II_PIDs
5. Leadtek Research Inc., GPS Protocol Reference Manual, p. 6

A Study on Sound Reproduction for Adaptive Mixed-Reality Space

Ho-Jin Lee, Ji-Woong Park, Soon Il Kwon*, Jong-Weon Lee, and Sung-Wook Baik

Dept. of Digital Contents, Sejong University, Seoul, Korea
{woogime,pjy8933}@naver.com, skwon@sejong.edu,
{jwlee,sbaik}@sejong.ac.kr

Abstract. Advertisements and campaigns on multiscreen with digital monitors are recently becoming very common in public places. However current virtual reality technology has limits in the feeling of physical space which is an important factor for feeling of real space. We present a study on adaptive audio reproduction for physical and spatial immersion with mixed reality technology. Especially we tried to adjust the distortion of audio signal spectrum that can occur at users dynamic locations. As a result, we need the differentiated adjustment audio signal in frequency as well as loudness according to a distance between a user and a speaker.

Keywords: Mixed Reality, Audio Sweet Spot, Spectrum Distortion.

1 Introduction

Nowadays, ads and campaigns on multiscreen are becoming very common through LCD and LED TV in a public place. But current virtual reality technology has limits in terms of a feeling of actual space of a content that describes feeling of real space. To overcome this problem, some smart TVs were used to make people feel the sense of space by installing multi-display contents on left and right side in the window of the train. They showed a scene that a plane came flying from the left window display in the train and flew into the right window display, thereby showing the scene, to match real space within users and spatial point of the landscape outside the window.(Fig. 1)

A study on new technology blending the actual physical space and virtual reality space has been processed first to maximize the feeling of actual space in Korea. This paper presents a study on adaptive audio reproduction optimization for physical and spatial immersion enlargement. Especially when user moved in the mixed reality environment, we studied to minimize the audio signal distortion from fixed or moved speakers to users. In the experiments, we adjusted the original sound beforehand by considering attenuation for each frequency band. Then we could confirm possibility of recreation that was same with original sound at user's location.

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_60,

Fig. 1. Samsung Electronics smart TV ads (CNN)

2 Interactive Architecture Based Adaptive Mixed-Reality Space

Our study based on Kinetic Architecture by considering aesthetic part and functional part introducing dynamic design concept is to make new adaptive mixed reality experience. This means that we are going to make new experience space which is physical space synchronized with user interaction and virtual reality contents. As a beginning try called "Moving Labyrinth" introduced interactive architecture concept, interior walls of labyrinth can rotate a full 360 degrees, and in can make Kinetic Architecture inner part from various motions and locations of walls programmed in advance.(Figure 2)

Fig. 2. Examples of interactive architecture

In this study, adaptive mixed reality experience space, VIP(Virtual reality, Interaction, Physical) space, designed based on interactive architecture is new conceptual experience space through interacting virtual reality contents to feel actual spatial immersion. The VIP space included interest factors such as virtual reality game contents and artistic factors in architecture is new conceptual experience space maximized studying effect of experience mixed reality contents and provides space typology fit the situation by changing dynamically smart media walls included display that express immersive contents, surround speakers, 3D cameras, etc. Users can experience the contents by absorbing naturally through interaction with virtual contents in this space. Figure 3 shows the example of mixed reality experience space. Smart media walls can control transfer

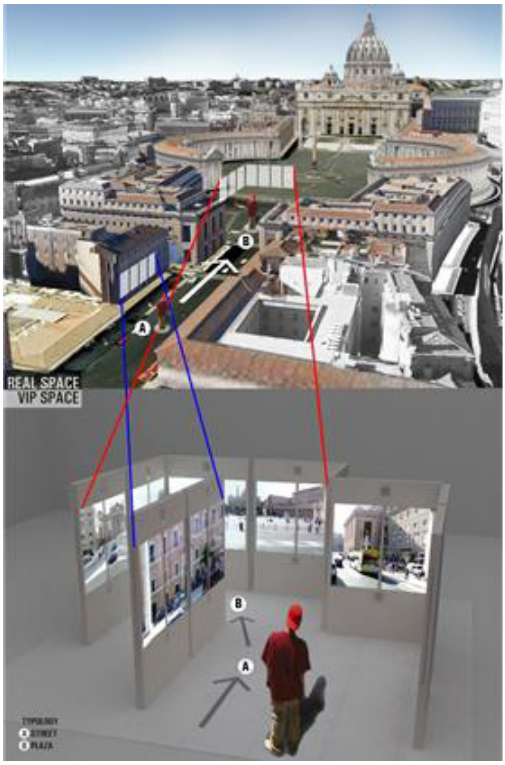


Fig. 3. An illustration of adaptive mixed reality experience space

process that is from Street(A) to Plaza(B) to feel the real space in VIP space dynamically. We can experience physical space similar with real space and representation of immersive virtual contents through smart media walls. In other words, this is a scenario that is similar to real world.

3 Sound Reproduction in Adaptive Mixed Reality Space

To provide Space Typology for suitable situation of contents by adaptive mixed reality experience space, smart media walls having inner speakers are moving dynamically. Basically when a user locates at special spot of VIP space, VIP controller controls multiple speakers for optimal stereo by considering the relative location between each speaker and the user. When the controller controls multiple speakers, it makes optimal audio sweet spot by considering the intensity of acoustic signal and time delay based on the relative position. But the location of smart media walls can change according to spatial type, we have to continuous resetting of the spatial correlation between a user and multiple speakers. To do upon processing, we have to relocate the speakers used for audio sweet spot and

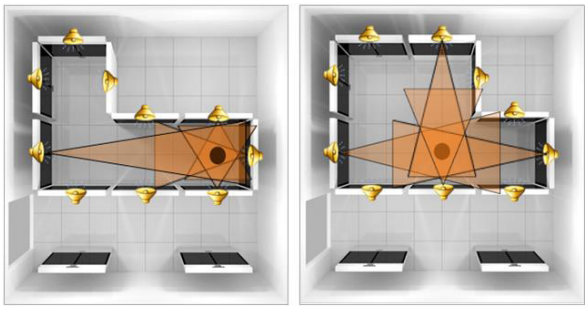


Fig. 4. An example of space typology and audio sweet spot according to user's location and speaker reallocation

adjust the acoustic signal to form audio sweet spot at allocated location.(Figure 4)[1][2][3]

We can get the results for sound reproduction by adjusting the volume inversely proportional to the user's location and each speaker in smart media walls. However we have to consider the fact that the decaying sine wave decreased according to inverse square law that every twice the distance reduced by 6dB.[4]

Generally when a user moves, VIP controller has to adjust attenuation levels until sine wave attenuation reaches a certain distance to reproduce the surround sound what we want. Equation (1) is the formular that to calculate sound pressure level in the half free-field(directional coefficent = 2). Equation (2) is the attenuation formular caused by air absorption in weather conditions.(ignoring the wind, the temperature is 20)

$$SPL = PWL - 10log(2\pi r^2), [Q = 2] = PWL - 20logr - 8dB \tag{1}$$

$$Aa = 7.4 * (\frac{f^2 * r}{\phi}) * 10^{-8}(dB) \tag{2}$$

In Equation (1), PWL is sound power level. Q is the directional coefficient of the free space not half free space on the ground. Therefore it is set to 2. In Equation (2), f is center frequency(Hz) in octave band, r is the distance(m) between the source and view point, ϕ is the relative humidity(%). We found out that the higher the frequency the lower the humidity, the attenuation of sound pressure levels is increased. Also the lower the temperature, the attenuation value is increased too.

In this paper we measured and analyzed the tendency of attenuation decaying sine waves in multi-channel speaker environment. Then we tried to adjust the attenuation levels of frequency components.[1][5][6] In case of attenuation levels are different in frequency, even though overall volume is adjusted, it cannot reproduce original sound at audio sweet spot since frequency distribution of the original sound is not preserved. Hence we need to compensate the degree of

attenuation to reproduce the original sound precisely at audio sweet spot by reproducing the frequency distribution of the original sound as it is.

4 Experiment and Result

We experimented at 7 X 10m general reverberation laboratory, and average environmental noise level of laboratory was measured around 30dB. We set microphone and speaker at 0.9m height from bottom of laboratory and, used 1kHz, 4kHz, 8kHz, 16kHz sine waves.

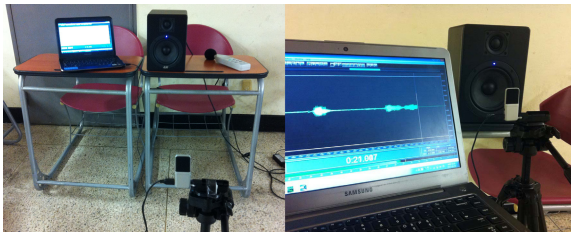


Fig. 5. Experiment environment and Instruments

In experience space for adaptive mixed reality, if we try to use point sound source attenuation equation that was used in previous studies in anechoic chamber, we have to separate the direct sound and the reflected sound from experimental sound.[7] However, in real world there is rarely anechoic conditions, and due to limitations of the experimental environment, it is impossible for us to record the direct sound precisely from experimental sound. Hence we recorded the experimental sound, then we analyzed the experimental wave form at the start time of recording, and measured the distance between speaker and microphone and walls to separate the direct sound and reflected sound from experimental sound by using the speed of sound at room temperature. In this manner we could get only direct sound from experimental sound. We recorded sine waves of a single frequency (1kHz, 4kHz, 8kHz, 16kHz) at 1m, 2m, 3m from speaker, and using the analysis manner, we could check attenuation differences according to frequency by distance at specific distances, and get attenuation levels(dB) of decaying sine waves according to frequency. Based on this, we adjusted the attenuation differences of decaying sine waves according to frequency, then we recorded and analyzed differences of sound pressure levels according to frequency.

In <a> graph, we measured sound levels of 1kHz, 4kHz, 8kHz, 16kHz experimental sounds at 0.15m, 1m, 2m, 3m from speaker. In <a>, Original is the standard sound levels to survey attenuation differences at each frequencies. So we recorded original sound 0.15m away form speaker. We can check attenuation differences according to frequency by distance.

In <b> graph, we adjusted the differences that obtained from <a> at each frequencies to original sound beforehand, then likewise the previous experiment,

we recorded the adjusted sound at 1m, 2m, 3m and analyzed the result using frequency filtering. In <b> all graphs are similar in appearance, this means that we can reduce sound distortion that caused by the difference in frequency depending on the distance. However, in <b>, the whole graphs not even, and the graphs are decrease as high frequency section. That is because of output degradation in the high frequency section.

As a result of the experiment, we could check that there are differences of attenuation of sound with distance according to frequency. However, except the reflected sound, as the attenuation equation due to air absorption(2), frequency is not proportional to distance. We supposed that the result can be explained because of the delay when played the experimental sound at start, so we couldn't separate the direct sound from experimental sound or because of output differences according to frequencies in speaker. In this phenomenon, to solve the

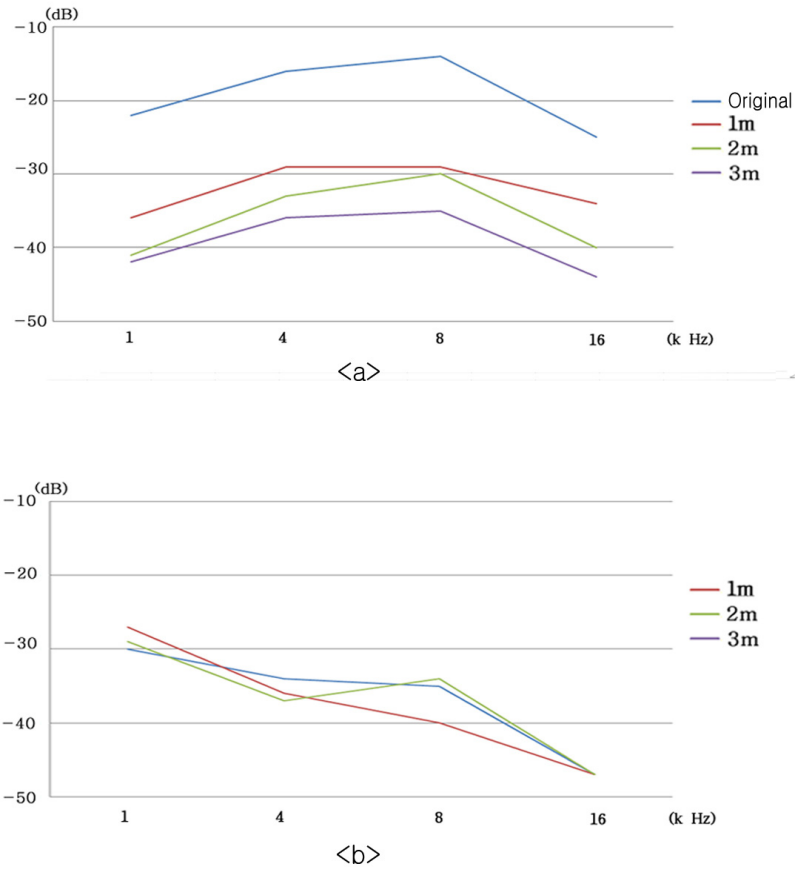


Fig. 6. Result: <a> Sound pressure levels of specific frequencies with distances. <b> Sound pressure levels of specific frequencies with distances after adjust.

problem, we are going to find out the speaker output differences in specific frequencies, and adjust the differences. We are going to study the manner of separating the direct sound from experimental sound precisely considering speaker feature. In addition, we are going to execute additional experiments to find out that differences of attenuation tendency according to frequency as we experiment with 1k, 4k, 8k 16k Hz single frequency sine waves independently and as we experiment with one complex sound of specific frequencies.

5 Conclusion

In this study, our final goal is to find the sound attenuation equation with distance at VIP space that general space exiting reverberation not in the semi-anechoic chamber. In this experiment, we found out the fact that because of the differences of attenuation levels with distance according to frequency, and distortion of spectrum shape of original sound, original sound can be transformed unintentionally. Also we can found out that it was effective to keep the tone of the original sound by reproducing the original sound with frequencies before play and playing after adjust through the speaker. If we use this technology, it will be possible to reproduce the tone of original sound at the center of Multi-channel-based audio Sweet Spot.

As we develop real-time optimized stereo technology by tracking dynamically the moving audience unlike providing an optimized stereo home theater technology for existing static audience, we expected to create a new concept to experience this dynamic space and create contents and use them. We are going to expand this studies to calculate adjustment coefficient of room size, structure, wall materials, etc that causes reverberation, we are going to continue research about manners to consider the reverberation and attenuation simultaneously.

Acknowledgments. This research is supported by the Industrial Strategic technology development program, 10041772, (The Development of an Adaptive Mixed-Reality Space based on Interactive Architecture) funded by the Ministry of Trade, Industry and Energy (MOTIE).

References

1. Song, M.S., Zhang, D., Florencio, D., Kang, H.G.: An Interactive 3-D Audio System With Loudspeakers. IEEE Transactions on Multimedia 13(5), 844–855 (2011)
2. Jeong, S., Lee, E., Yoo, S., Kim, S.-Y.: A Study on Propagation Characteristics of Acoustic Signals in Indoor Environments. KICS 36(2), 119–125 (2011)
3. Lee, J., Kim, Y., Yoo, S., Kim, S.-Y.: An Objective Performance Analysis of Crosstalk Cancellation Scheme for Sound Rendering Systems Based of Listener Position Tracking. KICS 36(2), 112–118 (2011)
4. Park, J.-H.: A Study on the Prediction of Plumbing Noise in the Machine Room Using Acoustic Simulation. Master's thesis, Wonkwang University, Architectural Engineering (2004)

5. Liebe, H.J., Manabe, T., Hufford, G.A.: Millimeter-Wave Attenuation and Delay Rates Due to Fog/Cloud Conditions. IEEE Transactions on Antennas and Propagation 37(12), 1617–1623 (1989)
6. Makivirta, A., Antsalo, P., Karjalainen, M., Valimaki, V.: Low-Frequency Modal Equalization Of Loudspeaker-Room Responses. In: Proc. of ARS 111th Convention, New York, NY, USA (2001)
7. Jeong, I.-R., Kim, J.-Y., Yoon, S.-C.: Latest Noise and Vibration Theory. Shin Kwang (2009)

A Study on the Resource Management against Availability Attacks in Cloud Computing

Sung-Min Jung[1], Jun-Kwon Jung[1], Tae-Kyung Kim[2], and Tai-Myoung Chung[1]

[1] Department of Electrical and Computer Engineering, Sungkyunkwan University,
300 Cheoncheon-dong, Jangan-gu, Suwon-si, Gyeonggi-do, 440-746, Korea
{smjung,jkjung}@imtl.skku.ac.kr, tmchung@ece.skku.ac.kr
[2] Department of Liberal Art, Seoul Theological University,
Sosabon-dong, Sosa-gu, Bucheon-si, Gyeonggi-do, 422-742, Korea
tkkim@stu.ac.kr

Abstract. In cloud computing, the computing resources in a different physical location are integrated into various services by a virtualization technology. Cloud computing has currently received significant attention due to its many advantages. However, it has security threats of the traditional IT and also has new security threats caused by the structural features of virtualized environments. In particular, availability attacks are very threatening to cloud computing due to the feature of resource sharing. In this paper, we present the scheme to analyze the workloads by allocation of new virtual machines against availability attacks on system resources. We discuss the additional role of a resource broker to threat detection and apply proper an allocation method. Also, we define some factors and analyze several equations to get a quantitative standard against attacks on system availability. System threats and workloads can be analyzed based on these factors. If some availability attacks are detected, then we can apply a proper allocation method against these attacks by using a proposed scheme.

Keywords: Cloud computing, Resource management, Availability attacks.

1 Introduction

Cloud computing provides various services by a virtualization technology associated with network, software, and hardware of the system. Cloud computing has currently received significant attention as a new computing paradigm which can replace the traditional client/server model. Cloud computing provides various virtualized resources to end users as the form of services through the network. The end users lease computing resources such as software, storage, and server as needed. Also, they are supported scalability in real time, and should pay as what they use.

In cloud computing, computing resources are in a different physical location, and they are integrated into various services by a virtualization technique[1]. Therefore, cloud computing has been recently emerging as an important technology. As interest increases in cloud computing, it is needed the research about new security threats of

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_61, © Springer-Verlag Berlin Heidelberg 2014

cloud services as well as existing security threats of a client/server model. It has to be addressed data confidentiality, integrity and availability of the stored data in cloud computing system. In particular, security threats on the availability of infrastructures such as servers and storages in data center are should be solved in order to provide reliable cloud services. Because interruptions or delays of cloud services can bring unexpected and serious damages, it is essential a management technology which can provide high availability and operate infrastructure resources based on proper threat analysis[2].

In this paper, we present the suitable scheme to detect threats, and apply a proper allocation method by a resource broker in cloud computing. We discuss several mathematical analyses to calculate the quantitative standard for the resource broker. The remainder of this paper is organized as follows. Section 2 introduces new security threats in cloud computing, and the concept of a resource broker. Section 3 discusses some mathematical analysis to detect security threats and to determine the suitable policy in the proposed scheme. Finally, section 4 concludes the paper.

2 The Security Threats in Cloud Computing

An availability attack generates large amount of traffic to specific system on the network. Therefore, it depletes the system resources, and end users do not receive reliable services. A distributed denial of service attack is one of availability attacks. There are many tools of this attack, and therefore it can be relatively quick and easy to attack. While it is possible to attack on various types, it is difficult to detect this attack or to find its source[3].

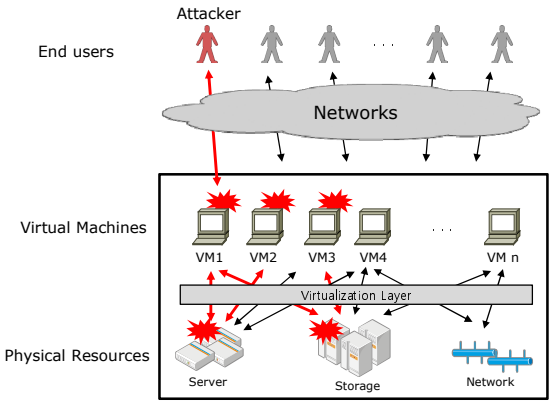


Fig. 1. The new security threat in cloud computing

In cloud computing, it is very vulnerable to an availability attack due to the structural features of cloud system. Figure 1 shows the structure of cloud system. There are physical resources such as a server, storage, and network. They are assigned as virtualization devices which shares physical resources, and provided to end users

through the virtualization layer. End users can connect to cloud system through the network, and cloud system operates several virtual machines(VM). In this case, if malicious user attacks against VM1 to reduce system availability, then it is difficult to provide reliable services due to system workloads. In figure 1, VM1 shares a processing unit with VM2, and shares a storage with VM3. Thus, the workloads of VM1 affect the performance of VM2 and VM3 at the same time. When physical resources are provided as logical resources via a virtualization layer, availability attacks on one virtual machine can affects other virtual machines which share the resources between them. Therefore, due to this feature that virtual machines share the physical resources, availability attacks on infrastructure are very intimidating. A cloud service provider ensures the certain level of service quality according to a service level agreement. Next we discuss the additional role of a resource broker to solve this problem.

There are three layers such as physical resources, virtual machines, and a resource broker in our system model. A resource broker is located between a virtual machine and end users. End users request the necessary resources over the network, and a resource broker verifies these requests and processes them. Also, the appropriate responses should be performed by a resource broker when availability attacks occur in cloud system. The resource broker monitors and analyzes network and system resources. Also, it determines the strength of threats and workloads of the system based on these results. We define some engines to analyze threats and workloads in the resource broker, and figure 2 shows these engines.

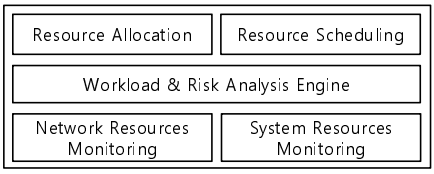


Fig. 2. The role of a resource broker

A network resource monitoring engine monitors network resources such as number of issued connection requests and replied connection requests in the system. Also, it analyzes the number of data transmission in bytes at regular time intervals. System resource monitoring engine monitors system resources such as response time, the number of virtual machines. The results are delivered to the workload and threat analysis engine. This engine analyzes the results and determines the strength of threats and workloads. Finally, a proper policy is selected, and resource allocation or scheduling engine applies this policy to manage the system resources.

It is needed to define several resources in order to measure quantitatively threats and workloads in cloud computing. We set transmitted data size, the number of connection requests, and the number of connection responses associated with network resources[4]. Also, we set execution time of tasks, the number of virtual machines associated with system resources. Base on these factors, we analyze packet drop rate and network bandwidth consumption to determine the system threats.

3 Analysis of Threats and Workloads

In our system model, we present the structure of resource management to analyze the system threats and workloads as shown in figure 3.

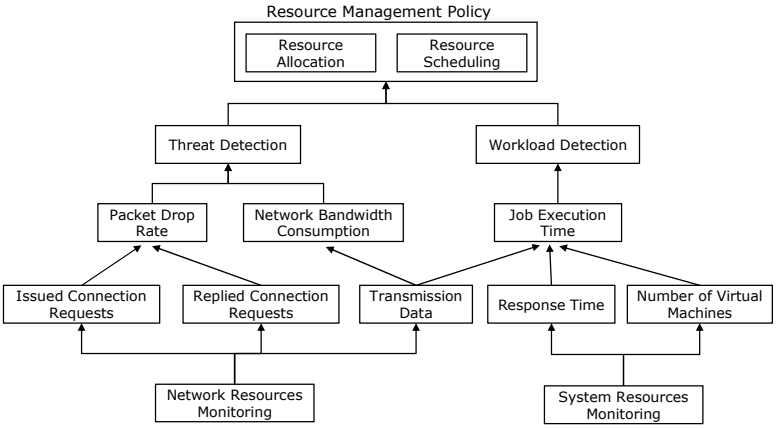


Fig. 3. The structure of resource management

A resource broker analyzes several factors such as the number of data transmission(TR), issued connection requests(CR), replied connection requests(RP), the execution time(ET), and the number of virtual machines(VM) in table 1.

Table 1. The parameters in equations

Type	Parameters
The number of data transmission in byte	*TR*
The number of issued connection requests	*CR*
The number of replied connection requests	*RP*
The execution time	*ET*
The number of virtual machines	*VM*

Packet drop rate NR_{dr} is represented as the rate of the total number of non-replied connection requests to the total number of issued connection requests. Equation (1) indicates the packet drop rate.

$$NR_{dr} = \frac{CR - RP}{CR} \tag{1}$$

Network bandwidth consumption NR_{bw} is defined as the average value between maximum and minimum numbers of data transmission in byte over a unit period of time. Equation (2) indicates the network bandwidth consumption. TR_{init} means the

initial number of data transmission in byte. We determine the strength of threats according to the product of NR_{dr} and NR_{bw}. Also, a resource broker needs a certain threshold to determine the strength of threat.

$$NR_{bw} = \frac{TR_{\max} + TR_{\min}}{2 \cdot TR_{init}} \tag{2}$$

At the same time, the workloads of a system also should be measured to manage cloud system properly. We should calculate the execution time to determine the workloads. It is assumed that the new distribution of virtual machines will linearly improve the system performance[5]. We define that the initial number of virtual machines is VM_{init} and the number of virtual machines in operation is VM_{cur}. VM_{init} virtual machines are allocated, and the execution time in each unit is uniformly distributed from ET_{min} to ET_{max} at first. ET_{min} is the lowest execution time and ET_{max} is the highest execution time.

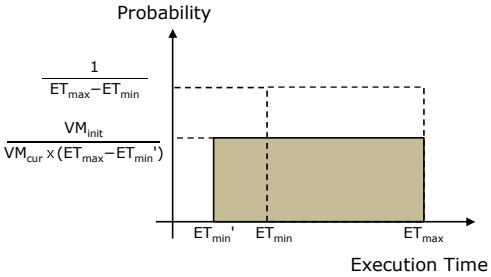


Fig. 4. The execution time in allocation case

In allocation case, new virtual machines are distributed to process tasks. So, the lowest execution time will be smaller than the previous one. The number of virtual machines and transmission data is VM_{cur} and TR respectively. Hence, one virtual machine is responsible for TR/VM_{cur} data, and the total execution time(ET_{total}) is represented as the product of TR/VM_{cur} and the expected value of execution time of one unit. Figure 4 illustrates that the change of values in the allocation case. Equation (3) represents the average of execution time.

$$ET_{total} = \frac{TR}{VM_{cur}} \left(ET_{\max} - \frac{ET_{\max} - ET_{\min}}{2 \times VM_{init}} \cdot VM_{cur} \right) \tag{3}$$

We can use these equations to detect the threat of the system and determine suitable method associated with workloads. Figure 5 shows the assumption of threats in the system. First, we set ET_{min} is 0.001 seconds. Also we define that ET_{max} is up to three times larger than the ET_{min} as the system threat, and different allocation method is applied as ET_{max}. The threshold about the ratio between ET_{max} and ET_{min} is 0.3 in our evaluation. Thus, up to 30% of the VM_{cur} is allocated to reduce workloads as equation (3).

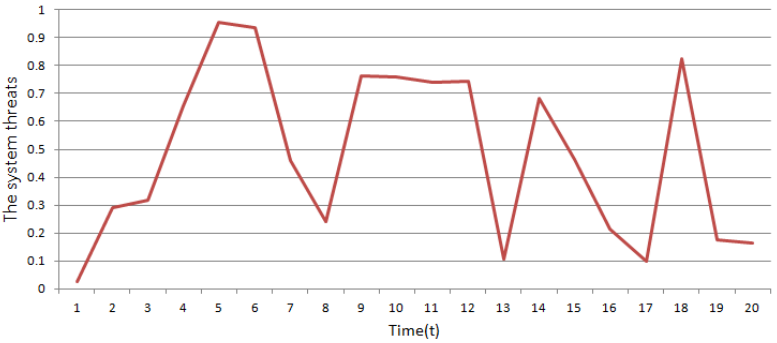


Fig. 5. The assumption of system threats

Figure 6 shows the results of evaluation of execution time in regular period. It shows the difference between the two cases. We compare the case with an allocation method and the case with a normal method. As the result of evaluation, the execution time can be reduced by using a suitable allocation method. In each case, the total execution time is 0.0666 and 0.0816 respectively.

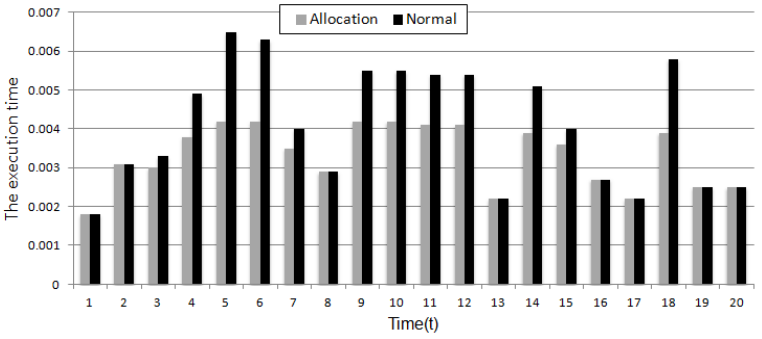


Fig. 6. The simulation result

4 Conclusion

Cloud computing has currently received significant attention because of its various advantages. However, cloud computing has security threat of the traditional IT and also new security threats caused by the features of resource sharing. In this paper, we discuss the additional role to determine the strength of threats and workloads in a resource broker. A resource broker has several engines to analyze the system threats and workloads. It provides the basis to apply proper allocation method. With this method, the suitable scheduling scheme is needed to operate efficiently the cloud system. However, the scheduling scheme is not considered in this paper. In the future,

we wish to research and apply the suitable scheduling scheme to maximize the performance of cloud computing.

Acknowledgements. This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2010-0020210).

References

1. Youseff, L., Butrico, M., Da Silva, D.: Towards a Unified Ontology of Cloud Computing. In: Proc. Grid Computing Environments Workshop (GCE), pp. 1–10 (2008)
2. Voorsluys, W., Broberg, J., Buyya, R.: Introduction to Cloud Computing, Cloud Computing Principles and Paradigms. Wiley (2011)
3. Douligeris, C., Mitrokotsa, A.: DDoS attacks and defense mechanisms: classification and state-of-the-art. Journal Computer Networks 44(5), 643–666 (2004)
4. Chen, X., Li, S., Ma, J., Li, J.: Quantitative threat assessment of denial of service attacks on service availability. In: International Conference on Computer Science and Automation Engineering (CSAE), pp. 220–224 (2011)
5. Yeo, S., Lee, H.-H.S.: Using Mathematical Modeling in Provisioning a Heterogeneous Cloud Computing Environment. IEEE Computer 44, 55–62 (2011)

Development of a Simulation Tool for the Face Recognition Using Feature Feedback

Nguyen Trong Nghia[1], Chang-Woo Park[1], Sang-Il Choi[2], Sang-Hoon Ji[3], and Gu-Min Jeong[1,*]

[1] Kookmin University, Korea
[2] Seoul National University, Korea
[3] Korea Institute of Industrial Technology, Korea
gm1004@kookmin.ac.kr

Abstract. In this paper, we present a novel simulation tool to analyze the results from Feature Feedback-based pattern recognition method especially for the face recognition. Using Feature Feedback, we can choose an important region that affects the classification and utilize the region for the recognition. To do so, a reverse mapping is used from the feature space to the original data space. However, in Feature Feedback, there are many parameters such as number of Fisherfaces, threshold value, etc. used to obtain better results. The simulation tool described in this paper makes it easier to evaluate the effect of those parameters on the recognition rate by optionally changing their values. It also helps the students have an intuitive look on Eigenface, Fisherface or feature mask which are important for the understanding of the Feature Feedback and pattern recognition theory.

Keywords: Feature Feedback, feature extraction, Fisherface, Eigenface, simulation tool.

1 Introduction

Pattern recognition has received much attention due to it theoretical challenges as well as applications in many science domain such as engineering, astronomy, biology, etc. There are a lot of methods have been proposed in the slast few decades. The efficiency of each method is decided based on the recognition rate, running time as well as occupied memory. In order to improve the recognition rate, some conventional feature extraction methods such as PCA or LDA were introduced. The main idea of these methods is that the high dimensional input data will be reconstructed to lower dimensional data, which contains only the most important features of original input data. With the reconstructed input data, the recognition process can be performed more effectively and the recognition rate can be improved.

Feature feedback based method has been proposed to select important region based on the selected features from feature selection methods. By using a reverse mapping

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*, Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_62,

from feature space to original space, we can reduce the dimensionality of the input data and considerably increase the recognition rate

Feature Feedback [1]-[4] has been applied to various applications such as face recognition, e-nose data recognition, etc. However, a few parameters such as number of Fisherfaces, threshold value, etc. should be adjusted in Feature Feedback to obtain better recognition results.

Considering these facts, in this paper, we present a simulation tool to show all the steps systematically and analyze the results depending on the parameters. The simulation tool enhances the understanding of pattern recognition and Feature Feedback as well as adjusting parameters to derive better recognition rates.

The rest of this paper is organized as follow: In section 2, we briefy overview related works. The description of simulation tool implementation is shown in section 3. Finally, Section 4 is the conclusion.

2 Related Word

2.1 Feature Feedback [1]

The process of a Feature Feedback-based face recognition system is modeled as in figure 1.

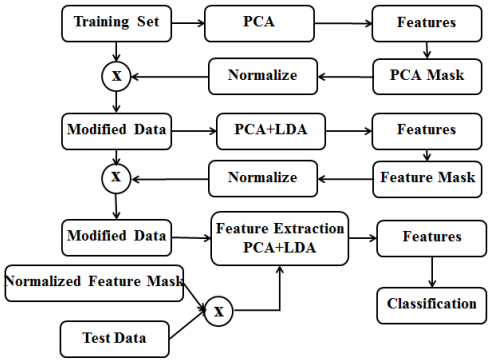


Fig. 1. Flow of Feature Feedback [1]

The detailed Feature Feedback procedure can be summarized as follows:

Step1: Obtain the projection vector from the original data by using PCA

Step2: On the basic of average value of projection vector, obtain the modified data where noise reduced

Step3: Extract Fisherfaces by using PCA+LDA to modify data. It defines Fisherfaces with the final transfer vector W.

Step4: Extract feature mask by taking binary to the final transfer vector W

Step5: Obtain the feature face data by multiplying the final mask to original data element by element.

In step 3 we take the final translated vector W through PCA+LDA. W is then divided into important and unimportant part by using threshold value T in (1)

$$\begin{Bmatrix} w_l \in FI_l \\ w_l \in FU_l \end{Bmatrix} \quad \begin{matrix} if \|w_l\| \geq T_l \\ otherwise \end{matrix} \} \tag{1}$$

We construct the final feature mask with the segmented Fisherfaces. Using the OR operation (2), we can obtain the feature mask.

$$FI = FI_1 \ (+) \ FI_2 \ (+) \ FI_3 \ (+) \ \ldots \ (+) \ FI_m \tag{2}$$

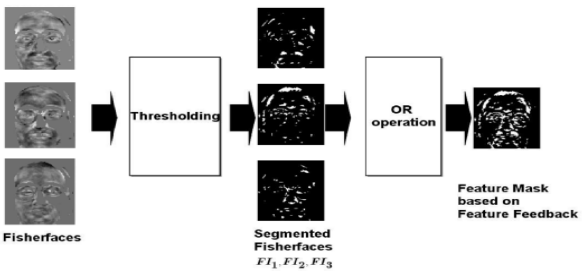


Fig. 2. Extracted feature mask (N=3)

The white pixels on the feature mask can be thought of as the important pixels in the face images. Masking images extracted by covering feature mask to original images are utilized as the input of the classification. This step is shown in Figure 3.

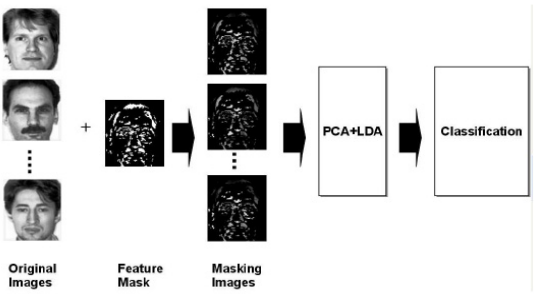


Fig. 3. Face recognition based on Feature Feedback

2.2 R-LDA Based Feature Feedback

In the proposed method, we use PCA+LDA to obtain the feature mask in Feature Feedback stage. By using PCA to reduce the input dimensionality, we can avoid the SSS (Small Sample Size) problem caused by LDA. But a potential problem can occur

is that the PCA criterion may not be compatible with the LDA criterion, thus the PCA step may discard dimensions that contain important discriminative information.

The regularized LDA (R-LDA) was developed in order to overcome this problem. The detail explanation as well as RLDA equations are shown in detail in [5]. Figure 4 shows the detail procedure of a face recognition system using RLDA based Feature Feedback.

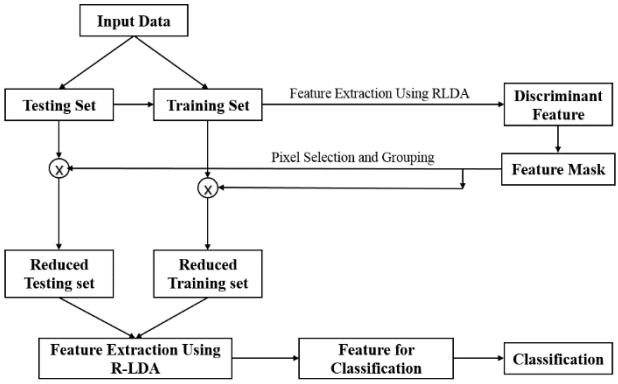


Fig. 4. The R-LDA based Feature Feedback procedure

From the previous experiments with Yale database, it is shown that we can archive a higher recognition rate with Feature Feedback compared with some popular face recognition methods such as PCA, LDA, etc. In addition, by selecting only important pixels for the classification, we can save considerable running time as well as usage memory.

2.3 Previous Experiments

To improve the classification rate of a face recognition system using Feature Feedback as well as evaluate the effect of some parameters on Feature Feedback method's efficiency, we have done several experiments and these are some noticeable obtained results:

- We can improve the efficiency of Feature Feedback by optimizing the number of Fisherface parameter[3]
- Using PCA instead of PCA+LDA in Feature Feedback, we can also get a higher recognition rate in comparison with PCA+LDA without Feature Feedback [2].
- A face recognition system using RLDA [4] for both Feature Feedback stage and classification stage can obtain a higher classification rate compared with PCA+LDA-based Feature Feedback.

3 Performance Enhancement by the Development of a Simulation Tool for Feature Feedback

3.1 Development of a Simulation Tool for Feature Feedback

In this section, we present a simulation tool for a face recognition system using Feature Feedback. Fig.4 shows the simulation tool's user interface.

The simulation tool provides following functions:

- We can change the method of obtaining feature mask in Feature Feedback, including PCA+LDA, PCA or RLDA. We can also choose which type of image (normal size, zoom in, zoom out, compressed non-compressed) we will use in our recognition system. All of these options are available in Operation Options section.
- In Input Parameters section, we can modify the value of Feature Feedback parameters such as: number of Fisherfaces N, threshold value T or number of extracted features used for classification process.
- We can also edit the personal information of output image and save it back to database, this makes the simulation tool more practical. This action can be done in Face Recognition Result section.
- Finally, we can see the running time and Feature Feedback component images (Eigenfaces, Fisherfaces, etc.) for each simulation time. This option can be chosen in Analysis Image section.

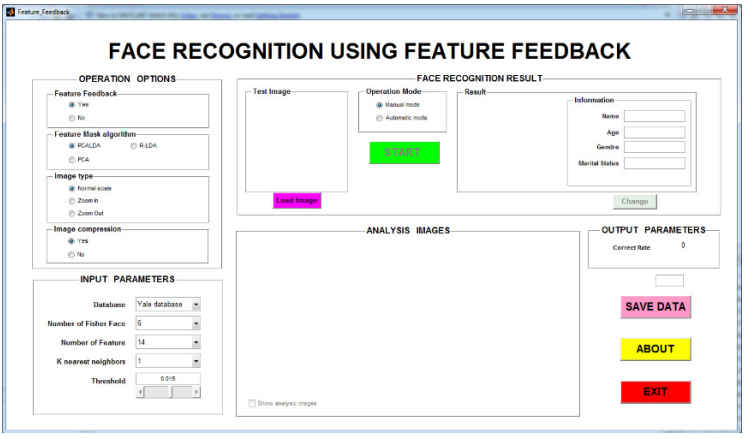


Fig. 5. Simulation tool's user interface

Fig.5 shows 15 images in testing set extracted from Yale database

Fig.6 shows the result after finishing an example simulation. We can see the adequate output person as well as his editable personal information. The running time for the whole simulation process (2.6914s) also is shown

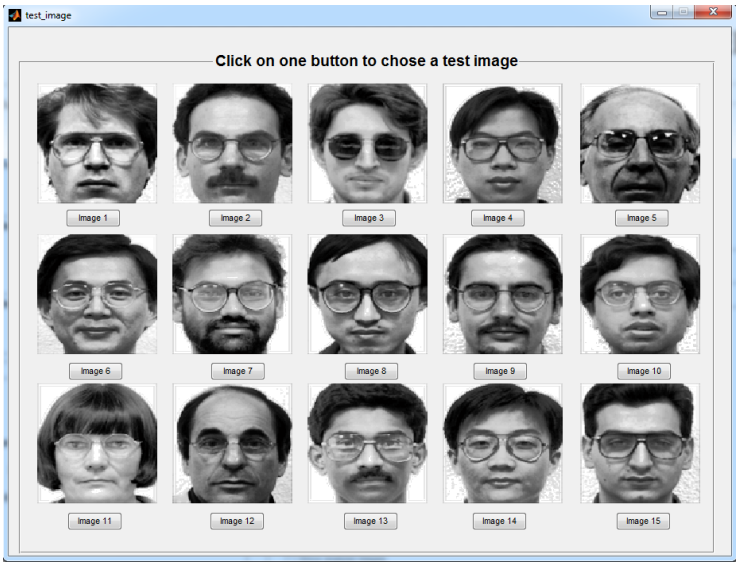


Fig. 6. Input images

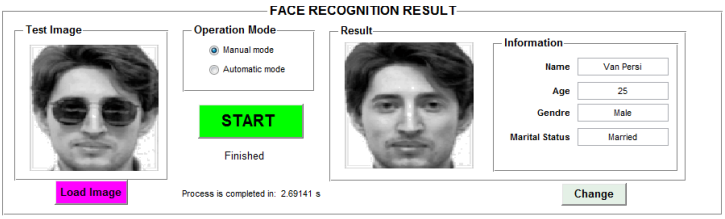


Fig. 7. Input and result image

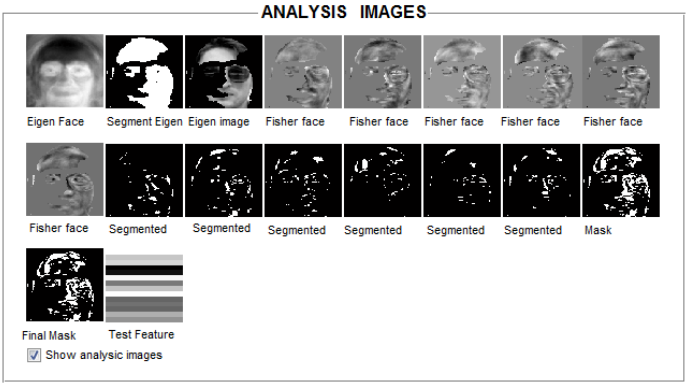


Fig. 8. Feature Feedback component images with 6 Fisherfaces

Fig.7 shows all the Feature Feedback component images such as Eigenface, Fisherfaces, feature mask, etc. By doing this, it will be easier to understand the Feature Feedback procedure described in section 2.1. Actually, the first image is Eigenface, which is the result from step 1. The modified image from step 2 in section 2.1 is the third image and so on. These images help us have an intuitive look at Feature Feedback and pattern recognition theory.

3.2 Performance Enhancement with the Simulation Tool

Using this simulation tool, we can easily evaluate the effect of Feature Feedback parameters on the classification rate by optionally changing their values and then find the optimized value for those parameters. Table1 shows an example when we want to find the optimized value of number of Fisherface parameter with a fixed threshold value 0.015. By changing the number of Fisherface parameter in simulation tool, it is shown that we can obtain the maximum classification rate with 6 Fisherfaces.

Table 1. Classification rate with different number of Fisherface

Number of Fisherface	Classification rate (%)	Running time (sec)
3	82.9	2.79
4	84	2.8
5	85.4	2.82
6	85.5	2.83
8	85.1	2.85

The simulation tool also gives us a convenient way to compare the performance of the algorithms used in our recognition system. By changing the options in Feature Mask Algorithm section of the simulation tool and adjusting the equivalent parameters for each algorithm, we can observe the optimized classification rate as well as the running time of them. With threshold value is 0.015, we can see the results in table 2.

Table 2. Classification rate with different feature extraction algorithms

Feature mask extraction algorithm	Classification rate (%)	Running time (sec)
PCA+LDA (no FF)	84.1	1.5
FF using PCA+LDA	85.5	2.88
FF using PCA	85.2	2.13
FF using RLDA	87	2.35

4 Conclusion

In this paper, we have presented a novel simulation tool to analyze a face recognition system using Feature Feedback. With many functions, the simulation tool helps to

have an intuitive look on Feature Feedback and it effects on a recognition system. By showing all Feature Feedback images after each simulation time, we can easily analyze the procedure of a Feature Feedback-based face recognition system. Furthermore, we have shown that this simulation tool is a convenient way to evaluate the effect of Feature Feedback parameters, such as number of Fisherfaces, threshold value, on the recognition rate.

For this paper, we require using others databases for the simulation such as CMU, Yale B, FERET, etc. We will also require developing a simulation tool for others pattern recognition system using Feature Feedback such as: vapor recognition, taste recognition, etc. These objectives remain as future works.

References

[1] Jeong, G.-M., Ahn, H.-S., Choi, S.-I., Kwak, N.-J., Moon, C.: Pattern recognition using feature feedback: Application to face recognition. International Journal of Control, Automation, and Systems 8, 141–148 (2010)

[2] Truong, L.B., Kim, S.-H., Jeong, G.-M.: Pattern Recognition with Feature Feedback: Feature Mask by Using PCA. In: KMMS 2010, pp. 14–16 (November 2010)

[3] Kim, S.-H., Truong, L.B., Jeong, G.-M.: A Study on the Recognition Rate Ac-cording to the number of Fisherfaces. In: MITA 2010 (2010)

[4] Truong, L.B., Choi, S.-I., Jeong, G.-M., Seo, J.-M.: An Improvement in Feature Feedback Using R-LDA with application to Yale database. In: Lee, G., Howard, D., Ślęzak, D. (eds.) ICHIT 2011. LNCS, vol. 6935, pp. 352–359. Springer, Heidelberg (2011)

[5] Lu, J., Plataniotis, K.N., Venetsanopoulos, A.N.: Regularization Studies of Linear Discriminant Analysis in Small Sample Size Scenarios with Application to Face Recognitoon. Pattern Recognition Letter 26, 181–191 (2005)

Dynamic Data Collection Algorithm with a Mobile Element in Wireless Sensor Networks

SungSuk Kim[1] and Young-Sik Jeong[2,*]

[1] Dept. of Computer Science, Seokyeong University, Korea
[2] Dept. of Multimedia Engineering, Dongguk Univeristy, Korea
sskim03@skuniv.ac.kr, ysjeong@dongduk.edu

Abstract. To improve energy efficiency in wireless sensor networks, there are lots of research efforts. In this paper, we utilize the notion of mobile element (*ME*), which moves and gathers data from sensors within communication range. And sensors resident in nearby area would form a cluster and select one among them to be their cluster node (*CN*). To determine the tour of *ME*, we first devise urgent message-based tour selection algorithm, and then supplement it by considering old *CN*s which do not generate any messages.

Keywords: Wireless sensor networks, cluster head node, mobile element, tour seletion, TSP.

1 Introduction

Wireless sensor networks (WSNs) consist of thousands of sensor nodes (*SN*s) that gather data from deployed environments [1]. Those SNs are typically restricted by the resources due to limited computational power and low battery supply and thus, energy saving techniques/algorithms must be considered. In general networks, multi-hop wireless communications are adopted to deliver gathered data to the sink node, which causes that nodes around the sink still have to consume much more energy than others due to heavier volumes of traffic transmitted by them. To solve the problem, another approach was proposed to utilize the controlled mobility of certain nodes, referred to as mobile element (*ME*) [2]. An ME can move through network areas, taking data through wireless communication technology from SNs nearby itself. By utilizing ME, not only more energy can be conserved and balanced on SNs, but also the communications and networking become possible in very sparse networks with "store-carry-and-forward" approach.

Recently, many related research efforts have appeared in the literature and lots of them focus on tour selection for ME. The tour selection problem with the consideration of the wireless communication range can be modeled as a traveling salesman problem networks (TSPN) [3]. As is well known, it is NP-complete problem and many researchers tried to develop approximated, heuristic algorithms. However,

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_63,

the main assumption in those efforts is that all sensor do the same work; that is, they periodically gather and store the same size data. Therefore, ME just starts from a fixed location (maybe sink node), visits all sensor node just one time along the predetermined tour path and finally comes back to the first location. If the locally stored data size in a sensor is different from that in other sensor nodes as time goes, another tour selection algorithm is needed and we tackle the problem in this paper.

2 System Model

In this paper, we assume cluster-based WSNs to gain a better energy utilization, where SN residents in nearby area would form a cluster and select one among them to be their cluster node (*CN*). *CN* selection is performed in the same manner suggested in our previous work [5]. Sensor nodes including *CN*s periodically gathers data and stored them into local storage. And *CN* relays data pieces received from its *SN*s when ME comes within its communication range [4, 5]. We make the following assumptions about the system:

- Every sensor node has a unique identifier and the local storage is the same size. Sensor nodes have selected a node as its *CN* at initial phase by normal manner. Sensing is done periodically (period : T).
- Each *CN* knows the locations of its neighbor *CN*s and thus it can transmit messages to them via multi-hop routing.
- *ME* knows the locations of all *CN*s. When it visits a CN_i, it aggregates data from all sensors who select CN_i as its *CN*. And then it determines next target dynamically by the proposed tour selection algorithm.

3 Dynamic Tour Selection Algorithm

3.1 Urgent Message-Based Tour Selection (*UTS*) Algorithm

When the local storage in a sensor is filled with data more than in a degree, the sensor has to inform the fact to its *CN* and then the *CN* also initiate urgent message (*Umsg*) to adjacent *CN*s. The message contains the following data:

Umsg = { *SenderID, receiverID, Type, Time, TTL* }

SenderID/receiverI	*CN*'s ID initiating/receiving *Umsg*
Type (Threshold	A or B (A < B)
Time	period value when the message
TTL (Time to Live)	Maximum hop limit during *UMsg*

Data *type* means how much local buffer is filled and can be A or B type. In case of B type, local buffer has less available free area than in case of type A. The message

can have one or more recipients. When neighbor *CN*s get this *Umsg*, it first decrease *TTL* by 1 and propagates the message to other neighbors if *TTL* is not zero. And then it also stores the messages locally to notify *ME* of those messages if it can contact with it. *TTL* value in case of type B is set to bigger than that in case of type A. This is because the sensor initiating the message has less time for its buffer to be full and thus to inform more *CN*s the fact.

When a *ME* visits CN_j, it has to determine the next target *CN* (see Figure 1). In the figure, a sensor sn_0 sends *Umsg* to its $CN(CN_i)$, which relays the message to its corresponding *CN*s and CN_j finally gets the message. After that, when *ME* visits CN_j, it comes to know that a sensor nearby CN_i suffers from storage full-up. Thus it has to consider the fact during determining next target *CN*. In this time, the remaining safe time (*rs_time*) is calculated (in case of type B, T_B is used instead of T_A):

$$\text{rs\_time} = (100 - T_A)/\text{Avg} + (\text{current\_time} - \text{Time}) \quad (1)$$

where T_A (or T_B) is threshold value (percentage value) which means T_A% is filled, *Avg* is the average size of periodically gathered data and Time is gotten from *Umsg*.

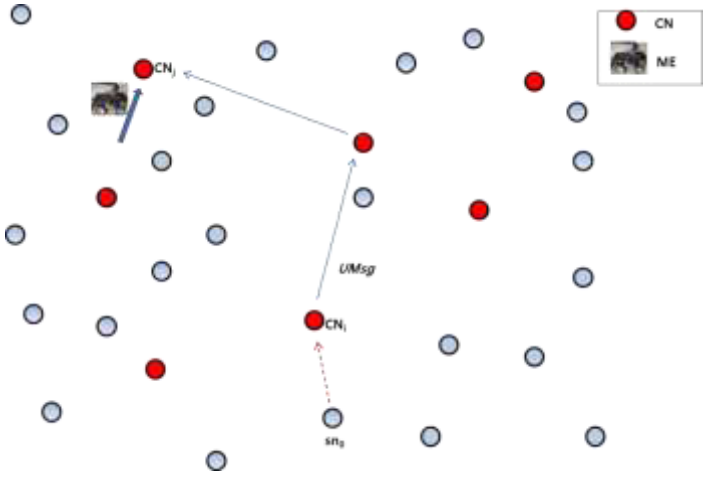


Fig. 1. The next visiting node of *ME* can be determined by urgent messages(*Umsg*)

For example, if a sensor makes a A-Type message when half of the storage is full(50%), T_A is set to 50. (100 - T_A)/*Avg* is constant and thus *rs_time* depends on *Time* value.

If *ME* gets two or more messages from a *CN*, the distance between the current location and the target and *rs_time* are considered together. In this paper, the *CN* near *ME* is chosen for next target in case of same *rs_time*. Of course, in case of no message in *CN*, *ME*'s target will be determined statically.

3.2 U*TS-Q* Algorithm

In the networks, lots of sensors do the same works periodically and the local storage usage pattern may also be similar among them. That means that messages burst may happen in some time later from nearby sensors. Thus data in some sensors may be lost since the local buffer becomes full before it deliver the data to *ME*. To deal with the situation, it is forced *ME* to visit other *CN*s which do not generate any message during long time and is located near the path to the next target *CN*.

As shown in Section 2, *ME* stores the predetermined tour path at the initial setup to use in case of no message. In addition, it has a queue(Q_v) with size n (= number of all *CN*s) to store the list of visited *CN*s. When *ME* visits a CN_i ($1 \le i \le n$), the node will be inserted into Q_v. If it was already inserted, it is first deleted from the queue and reinserted. Therefore the nodes in head side are inserted long time ago and we named it *Old_CN* (figure 2).

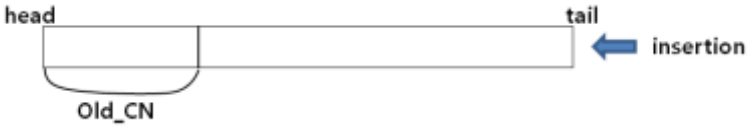


Fig. 2. Visiting Queue (Q_v) stored in ME

When *ME* is going to the next target node CN_t, it also finds the other nodes (CN_{n0}, CN_{n1}) which are located near the path to target. If the node is also a member of *Old_CN*, *ME* visits them before CN_t.

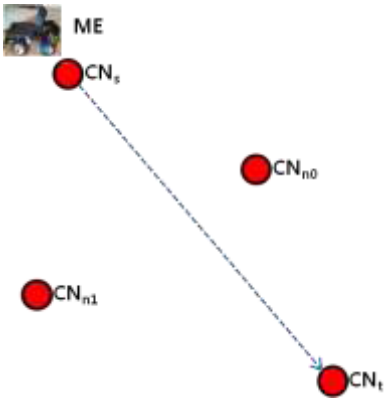


Fig. 3. ME also considers other CNs near the path to target CN_t

4 Performance Evaluation

In this section, we evaluate our algorithms with simulation works, which was developed in C# on Windows 7. To show how much data loss rate decreases, the basic algorithm with no-message mechanism and static path was also implemented.

In the experiments, total field is a rectangle region 500×500 where *N* sensor nodes are deployed randomly and the communication range of the sensors is 5. They works periodically(period : 10) and has a local buffer with the size of 100. The average gathered data size is 1±0.2. T_A and T_B for the type in *Umsg* is set to 50 and 80, respectively. *ME* moves at 30/time speed and is equpped with big size buffer.

At first, table 1 shows the number of selected *CN*s from all sensor nodes, which means that *ME* only visits some nodes while there are more sensors in the networks. From the table, we come to know that there are less *CN*s than the number of sensors, which means that *ME* does not have to move not long distance.

Table 1. Initial setup for CN

# of sensors	500	1000	1500	2000	2500	3000
# of CNs	312	441	495	527	544	570

Fig. 4 shows the experimental results. In the experiments, failure rate is our criterion for evaluation. It is the number of sensors of which the buffer is full over the number of total sensors during experiments. X-axis is the number of sensors and y-axis is the failure rate.

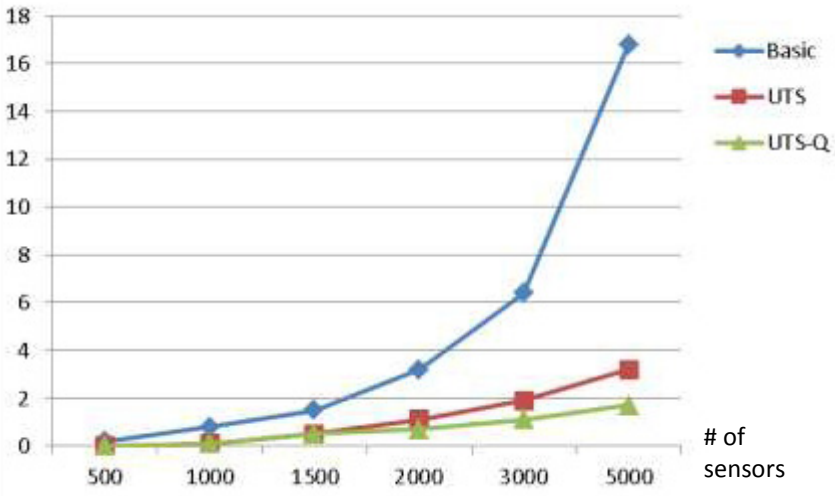


Fig. 4. The comparison of failure rate

From the figure, it is not good for *ME* to move along the predetermined path repeatedly even if the path is optimal (Basic). In contrast, the failure rate decreases when each sensor generates messages to inform that its buffer contains too much gathered data. In addition, the result becomes better when *ME* takes care of the *CN*s that will have a problem in the near future.

5 Conclusion

In this paper, we propose two data gathering algorithms in wireless sensor networks. All sensors are grouped and *CN* is selected as the representative node. And *ME* moves and collects data from nearby sensors. Unlike the other techniques, sensors generate messages to prevent the problematic case. In the future, we will do more experiments under the various kinds of parameters. And we have interests in the dynamic path of *ME*.

References

1. Xu, X., Luo, J., Zhang, Q.: Delay Tolerant Event Collection in Sensor Networks with Mobile Sink. In: Proc. IEEE INFOCOM (2010)
2. Shah, R., Roy, S., Jain, S., Brunette, W.: Data MULEs: Modeling a Three-Tier Architecture for Sparse Sensor Networks. Elsevier AdHoc Networks Journal 1, 215–233 (2003)
3. Gudmundsson, J., Levcopoulos, C.: A Fast Approximation Algorithm for TSP with Neighborhoods. Nordic J. Computing 6(4), 469–488 (1999)
4. Lin, Y.-H., Chang, S.-Y., Sun, H.-M.: CDAMA: Concealed Data Aggregation Scheme for Multiple Applications in Wireless Sensor Networks. IEEE TKDE 25(7), 1471–1483 (2013)
5. Kim, S., Yang, S.-O., Gil, J.-M., Jeong, Y.-S.: Determining the route of a mobile collector based on connecting sensor nodes. In: Proc. IWUCA (2012)

Implementation of the Android-Based Automotive Infortainment System for Supporting Drivers' Safe Driving

Minyoung Kim, Jung-eun Lee, and Jong-wook Jang

Dept. of Computer Engineering, Dong-eui Univ., Gaya 3-dong, Busanjin-gu, Busan, Republic of Korea
kmyco@nate.com, neobart@naver.com, jwjang@deu.ac.kr

Abstract. The automotive infortainment systems providing the navigation and multimedia functions have been developed. Regrettably, none of the systems that have been released up to date have any function seeking drivers' safety. Now, it is required to develop an automotive infortainment system performing an action to support drivers' safe driving in addition to providing convenience for drivers during the course of driving. This paper presents a study on implementation of the automotive infortainment system based on the android platform embedded hardware. The system designed in this paper have additional functions supporting drivers' safe driving (i.e., black box and self-diagnosis functions) as well as functions of the existing automotive infortainment systems (i.e., navigation and multimedia).

Keywords: Infortainment, Block Box, Self-diagnosis, Android, OBD.

1 Introduction

The recent rapid development in information technology has brought various changes. At present, drivers are provided with services through embedded devices while driving. The representative device thereof is the automotive infortainment system that combines the functions of navigation and multimedia (music and video player).

The current automotive infortainment systems provide the functions of navigation and multimedia, and drivers can buy and install such systems by themselves. Some fancy automobiles are provided for drivers with built-in automotive infortainment systems, and such systems also function as the center fascia (controlling panel located between the seats of driver and passenger).

The existing automotive infortainment fails to perform functions seeking drivers' safe driving. The representative functions allowing for drivers' safe driving include the 'automotive black box' function that collects and saves the surrounding environmental information in response to accidents that may occur while driving and the 'automotive self-diagnosis' function using the internal automotive data.

The existing automotive infortainment systems are not designed to provide additional functions. If a driver needs any other functions besides the existing

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*, Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_64,

functions of the infortainment during driving, the driver should purchase and use products with such functions that meet his need. Most of the supplementary functions apply to hardware parts which need to be installed in the automobile before use. If the installation method of the purchased products is complicated, the driver will have to calls in an installation expert. In this case, he should bear the cost thereof.

This paper presents a study on implementation of the automotive infortainment system based on the android platform embedded hardware. The system designed in this paper provides additional functions supporting drivers' safe driving (i.e., black box and self-diagnosis functions) as well as functions of the existing automotive infortainment systems (i.e., navigation and multimedia). Also, it provides an environment where functions can be added at the driver's command, which is not possible under the existing automotive infortainment system.

2 Related Studies for This System

2.1 Existing Automotive Infortainment

The automotive infortainment which is currently released in Korea provides various services through the embedded devices based on MS Windows CE-based. Navigation is served as the main function, and multimedia as the supplementary function. Moreover, by connecting with the driver's smartphone, a smart navigation service that plays the multimedia files saved in the driver's smartphone or provides optimized rout based on the real traffic information through internet is offered.

Each company uses various platforms for the use of the existing automotive infortainment systems but there are barriers in developmental environment. This requires time and costs that the developers have to bear in making researches on the developmental environment and tools for the concerned platforms in order to develop new functions based on the automotive infortainment [1].

Currently, automotive manufacturers and major IT companies play a key role in developing open platforms for automotive infortainment. GENIVI (GENAVA In-Vehicle Infortainment) is one of the representative platforms which have been developed by European automotive manufacturers. GENIVI also functions as a platform for embedded devices and will be possibly used without restriction only if the development related standards are performed. The standards for GENIVI are being developed at present and form the basis of researches on UI (User Interface) and application development.

2.2 Android Platform

Android is a Linux-based platform for mobile devices, which is offered by Google. It is currently used for smartphone and tablet PC, and through Google Play (former Market), users are supplied with applications accessible on Android at a cost or without charge [3].

This paper deals with development of the automotive infortainment for which Android is installed in the embedded hardware. This is intended to provide the

multimedia and navigation functions through the existing Android applications and qualified applications of Market.

This paper aims to develop such infortainment because it provides an environment where users/drivers may add any functions they want according to their own tastes through Android Market. With the applications verified by Market, drivers can be provided with excellent functions of navigation and multimedia and also add the functions they want.

Furthermore, as an operating system installed in smartphone and various embedded devices, this infortainment offers drivers with a stable environment for mobile operation and a familiar environment for mobile use so that the drivers may use this infortainment easily without inconvenience.

2.3 Collection of Driving Information

It is necessary to collect the internal automotive data in order to perform the black box and car self-diagnosis functions of the infortainment implemented in this paper. The internal automotive data is collected through the automotive ECU (Electronic Control Unit) having the major system information and the OBD (On-Board Diagnostics) protocol confirming the existence of breakdowns.

3 System Configuration and Components

The automotive infortainment system designed in this paper consists of two components on a large scale: 'Embedded Hardware and Touch Screen' which directly provide drivers with infortainment services and 'OBD Data Collector' which collects the internal automotive data through the OBD protocol for transmission to the embedded hardware (Fig. 1).

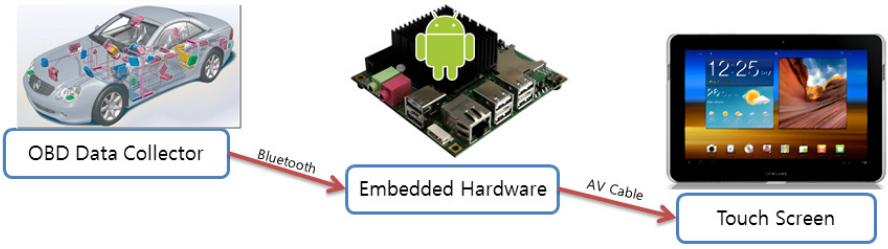


Fig. 1. Diagram of System Configuration

3.1 Hardware and Touch Screen

The infortainment system of this paper uses the Embedded Hardware installed with the Android platform as the operating system. The development presented in this paper is based on ODROID-X2 of Hard Kernel with CPU including 1.7GHz Exynos4

Quad Core. The major features of this product involve compacted size, expendability with various ports of input and output, ultrafast Exynos4-based CPU, and built-in Android 4.0(ICS) [5].

For Touch Screen, a 7-inch product available on ODROID-X2 that supports multi-touch is used so that users may select the functions they want on the screen (Fig 2).

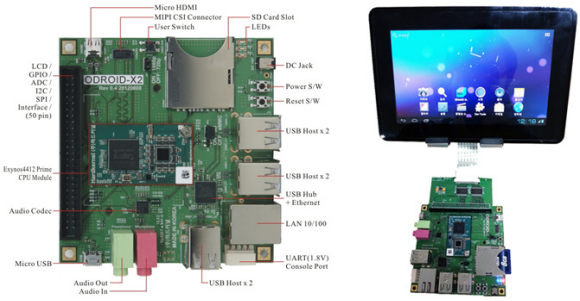


Fig. 2. Appearance of ORDROID-X2 (Left) and ODROID-X2 with 7-inch touch LCD (Right)

3.2 OBD Data Collector

The OBD Data Collector (hereinafter, the "Collector") used in this paper is physically connected to the internal automotive network and collects data through the OBD protocol to transmit the collected data to the service provider hardware via Bluetooth.

This equipment uses highly efficient 32Bit MCU (Micro Controller Unit) based on ARM CORTEX 9 core in order to collect the internal automotive data at high speed, normalize the transport protocol, and the MCU uses the OBD Interpreter Chip in order to search and collect the OBD data with speediness and correctness (Fig. 3).

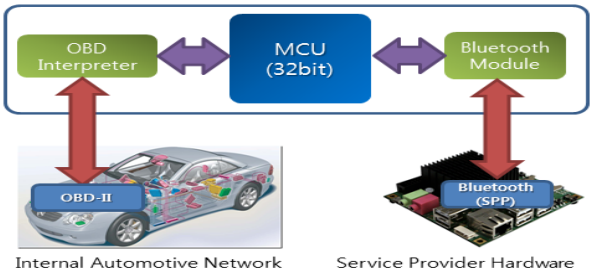


Fig. 3. Block Diagram of 'OBD Data Collector' Inner Structure

3.3 Service Provider Function

The functions of the automotive infortainment system implemented in this paper are carried out by applications for the Android platform.

The multimedia function is replaced with the applications such as 'Music', 'Movie Player' and 'Gallery' which the Android platform basically provides.

The navigation function is replaced with the application such as 'Navigation' which the Android platform basically provides.

Drivers may use the functions of multimedia and navigation by way of installation of the applications that meet their own tastes through 'Google Play Store', instead of the existing applications. If drivers install additional applications, while driving, they may use the functions that they want under the infortainment system of this paper (Fig. 4).

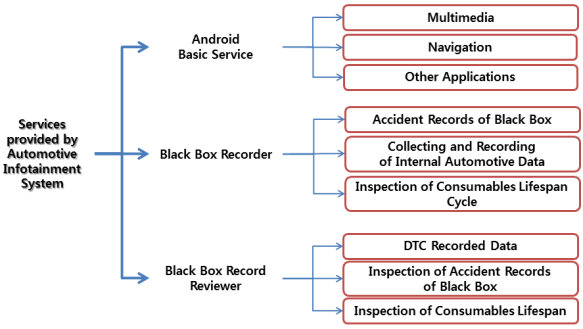


Fig. 4. Composition of Functional Services Provided by Automotive Infortainment System

The functions of black box and automotive self-diagnosis are provided through two applications such as 'Black Box Recorder' and 'Black Box Record Reviewer' which have been autonomously developed.

The Black Box Recorder application operates at all times in order to record traffic accidents that may occur while driving. The Black Box Recorder application saves the internal automotive data collected by the Collector, together with driving movie data through a built-in camera of the service provider hardware. Moreover, while driving, this application inspects consumables life cycle by using the information of the Collector and notifies users of the consumables whose lifespan is less than 1000km only. In addition, while driving, it provides the function allowing the drivers to know internal automotive breakdowns in real time by using DTC (Diagnostic Trouble Code) through OBD.

The Black Box Record Reviewer application functions as a reviewer of black box's accident records recorded while driving, DTC recorded data, and consumables lifespan. The accident records are provided by playing the recorded video and printing out the collected internal automotive data and the GPS information so that the scene of traffic accidents may be recreated. DTC recorded data are provided according to the found date by printing out the list and details of DTC occurring while driving. The consumables lifespan cycle enables an inspection of the consumables lifespan based on the mileage measured while driving. Also, there is a function that allows users to

determine the standards for mileage and lifespan at their own discretion. Detailed explanation on each part is also provided to enhance drivers' understanding in regard to the importance of parts.

4 System Implementation and Its Result

To verify the functions provided by the automotive infortainment system implemented in this paper, the system was actually installed in a car and a test was conducted.

Fig. 5. Actual Test (Operation of Black Box Recorder and Navigation)

Considering the fact that it is difficult to install in an actual car the hardware and monitor mentioned in this paper, a smart tablet (Samsung Galaxy Note 10.1) was substituted for the hardware in the actual driving test (Fig. 5) in order to examine whether two applications of the 'Black Box Recorder' (Fig. 6) and the 'Black Box Record Reviewer' (Fig. 7, 8 & 9) perform normal functions.

It was confirmed that while driving, the record program operated without problem and that the driver was provided with the functions of navigation and black box.

Fig. 6. Black Box's Recording Function and Its Screen Composition

Fig. 7. Consumables Exchange Cycle Review Function and Its Screen Composition

Fig. 8. DTC Found Record Review Function and Its Screen Composition

Fig. 9. Black Box's Accident Review
(Accident Record List (Left) & Review of Accident Record (Right))

5 Result and Discussion Conclusion

This paper implemented an automotive infortainment system based on the Android platform embedded devices which performs the functions of black box and automotive self-diagnosis by using internal automotive data as well as the functions of the existing systems.

This paper verified the possibility that the Android platform can be appropriate for the automotive infortainment system by way of using the expendability of the Android platform.

However, in order to provide drivers with a perfect automotive infortainment system, it is necessary to reinforce UI and UX for the additional functions such as black box and automotive self-diagnosis and to find out and correct the errors occurring during the course of operation. Therefore, the automotive infortainment system shall be provided for drivers after adding the functions of multimedia and navigation which are more upgraded than the basic functions supplied by the Android platform.

Acknowledgments. This work was supported by the Brain Busan 21 Project in 2013.

References

1. Lee, C. (MDS technology): Current status of the in-vehicle infotainment technology, how far evolution. EP&C NO.292, Korea Electronic Components News, Republic of Korea, 18–26 (June 2012)
2. Kim, J., Han, T.: Trends of the Standard Open Platform for In-Vehicle Infotainment and GENIVI based Human Machine Interface. Journal of KIISE: Software and Applications 39(6), 444–452 (2012)
3. Wikipedia Korean Pages, http://ko.wikipedia.org/wiki/
4. Beak, S.-H., Jang, J.-W.: A implement of vehicle Blackbox system with OBD and MOST network. In: Proceedings of the Korean Institute of Information and Communication Sciences Conference 2010, Republic of Korea, vol. 14(2), pp. 66–69 (2010)
5. ODROID Web Site, http://www.hardkernel.com

Design and Implementation of an Around View Image Storing System Based on Car PC

Sang-uk Seo, Sweung-hwan Cheon, and Si-woong Jang

Dept. of Computer Science, Dong-eui Univ., Gaya 3-dong,
Busanjin-gu, Busan, Republic of Korea
{ssu382,perari0}@nate.com, swjang@deu.ac.kr

Abstract. The automotive electronic systems for effective driving include a navigation reflecting real-time traffic information, a blind spot camera for accident prevention, a black box for recording driving image prior and after accidents, an around view monitoring system enabling all the sides to be seen, etc. The existing around view monitoring system among them was developed based on the embedded system in order to provide real-time information around a car in blind spots and narrow spaces. In this paper, we have implemented an around view monitoring system which enables image storing by combining the strong points of the around view monitoring system and those of the black box. In this system, location information which is not provided by the existing AVM systems of automotive companies is added to the around view monitoring system based on Car PC by way of using a USB GPS receiver with which location information and speed can be detected. This study has implemented an image storing system in which latitude, longitude and speed information are detected by using the RMC sentence summarizing the minimum data of GPS through the GPS receiver, and images from four cameras are input and synthesized and then, saved altogether with location and speed information in the storing device.

Keywords: AVM(Around View Monitoring System), Black Box, GPS, Homography transformation, Image distortion correction.

1 Introduction

At present, an around view monitoring system has been developed as a driving auxiliary system to prevent accidents from occurring in blind spots while driving at low speed and from frequently occurring in narrow spaces. The around view monitoring system is a monitoring system which displays the surroundings of a car to be seen at a look through user interface while driving, parking and reversing on a narrow road [1].

The around view monitoring system provides a real-time image displaying function but not an image storing function. In this paper, an around view monitoring system which enables image storing has been designed and implemented by way of combining the strong points of the existing around view monitoring system and those of the black box. Also, in the system, location information (latitude and longitude)

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_65,

and speed information can be seen on the monitor screen by way of adding such information to the functions of the existing around view monitoring system. In Chapter 2 of this paper, studies related to the existing around view monitoring system and black box will be explained, in Chapter 3, design of an around view image storing system, in Chapter 4, implementation of an around view image storing system and its result, and in Chapter 5, conclusion

2 Related Studies

2.1 Analysis of the Existing System

An around view monitoring system is a system to monitor the surroundings of a car at a look in which such surroundings are displayed on the monitor while driving forward, backward and parking, as in Fig. 1.

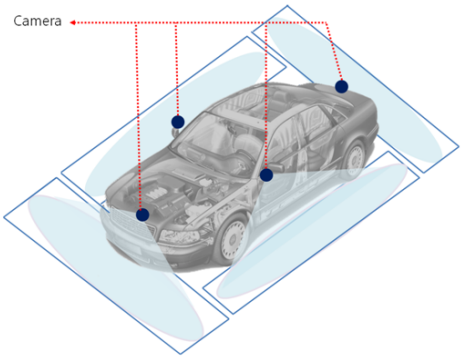


Fig. 1. Installation of Cameras for Around View Monitoring System

In order to display around view images on the monitor, it is necessary to receive images from 4 cameras as located in Fig. 1 and to integrate such images with the automotive images. In order to see 180° images in every direction, wide-angle cameras are required. The use of wide-angle cameras leads to distorted images, so it is necessary to synthesize images after correcting distortion.

2.2 Studies on Image Distortion Correction

A number of studies on image distortion correction have been progressed in Korea. The study proposed a plan to simplify lens correction process and implemented an embedded correction module. It confirmed that only with lower priced fisheye lens, it is possible to restore images to the extent that algorithm used for general robot vision and computer vision may be applied in real time [2]. 'Design and Implementation of Digital Filter Connected with Smart Phone for Bio-signal Distortion Correction' aimed for correction of distorted bio-signals, which is far from correction of distortion in regard to automotive cameras [3].

2.3 Studies on Around View Image Synthesis

'An Embedded System for Vehicle Surrounding Monitoring' contributed to image quality by presenting an algorithm for determining dynamic areas in order to decide which camera image should be outputted in regard to the overlaps between multiple cameras [4].

'Comparison of Image Qualities according to Selection of Pixel Values in the Homography Algorithm' improved the degree of precision by way of dividing an image by the lattice, instead of using the entire image, when performing homography transformation in order to increase the degree of geometric precision in integrating images, but this had a bad point requiring manual labor in part. The reason for requiring such manual labor is that the selection of location of lattice should be accurate in order to achieve accurate homography transformation by the lattice even though the lattice unit is automatically perceived by the program; manual labor may correct parts which are inaccurately perceived so that the degree of precision would increase [5].

2.4 Studies on the H/W Platform and System for Around View Image Processing

'Design and Implementation of a 4-sided Monitoring System Using a Car PC' explained a method to implement around view images on the Car PC platform [6]. 'Design of a Four-sided Monitoring System based on MOST Network' presented a method by which cameras are connected to MOST network in order to obtain around view images; images from the cameras are delivered to a master node through MOST network; and the images are corrected and synthesized on the master node accordingly [7].

3 Design of an around View Image Storing System

The around view image storing system in this paper consists of cameras receiving images, a frame grabber sending images to PC, a Car PC where software works can be done, a GPS receiver to obtain location information and speed, and a displayer showing the completed images. As in the Fig. 2, firstly, the real-time images from four cameras are sent to the Car PC through the frame grabber. In the Car PC, the received data of four images are corrected and synthesized. After the course of synthesis, the data of GPS are collected and put into the synthesized image. After synthesis, the synthesized image and the respective images from cameras are saved.

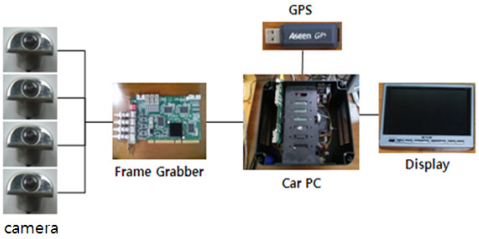


Fig. 2. Concept Graphic of H/W

3.1 Design of an Around View Image Storing System

As in Fig. 3, four images are inputted and go through distortion correction, homography transformation, image synthesis and GPS RMC collection in order to be seen on the displayer. During the course of distortion correction, the images provided by cameras are distorted by physical features of the cameras such as lens, etc. Distortion correction coefficient for the respective cameras in every direction is pre-calculated by using inverse-perspective transform in order to integrate into one image the images from the respective cameras after distortion correction. Four images from the front, rear, left and right sides which performed distortion correction and inverse-perspective transform are shown as one image by way of using α-blending algorithm. In the GPS RMC collection, GPRMC (Recommended Minimum data) provides minimum data which are recommended.

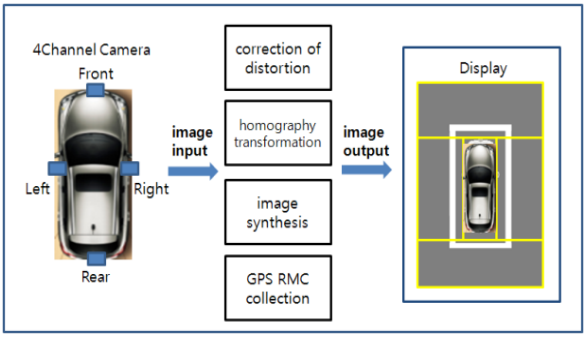


Fig. 3. Design of an Around View Image Storing System

4 Implementation of an around View Image Storing System and Its Result

In the around view morning system implemented by this paper, around view images are created by synthesizing of the images from all the directions through homography transformation in order to map the removed distorted images with other planar images, after removing distortion on the respective images from four cameras in every direction.

4.1 Distortion Correction

Firstly, image distortion should be corrected in order to implement the system. Distortion is corrected by way of distortion correction algorithm with original image obtained. A sphere-based distortion correction method has been used to correct distortion.

Fig. 4. Distortion-corrected Images

4.2 Homography Transformation

Homography transformation is performed to map the respective images after distortion correction into planar images. In this paper, the homography transformation has been performed by using the inverse-perspective transformation as explained in the said design of the system. Firstly, the work scope should be set in order to perform the homography transformation. As in Fig. 5, the initial four dots are placed on the image of distortion correction, based on which precise coordinates are calculated.

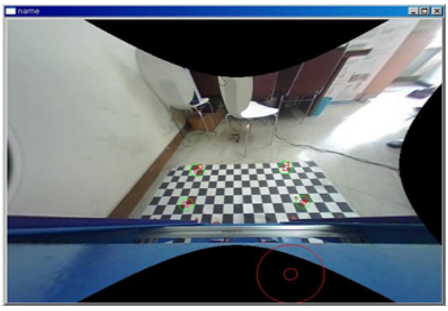


Fig. 5. Designation of Work Scope by Placing Initial Four Dots

After such calculation, the location information of the respective dots of coordinates is input from the inside of source as in the box of Fig. 6, so that the approximate value of each coordinate can be settled.

In addition, homography is calculated by the use of the initial four dots, and inverse matrix is made with such homography. Then, the respective approximate values of dots on the images of distortion correction are calculated by way of multiplying the inverse matrix by the respective approximate coordinates.

```
load = cvLoadImage("input/h0.jpg");

a.x = 0; a.y = 0;
b.x = MX1; b.y = 7;

cvmSet(dst, 0, 0, 3+scale+tx);
cvmSet(dst, 1, 0, 0+scale+ty);
cvmSet(dst, 2, 0, 1.0f);

cvmSet(dst, 0, 1, 15+scale+tx);
cvmSet(dst, 1, 1, 0+scale+ty);
cvmSet(dst, 2, 1, 1.0f);

cvmSet(dst, 0, 2, 15+scale+tx);
cvmSet(dst, 1, 2, 5+scale+ty);
cvmSet(dst, 2, 2, 1.0f);

cvmSet(dst, 0, 3, 3+scale+tx);
cvmSet(dst, 1, 3, 5+scale+ty);
cvmSet(dst, 2, 3, 1.0f);
```

Fig. 6. Internal Source of Homography

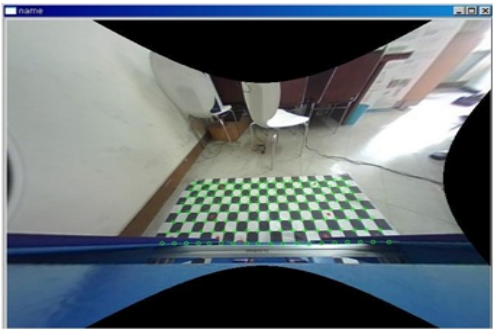


Fig. 7. Newly Observed Dots Storing and Displaying

Homography is calculated again by the use of newly observed dots and then, such calculation is applied to each pattern, which is stored in a form of table as homo.txt. The left of Fig. 8 shows the inside of txt file storing each value of homography in a form of table by the pattern. The right of Fig. 8 shows the preview image where the homography results calculated with the newly observed dots are applied to an image of distortion correction.

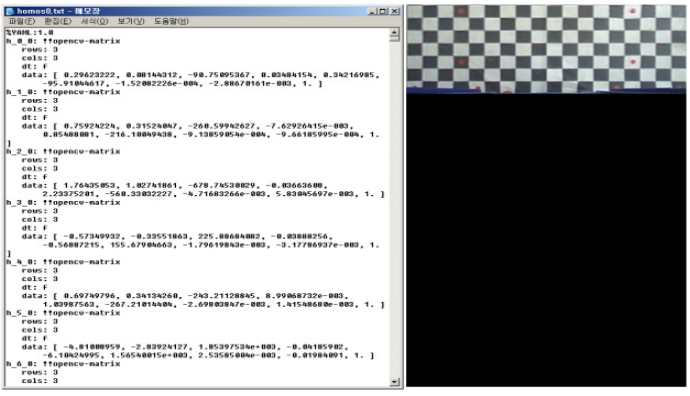


Fig. 8. homos0.txt(left) and g0.jpg(right)

4.3 GPS Information

After image synthesis, location information should be added to the synthesized image. A GPS module for location information is connected to a COM port. As described in Fig. 9, a code is used in order to connect GPS to the COM port.

```
m_Sio.InitComm(9, 9600, NOPARITY, 8, ONESTOPBIT, FALSE);
```

Fig. 9. COM Port Connection

The buffer receiving GPS information should be opened after the port connection and at this time, location information may be continuously received through application of counter codes.

```
BYTE pBuff[256];
DWORD dwBytesRead;
dwByteRead = m_Sio.Read(pBuff, 255);
m_NMEAparser.ParseBuffer(pBuff, dwBytesRead);
```

Fig. 10. Location Information Counter Codes

The received GPS information such as latitude, longitude, speed, etc. is displayed on the monitor after being added to the bottom of the around view image. Fig. 11 shows codes by which the GPS information codes may be outputted on the around view image.

```
m_t.Format("X:%f", m_NMEAParser.m_dGGALatitude);
cvPutText(m_bf, (LPCSTR)m_t, cvPoint(10,477), &m_fnt, cvScalar(255, 0, 0));
m_t.Format("X:%f", m_NMEAParser.m_dGGALongitude);
cvPutText(m_bf, (LPCSTR)m_t, cvPoint(120,477), &m_fnt, cvScalar(255, 0, 0));
m_t.Format("X:%f", m_NMEAParser.m_dGGAGroundSpeed);
cvPutText(m_bf, (LPCSTR)m_t, cvPoint(240,477), &m_fnt, cvScalar(255, 0, 0));
```

Fig. 11. Location Information Counter Codes

4.4 Image Storing

This is a source to save the synthesized result as a movie. Through the cvCreateVideoWriter function, the synthesized image from four channels is saved and also, the images from the four cameras are saved as four files, respectively.

```
If(m_vsave)
{
   If(m_vsaved == 0)
      {
         m_vw = cvCreateVideoWriter(movieName, CV_FOURCC('D', 'I',
         'V', 'X')
         m_vsaved = 1;
       }
   cvWriteFrame(m_vw, m_bf);
}
```

Fig. 12. Image Storing

4.5 Results of System

The test bed for this study is that four cameras were installed on the front, rear, left and right sides of a metal cart; and the Car PC and the monitor were supplied with electricity through batteries. The screen of the 4-channel around view image storing system using Car PC and GPS which has been implemented in such test bed is as in the left of Fig. 13. This is an image in which X(latitude) and Y(longitude) and SP(speed) are printed out and in such a way, GPS information is outputted. The right of Fig. 13 shows an image captured from the movie saved in an actual automobile. In an actual vehicle, the location and speed (km/hr) can be noticed based on the information of latitude and longitude.

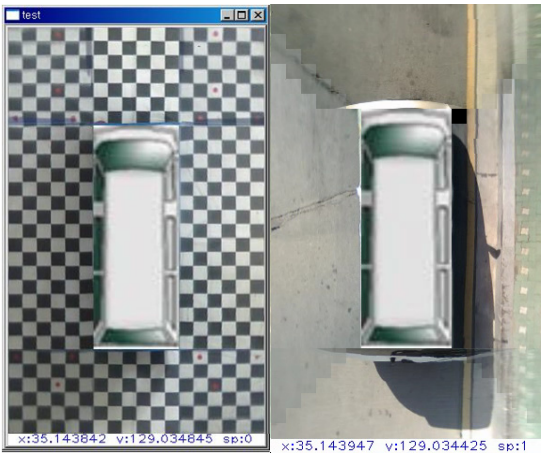


Fig. 13. Image Captured from Movie

5 Conclusion

As auxiliary systems for drivers to prevent accidents from occurring in blind spots and narrow spaces, etc. have became essential, studies thereon have been actively progressed. However, the automotive black box which has been researched up to date saves image/sound information only, but it will be required to be combined with GPS as well as various sensors in the future and accordingly, related studies will need to be performed.

This thesis implemented a 4-channel around view image storing system using Car PC and GPS as well as 4-direction monitoring function. In such implementation, the location information which is not provided by the existing AVM systems of automotive companies is added by way of using a USB GPS receiver with which location information and speed can be detected and an around view image storing function is also added. This system may provide 30 or more frames of around view images per second in real time, so that regardless of the speed of a vehicle, it is possible to monitor the surroundings of the vehicle even while driving at high speed. Also, there is another strong point that functions may be easily added by only software under the environment where Car PC is installed.

The future studies will deal with a graphic related area and implement an easy and clear interface; add various infotainment systems in regard to GPS and image storing functions and strengthen the advantages of Car PC in various forms for such implementation; and upgrade image quality through improvement of distortion correction.

Acknowledgments. This work was supported by the Brain Busan 21 Project in 2013.

References

1. Nissan Motor Co., Ltd., Device and Method for Monitoring Vehicle Surroundings, European Patent
2. Kang, J.-A., Park, J.-M., Kim, B.-G.: The Technical Development for the Fish-Eye Lens Distortion Correction. In: Conference of KOREA Spatial Information Society, pp. 133–138 (2007)
3. Kim, J.-H., Kim, K.-S., Shin, S.-W., Kim, H.-T., Lee, J.-W., Kim, D.-J.: Suppression of Noisy Characteristics of Bio signals by Implementing Digital Filters with an Android Smartphone Platform. The Transactions of the Korean Institute of Electrical Engineers 61(10), 1518–1523 (2012)
4. Chen, Y.-Y., Tu, Y.-Y., Chiu, C.-H., Chen, Y.-S.: An Embedded System for Vehicle Surrounding Monitoring. In: Proc. of Power Electronics and Intelligent Transportation System (PEITS), pp. 92–95 (2009)
5. Yoon, H.-D., Yu, Y.-H., Jang, S.-W.: Comparison of Image Quality according to Choice of Pixel Values in Homography Algorithm. In: Proceedings of the Korean Institute of Information and Communication Sciences Conference, pp. 503–506 (October 26, 2011)
6. Jang, S.-W., Lee, S.-J., Chung, L.: Design and Implementation of a 4-sided Monitoring System using a Car PC. In: International Conference of SDPS 2011 (2011)
7. Jang, S.-W., Yu, Y.-S., Jeon, Y.-J.: Design of a four-sided monitoring system based on MOST Network. In: 2009 International Conference of Maritime Information and Communication Sciences (2009)

Mobile Cloud Computing Architectural Design Taxonomy toward the 'Cloud Computing in Hand' Era

Yoojin Lim and Eunmi Choi

Graduate School of Business IT, Kookmin University, Republic of Korea
emchoi@kookmin.ac.kr

Abstract. Beyond the hybrid cloud computing and its integration service, Mobile Cloud Computing (MCC) starts to cultivate personal cloud computing areas with enormous services and applications with rapid increase of personal use of mobile devices. This paper presents the motivation of MCC with unique features which are discriminated from those of traditional cloud computing, and also shows the architectural design strategies with recent technologies to realize the MCC. We organize the taxonomy of architectural design strategy under five focuses: data-intensive architectural design, mobile processing-oriented design, service-aware design, privacy-sensitive design, and analytical-purpose design. We also outline the design metrics to archive a smart MCC.

Keywords: Mobile Cloud Computing, Cloud Computing, Taxonomy.

1 Introduction

The Cloud Computing paradigm permeates IT applications and services whether we can recognize it easily. Under the personal environment or enterprise delivery, various cloud services provide easy-to-handle processing with a reasonable cost. The Mobile Cloud Computing, so-called MCC, is the area to be an extension of cloud computing in which the foundation hardware consists at least partially of mobile devices, that is defined by E.E. Marinelli [1]. And MCC is also defined as a rich mobile computing technology that leverages unified elastic resources of varied clouds and network technologies toward unrestricted resources to serve mobile devices regardless of heterogeneous environments based on the pay-as-you-use principle, by R. Buyya [2].

Technologies in cloud computing become mature in the infra systems and cloud server-level service providers, which have kept the infra system technologies as the data centers, server clusters, virtualization, service providers by extending services of Private Cloud. To make the Public Cloud popular, the Amazon AWS has pioneered with the EC2, S3, EBS, and various cloud services toward the SME companies and computing intensive service-oriented markets. On top of mature infra system technologies, application service is the next paradigm to launch 'the cloud computing in hand'. We define the cloud computing in hand as personal cloud under convenient environment with ambient intelligence. The Gartner group says that personal cloud

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_66,

offerings expands in scope and capabilities continuously, and it can empower the mobile device user and enable new mobile collaborations in predicts 2013 [3].

As a representative of mobile devices, usage of smart phones spreads over the world-wide people with rapid increase. The International Data Corporation expects strongly that the number of smartphones shipped worldwide will reach approximately 920 million this year, up about 27% from the 722.4 million units shipped in 2012 [4]. To build up the strategies of MCC services and providers for the high population of smart phones, we need to prepare for the next upcoming era of cloud computing in hand. In this paper, we introduce a number of aspects of architecture views on the MCC, criteria of allocating workload modules, and topology of data and service processing to enhance the capability of MCC.

Table 1. Viewpoints of Architectural Design Strategy

Architectural Design Strategy	Considerable Viewpoints
Data-intensive architectural design strategy	• Data storing mechanism and location • Data intensive service • Data management and controller • Data-driven operation
Mobile processing-oriented design strategy	• Mobile devices' characteristics • Ad-hoc topology • Mobile cooperative services and operation • Local segmented processing
Service-aware design strategy	• Cloud service operation and management • VM controller and management
Privacy-sensitive design strategy	• Personal information management • Security mechanism
Analytical-purpose design strategy	• Data collection and logging • Domain purpose analysis • Filtering and aggregation

2 Architectural Criteria for the Mobile Cloud Computing

The MCC has the following considerations that are discriminated from any traditional cloud computing features.

- MCC is for mobile devices like smartphones, laptops, tablets which have network connectivity.
- They have to present inherent lack and poverty of resources even the state-of-the-art devices have excellent specifications of resource such as CPU speed, memory capacity, battery volume, and data communication speed [1][2][5].
- Wireless communication is less stable than wired one and has low connectivity because there is fluctuation depending on the number and the location of the moobile devices. And it would be limited by the telecommunication company in which they are registered [2][5].

- The data what is created in mobile devices for personal usage has a private feature, and its type is multimedia-intensive such as audio, image, and video in large-size data [6].

In these unique features of MCC, we need to consider the architectural hierarchy design to adopt the cloud services into our hands. This paper shows the architectural focuses of those design strategies with recent technologies to realize the MCC. Before showing the detailed contents, we present the consideration viewpoints of architectural design strategies in Table 1.

3 Architectural Design Strategies for MCC

3.1 Data-Intensive Architectural Design

The mobile cloud computing operates with mobile devices, which tends to collect various contents such as audios, videos, and photos. With its native mobility and extended functionality, a huge size of data is accumulated in the device. Also, various applications connecting to web services makes user data to be continuously piled in a mobile device, so the device requires a large capacity of storage. In this context, cloud service is used for this mobile device. In this section, we focus on the architectural design by considering data intensive characteristics that requires extra storage services beyond the mobile device.

Data-Storage Service. A mobile device can generate and store data manually or automatically in cloud storage via network. The purpose of this operation is literally storing. The mobile device has the role of data-sensitive collector, which moves the data from its local file system to the cloud in order to complement its weakness of the storage. For example, a smart phone, which has lot of personal data such as photos and video clips, stores a portion of the personal data in the cloud that provides storage services through network like dropbox[1], amazon s3.[2] The data-storage service does not offer any other extended service.

Data-Oriented Service. Data-intensive service, the extended form of storage service, is to provide specific services according to the type of data which is stored in cloud. The cloud service provider provides not only simply storage service but also more advanced services such as encoding video clips, image retrieval, and connecting each data file with suitable applications. As a result, the provider is able to open convenient service-chains which have combinations of each service such as ndrive[3] and google drive[4].

[1] https://www.dropbox.com/
[2] http://aws.amazon.com/s3/
[3] http://ndrive.naver.com
[4] https://drive.google.com

Data-Circulating Service. When mobile devices upload data to cloud, the service provider uses and combines the data with other information and consequently extracts specifically focused information from the combined data through data processing. The extracting processing applies filter, complex formula, and the analytical rules. Thus, the MCC devices utilize the latest information as necessary. Forensic Cloud [7] is proposed to share digital forensic data and to collaborate with others. When a forensic examiner uploads various file formats to the Forensic Cloud, a file filtering module extracts and saves plain text in HBase[5]. Other examiners probe cases and collects information by searching, merging and analyze cases correlatively.

3.2 Mobile Processing-Oriented Design

With the limited CPU, memory, battery, and hardware configuration, mobile devices will have burden to process large and long computations. They can share the burden by means of their cooperation or interoperability with remote cloud services. In this section, we focus on the architectural design by considering mobile processing.

Cooperative Topology for Mobile Processing. Mobile device becomes part of resource provider and acts like a node of grid computing. On the same topology, the devices can divide workloads and combine results of each computation. Hyrax distributes workload of its computation to the nodes by using server and smartphone based on android platform. And the Hyrax [1] experiments the feasibility of mobile devices as resource provider on its own multimedia searching and sharing service.

Cloud-Connected Topology for Mobile Processing. In this topology, cloud contains high-level servers and supplements mobile device which has the relative poverty of computing capability. And this topology is adopted in various areas to settle the insufficiency and shortage of the resource. In order to provide more suitable video clips for mobile devices, HDVTS [6] proposes the cloud server-side transcoding that encodes various video formats to MPEG4 format by using of the MapReduce framework on Hadoop Distributed File System. Recently, with green energy concept, this topology aims the efficiency of resource usage toward entire system. GEMCloud [8] distributes server-side workload to smart devices which are based on android platform such as smartphones and tablets by means of using their idling computing power.

Agent-Based Topology for Mobile Processing. Remote cloud provides agents, which are stable computers, in several local areas where mobile devices move. The devices do not access and use cloud services directly. Instead, they delegate the nearby agent to request and response in order to use cloud services more easily and stably. Cloudlet [9] is a computer that can be placed in common area like datacenter for connectivity purpose between mobile devices and cloud server. And it offloads workload of the mobile devices and alleviates the network problem of latency and bandwidth.

[5] `http://hbase.apache.org`

3.3 Service-Aware Design

What a client recognizes about a cloud service is depending on the group, the client involved in. In terms of access control, service user, service developer, and service administrator have different permission levels to access the cloud server with different view levels. The cloud service provider also has to provide different ways of services based on the user level.

Service-only. Mobile devices just use cloud services but do not recognize the existence of VMs where modules of the cloud services are running. So the mobile users only need to carefully deal with data.

VM-aware. Mobile devices can recognize each VM individually and operate it directly. When the mobile users want to configure it or to provide own services on the cloud, the users should access the VM by means of APIs, web pages, or terminal program with ssh.

VMM-aware. Cloud service provider grants authority to manage and operate many VMs to a few privileged users, so the users can operate functions of hypervisor such as managing lifecycle of VMs, monitoring of VMs, and backup. In this case, the user allowed to access via web pages or dedicated applications in order to use complicated functions.

3.4 Privacy-Sensitive Design

Even without any users' intentional input, sensing ability of mobile devices can make extensible information such as location, device state, and gesture. And mobile cloud can play a role as provider of context awareness by means of utilizing the sensing information from its mobile devices. For example, spatial context-awareness[6] would be composed of GPS data, 3G, time zone, configuration of its application, and existing map data in cloud. And context-awareness of mobile cloud is applied in augmented reality service, location based service, and crowdsourcing service.

Under the restricted situation, sensing data is usually personal and private. It makes MCC architecture need the systematic approach to protect the private data. A running application utilizes current location data to provide appropriate service to the mobile user, but it is also a matter that the user is concerned about the privacy issue. Many applications should notify the access to private local data but protection system is still needed.

Location Trusted Server (LTS) [10] proposes a systematic solution about privacy issue. LTS is located between mobile devices and location based services in cloud, and receives mobile devices' request and conceals users' information. Then LTS queries to the location based services only with general region information, and the result passes through LTS. Other researcher implements the similar algorithm which can cloaks user's information on the user's mobile device [11].

[6] http://en.wikipedia.org/wiki/Spatial_contextual_awareness

3.5 Analytical-Purpose Design

A multitude of mobile devices accumulate various forms and large volumes of data including personal sensing data, and MCC needs to analyze the data set in order to launch services that use it together with other data set.

Data Analysis on Mobile Devices. Mobile devices, the data collectors, participate in the analysis work as resource providers which use no external resource. It is essential to port common function for data analysis on the heterogeneous devices, but the devices need a virtual mobile cloud platform, which can support essential functions such as monitoring and managing of limited resources, and partitioning and offloading of workload.

A few research teams [1][12] adopt the MapReduce on Hadoop as a tool of data analysis and port it on their devices, but there is less consequent research. It is because the MapReduce requires a huge amount of resources and Hadoop is not designed for mobile devices by nature.

Data Analysis on High-Performance Cloud Server. Recently people are interested in big data, and many researches, which try to use Hadoop eco-system with powerful machines, are ongoing because MapReduce mechanism is a de facto standard for big data analysis in distributed environment. Because the MapReduce require a lot of computing and network bandwidth, it is rational that the mechanism should be running on cloud which have freedom of resource usage relatively and that mobile devices use the result.

- Apache Hadoop[7] is an open-source software library which permits for the distributed processing of huge data sets across clusters of computers.
- MapReduce[8] is a programming model which is associated implementation for processing and generating huge data sets.

The travel recommendation system [13] proposed by S.R. Yerva offers the predicted mood information of people on where and when users wish to travel based on social and sensor data in the cloud. This system fuses weather-dependent people's mood information from Twitter and meteorological sensor data streams.

4 MCC Design for Performance

From the observation of architectural focuses of MCC design strategies, we have studied that MCC services can be categorized according to the operational characteristics. To achieve the smart cloud computing for mobile environment, we extract the performance metrics to maximize the MCC architectural design benefit. It is necessary to consider the metrics in the design stage for effective MCC operations on various cases and supportive areas as in Table 2.

[7] http://hadoop.apache.org
[8] http://research.google.com/archive/mapreduce.html

Table 2. MCC Design Metrics

Metrics	Supportive Cases
Energy saving measurement	• GEMCloud [8] analyzes energy-efficiency of the entire system by using the idling computing power of mobile devices.
Personalized privacy level	• Location trusted server with cloaked region hides user's data when it queries requests to cloud service as a delegate [10].
VM provisioning management	• Amazon Web Service distinguishes authentication levels for accessing VMs.
Storage utilization on mobile device	• Dropbox and ucloud[9] provide data storage space and synchronize mobile user's file system with the other storage. • Google drive connects data files with google docs.
Communication frequency	• Cloudlet[9] offloads workload of the mobile devices and alleviates the network problem of latency and bandwidth.
Data availability and tracing for analysis	• For analysis and business intelligence of accumulated data, applications such as mobile millennium system [14], the travel recommendation system adopts Hadoop projects.

5 Conclusion and Future Work

In this paper, we observe the major characteristics of mobile cloud computing, such as resource limitation of mobile devices, unstable connectivity from wireless communication, and data type of personal usage. Based on those features, we proposed the architectural design strategies to build the taxonomy of MCC, and introduced its cloud services in the views of data-intensive architectural design, mobile processing-oriented design, service-aware design, privacy-sensitive design, and analytical-purpose design. This taxonomy gives the architectural design approach to make MCC services and applications. In this context, we derived the metrics which suggest the useful consideration for an effective MCC design.

In the near future, mobile devices will be more diverse and more ubiquitous via advancement of embedded system technology. With the advanced technology, personal cloud takes an important role to reflect user's context and generate mobile services via cloud computing from huge multiform data in various mobile clouds. In order to correspond with the diversity and new services, MCC research on security and trust should be pursued in the view of data integrity. Furthermore, in order to achieve interoperable and trustable architecture of various mobile clouds, research on common or referential platform of MCC should be considered.

[9] https://en.ucloudbiz.olleh.com/

Acknowledgment. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology. (Grant Number: 2011-0011507).

References

1. Marinelli, E.E.: Hyrax: Cloud Computing on Mobile Devices using MapReduce. Master Thesis, Carnegie Mellon University (2009)
2. Sanaei, Z., Abolfazli, S., Gani, A., Buyya, R.: Heterogeneity in Mobile Cloud Computing: Taxonomy and Open Challenges. Communications Surveys & Tutorials (99) (2013), doi:10.1109/SURV.2013.050113.00090
3. Smith, D.M., Plummer, D.C., Bittman, T.J., Bova, T., Basso, M., Lheureux, B.J., Prentice, B.: Predicts 2013: Cloud Computing Becomes an Integral Part of IT. Research Note Gartner. (2012)
4. Llamas, R.T., Stofega, W.: Worldwide Smartphone 2013-2017 Forecast and Analysis. International Data Corporation (2013)
5. Dinh, H.T., Lee, C., Niyato, D., Wang, P.: A Survey of Mobile Cloud Computing: Architecture, Applications, and Approaches. In: Wireless Communications and Mobile Computing, Wiley (2011), doi:10.1002/wcm.1203
6. Kim, M., Cui, Y., Han, S., Lee, H.: Towards Efficient Design and Implementation of a Hadoop-based Distributed Video Transcoding System in Cloud Computing Environment. International Journal of Multimedia and Ubiquitous Engineering 8(2), 213–224 (2013), SERSC
7. Lee, J., Hong, D.: Pervasive Forensic Analysis based on Mobile Cloud Computing. In: Proceeding of the 3rd International Conference on Multimedia Information Networking and Security, pp. 572–576 (2011)
8. Ba, H., Heinzelman, W., Janssen, C.A.: J Shi: Mobile Computing – A Green Computing Resource. In: Proceeding of IEEE Wireless Communication and Networking Conference, pp. 4474–4479. IEEE Communications Society, New York (2013)
9. Satyanarayanan, M., Bahl, P., Caceres, R., Davies, N.: The Case for VM-Based Cloudlets in Mobile Computing. Pervasive Computing 4(4), 14–23 (2009)
10. Zhangwei, H., Mingjun, X.: A Distributed Spatial Cloaking Protocol for Location Privacy. In: Proceedings of the 2nd International Conference on Networks Security Wireless Communications and Trusted Computing, vol. 2, pp. 468–471. IEEE Computer Society, Los Alamitos (2010)
11. Wang, S., Wang, X.S.: In-Device Spatial Cloaking for Mobile User Privacy Assisted by the Cloud. In: Proceedings of the 10th International Conference on Mobile Data Management, pp. 381–386. IEEE Computer Society, Washington DC (2010)
12. Huerta-Canepa, G., Lee, D.: A Virtual Cloud Computing Provider for Mobile Devices. In: Proceedings of the 1st ACM Workshop on Mobile Cloud Computing & Services: Social Networks and Beyond, vol. 6, pp. 5–9. ACM, New York
13. Yerva, S.R., Saltarin, J., Jeung, H., Aberer, K.: Social and Sensor Data Fusion in the Cloud. In: Proceedings of the 13th International Conference on Mobile Data Management, pp. 276–277. IEEE Computer Society, Washington DC (2012)
14. Hunter, T., Moldovan, T., Zaharia, M., Merzgui, S., Ma, J., Franklin, M.J., Abbeel, P., Bayen, A.M.: Scaling the mobile millennium system in the cloud. In: Proceeding of the 2nd ACM Symposium on Cloud Computing, ACM, New York (2011), doi: 10.1145/2038916.2038944.

Analysis of Discriminant Features in Fourier Domain Compensating Shadow Areas on Facial Images

Phuc Truong Huu[1], Sang-Il Choi[2], Sang-Hoon Ji[1], Hong-Seok Kim[1], and Gu-Min Jeong[3]

[1] Korea Institute of Industrial Technology, Korea
{whitelion_pc,robot91,hskim}@kitech.re.kr
[2] Dankook University, Korea
choisi@dankook.ac.kr
[3] Kookmin University, Korea
gm1004@kookmin.ac.kr

Abstract. This paper proposes a novel compensation method for shadow areas on the human face by analyzing magnitude components of the facial images on the Fourier domain. A feature extraction algorithm based on PCA+LDA is utilized to extract features of the magnitude components, and create a compensation handling mask which identifies and catalog the necessary compensation levels of darkness pixels. The proposed algorithm is applied to the facial data for the face recognition. The experimental results demonstrate that the proposed algorithm can effectively compensate for the dark areas in the human image as well as improve the accuracy of face recognition.

Keywords: Face Recognition, Feature Extraction, Fourier Transform.

1 Introduction

Illumination is a regular challenge in the face recognition discipline because of its variation. The changes induced by illumination, such as cast shadows or attached shadows, can be larger than the innate differences between individuals. There are numerous methods proposed to deal with it, however, they have not yet solved the illumination completely. In this paper, we suggest a new method for shadow compensation in the facial images to improve the performance of the face recognition system. The proposed compensation method deals with the shadow variation by transforming the image to the Fourier domain, then, analyze to find dark areas and efficient modification factors.

The rest of this paper is organized as follows. Section 2 and section 3 explain the procedure of the algorithm to identify and compensate dark pixels on facial images. Section 4 consists of the experiment results, followed by the conclusion in Section 5.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_67, © Springer-Verlag Berlin Heidelberg 2014

2 An Overall of Discriminant Features in Fourier Domain Compensating Shadow

The proposed algorithm compensates shadow areas on images based on a feature based multilevel mask. This mask is created by analyzing the ratio of each element of the features that are extracted from the magnitude component of the image. These features are transformed into a unit space to calculate the main projector vector, and to compare the importance on each unit vector component to the main projector vector. The vectors, which have a high importance in the unit space, correspond to the important pixels of the image in the spatial space. Therefore, brightening these pixels in the spatial space is equivalent to increasing the magnitude of these vectors in the Fourier space.

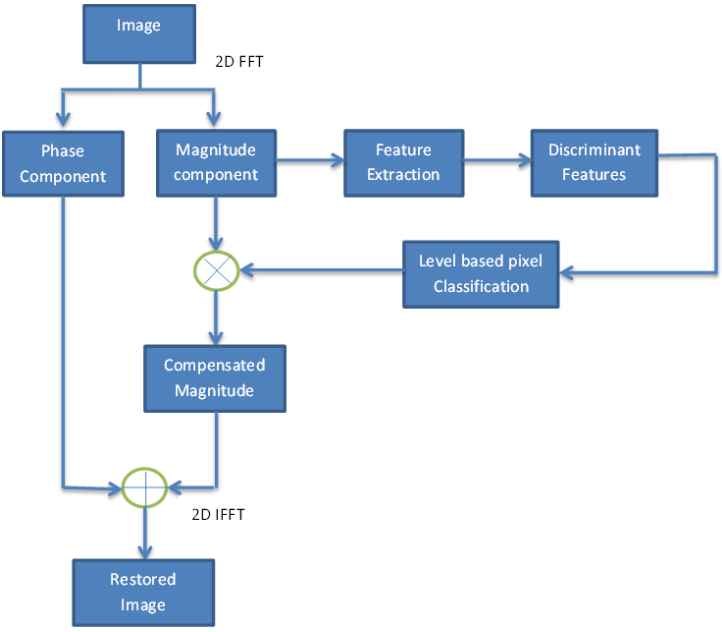


Fig. 1. The proposed shadow compensating algorithm for facial images

3 Procedure of the Proposed Shadow Compensation Algorithm

Step 1: Calculate Fourier Transformation value of the training set.

The image in the spatial domain can be represented in the Fourier domain using 2D Fourier transformation. Suppose that the intensity of the gray-scale image with $M \times N$ pixels be $I(x, y) \in \mathbb{R}^{M \times N}$, the transformation of $I(x, y)$ is given by

$$F(u,v)=\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}I(x,y)e^{-j2\pi(ux/M+vy/N)}, \tag{1}$$

with $u=0,1,...,M-1$ and $v=0,1,...,N-1$.

There are 2 parts composed from each image through this transformation: magnitude $\|F(u,v)\|$ and phase components $\phi(u,v)$. Oppenheim et al. in [5] shows that whereas the magnitude component is much affected by the shadow, the phase one in the frequency domain, which contains the structural information of the image, is less prone to the effects of illumination variations. Therefore, the phase component will be keep for the reconstruction step without any change. The magnitude component is used to improve the quality of illumination on the image by extracting and analyzing important features of this component.

Step 2: Find a correlation matrix from features.

First, we apply PCA to the magnitude component set of training images set to get the general vectors of the training set. Then, a LDA algorithm is applied to these vectors to get important features from magnitude set. Suppose that we have the features obtained from LDA as follows.

$$\begin{aligned} f_1&=\{a_{11}e_1,a_{12}e_2,...,a_{1n}e_n\}\\ f_2&=\{a_{21}e_1,a_{22}e_2,...,a_{2n}e_n\}\\ f_k&=\{a_{k1}e_1,a_{k2}e_2,...,a_{kn}e_n\}\end{aligned} \tag{2}$$

The main projection vector is created by calculating the mean value of these feature vectors.

$$f_m=\sum_{i=1}^{k}f_i=\{a_{m1}e_1,a_{m2}e_2,...,a_{mn}e_n\} \tag{3}$$

The correlation matrices c_i are estimated by indexing the comparison result of each component of each vector f_i with each corresponding one of the main projection vector f_m. For example, if we have two vectors

$$f_1=\{a_{11},a_{12},...,a_{1l},a_{1(l+1)},...,a_{1n}\} \tag{4}$$

and $f_m=\{a_{m1},a_{m2},...,a_{ml},a_{m(l+1)},...,a_{mn}\}$,

suppose that

$$\begin{cases}a_{11}>a_{m1},a_{12}>a_{m2},...,a_{1k}>a_{ml}\\ a_{1l}<a_{ml},a_{1(l+1)}<a_{m(l+1)},...,a_{1n}<a_{mn}\end{cases}\quad(1\le l\le n), \tag{5}$$

the correlation matrix c_1 will be

$$\begin{aligned} f_1 &= \{a_{11},\ a_{12}\ ,\ldots,\ a_{1l}\ \ ,a_{1(l+1)}\ ,\ldots,\ a_{1n}\} \\ f_m &= \{a_{m1},\ a_{m2},\ldots,\ a_{ml}\ ,\ a_{m(l+1)},\ldots,\ a_{mn}\} \\ c_1 &= \{1\ \ ,\ 1\ \ ,\ \ldots,\ 1\ \ ,\ 0\ \ \ \ ,\ldots,\ 0\} \end{aligned} \tag{6}$$

The final correlation mask is a summation of all these correlation matrices c_1.

$$c = \sum_{i=1}^{n} c_i \tag{7}$$

Step 3: Create a multi-level mask.

We use 2 threshold levels to evaluate the dark level of a pixel through analyzing its magnitude component in the Fourier domain. This means that a pixel is catalogued to 3 levels: acceptable, low and very low brightness. The level definition of all pixels of a face image is stored in a matrix m, which is called the multi-level mask. The multi-level mask m is constructed by an analysis on the space formed by the unit vectors $(e_1, e_2, ..., e_n)$.

$$\mathrm{m}_j = \begin{cases} 0 & \mathrm{m'}_j < threshold_1, (j = \overline{1,n}) \\ 1 & threshold_1 < \mathrm{m'}_j < threshold_2 \\ 2 & \mathrm{m'}_j > threshold_2 \end{cases} \tag{8}$$

Step 4: Recovering the image in the spatial domain.

After cataloging for each pixel, we compensate the image's shadow areas and recover the image in the spatial domain. The compensation process is implemented by multiplying the magnitude component of each pixel in the Fourier domain with the pre-defined factors. The factor of each pixel is determined by the corresponding value in the multi-level matrix m.

$$t(m_j) = \begin{cases} 1 & m_j = 0 \\ 1.25 & m_j = 1 \\ 1.5 & m_j = 2 \end{cases} \qquad (j = \overline{1,n}) \tag{9}$$

The image recovery is conducted by applying an inverse Fourier transform on the magnitude component and the original phase component.

$$I(x,y) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} F(u,v) e^{j2\pi(ux/M + vy/N)}. \tag{10}$$

4 Experimental Results

The main purpose of the shadow areas compensation in this paper is to improve the result of face recognition. We, therefore, evaluate the quality of compensation algorithm by estimate the final result of face recognition on a database. Specifically,

we applied the proposed method on the Yale database [3] with $threshold_1 = 0.4$ and $threshold_2 = 0.6$. To evaluate the effect of the proposed method, we compare its result with one of the mean based method [2]. In this method, compensating for the magnitude component is conducted by adding a predetermined shadow adjusting part into each testing image. This shadow adjusting part is composed by obtaining the average magnitude of all images in the training set. Figure 2 shows a comparison of the proposed method with the compensation based on the mean value of magnitude components in term of classification rate.

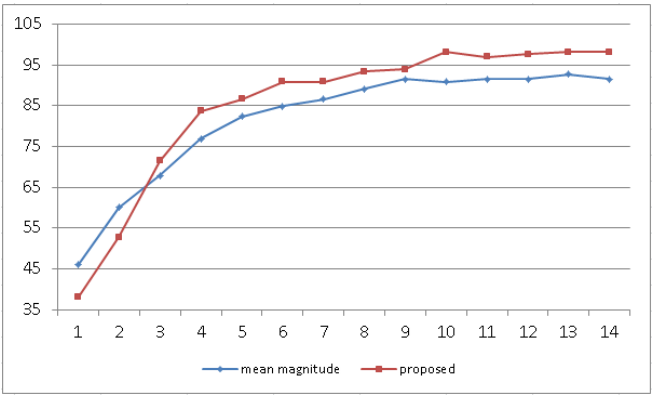


Fig. 2. Results of the proposed and mean based compensation method

It is clear that the proposed method can provide a better result of face recognition compare to the mean based compensation method. Specifically, when applied to the Yale Database, the average correction rate of the proposed method is 85.07%, whereas, the mean based compensation algorithm provides 81.69% accurate.

5 Conclusion

In this paper, we have proposed a novel method to compensate dark areas on the facial images. The method transforms images in the spatial domain to the Fourier domain using 2D Fourier transformation. The phase component that contains structural information and is less effected by illumination variation is kept to reconstruct images, whereas, the magnitude component is utilized to analyze and compensate shadows for the image. The experimental results demonstrated that the proposed technique provided a better solution compare to the mean based compensation method when applied to the Yale Database.

Acknowledgement. This work is supported by the Ministry of Trade, Industry and Energy (MOTIE), South Korea.

References

1. Jeong, G.-M., Ahn, H.-S., Choi, S.-I., Kwak, N.-J., Moon, C.: Pattern recognition using feature feedback: Application to face recognition. International Journal of Control, Automation, and Systems 8, 141–148 (2010)
2. Choi, S.-I., Kim, C., Choi, C.-H.: Shadow compensation in 2D images for face recognition. Int. Pattern Recognition 40(7), 2118–2125 (2007)
3. Bellhumer, P.N., Hespanha, J., Kriegman, D.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transaction on Pattern Analysis and Machine Intelligence, Special Issue on Face Recognition, 711–720 (1997)
4. Choi, S.-I., Choi, C.-H., Jeong, G.-M.: Pixel selection in a face image based on discriminant features for face recognition. In: IEEE International Conference Automatic Face and Gesture Recognition (2008)
5. Oppenheim, A.V., Willsky, A.S., Nawab, S.H.: Signal and Systems, 2nd edn., pp. 303–304. Prentice-Hall, Upper Saddle River (1996)

An Efficient Routing Scheme Based on Social Relations in Delay-Tolerant Networks

Chan-Myung Kim[1], In-Seok Kang[1], Youn-Hee Han[1,*], and Young-Sik Jeong[2]

[1] Advanced Technology Research Center,
Korea University of Technology and Education, Korea
{cmdr,iseka,yhhan}@koreatech.ac.kr
[2] Dongguk University, Korea
ysjeong@dongguk.edu

Abstract. In delay-tolerant network (DTN), the message forwarding and routing are important research issues, since the network topology changes dynamically and there is no guarantee of continuous connectivity between any two nodes. In this paper, we propose an efficient DTN routing scheme by using a node's social relation where each node chooses a proper relay node based on its contact history. In order to enhance the routing efficiency, the expanded ego-network betweenness centrality is used when a relay node is selected. We have demonstrated that our algorithm performs efficiently compared to the existing epidemic and friendship routing schemes.

Keywords: Delay-Tolerant Networks, Social Networks, Routing, Expanded Ego Network.

1 Introduction

A delay-tolerant network [1] is a network architecture designed to operate effectively as an overlay on top of regional networks, or as an interplanetary internet. DTNs promises to enable communication between challenged networks, which includes deep space networks, sensor networks, mobile ad-hoc networks. Network topology change dynamically, and the lack of end-to-end connectivity poses a number of challenges in routing in DTNs. A message is delivered through hop-by-hop communication and the message delivery probability is unpredictable. Designing efficient routing protocols in DTNs can tackle various issues arising due to lack of continuous network connectivity.

Utilizing social network characteristic has been recently studied in designing efficient routing protocols. Many studies have shown that nodes tend to have mobility patterns influenced by their social relationships and/or social behavior [1,3]. By examining the social network of the DTN nodes, it may be possible to optimize data routing by forwarding data to nodes that are much socially related.

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_68,

In social based DTNs, nodes encounter other nodes and store contact information (e.g. when they met and when they are separated) to their buffers. Thus, many studies have examined contact information when they analyze social relationships of nodes. E. Bulut et al. introduced a new metric to detect the quality of friendships of each nodes accurately [7]. They calculate the link weight based on contact information and use it when constructing the friendship community where the set of nodes have close friendship between each other. They also presented a new sociality-based routing scheme, called *Friendship Routing*, which utilizes the link weight to make the forwarding decisions of messages.

In this paper, we propose an efficient DTN routing scheme where each node chooses a proper relay node based on the contact history. In order to enhance the routing efficiency, the expanded ego-network betweenness centrality [8]. The expanded ego betweenness centrality can be calculated locally in a node and used to identify a bridge node within the network. We have demonstrated that our algorithm performs efficiently compared to the existing epidemic and friendship routing schemes.

The rest of this paper is organized as follows. Section 2 explains how to construct a social network in each node and how to get the expanded ego betweenness centrality, and Section 3 presents the proposed routing scheme. Section 4 shows a simulation analysis, and Section 5 finally concludes this paper.

2 Local Information-Based Social Network Construction

In this paper, we consider a network constituted by nodes with mobility, so the network topology changes dynamically. Let us assume that r transmission range of a node. We assume that when a node i ends a message to any node within a distance out any failure. A node i can *encounter* another node j when the node i comes close to the node j and receive a first hello message broadcasted by the node j. If the node i stays within the transmission range of the node j, the node i can hear a periodic hello message from the node j. When the node i does not hear a predefined number of the node j's hello messages continuously, the node i considers that it leaves the node j. When the node i meets the node j at a time α and *leaves* the node j at a time β, we define $\beta - \alpha$ as the node i's *contact duration* for the node j.

2.1 Social Network Construction Based on Contact History

We assume that each node records the contact duration information per node which it has encounters. Each node allocates *contact window* in its buffer. Its size is called *contact window size* W_s which is pre-determined by a node. A node keeps the contact duration information for a node which it has encountered within the range of W_s. As time goes on, the contact window slides by one unit time to keep the recorded contact information fresh.

With these information, a node i constructs its social network SN_i, where each vertex corresponds to nodes which the node i has encountered frequently and each

edge corresponds to the relation between the node i and the frequently-encountered nodes. Between the nodes i and j, the following weight $w_{i,j}$ is allocated:

$$w_{i,j} = \frac{W_s}{\int_{t=0}^{W_s} f(t)dt} \tag{1}$$

where f(t) returns the remaining time to the first encounter of the node j after time t. For example, let us assume that the contact window size W_s is 10. If the node i is in contact with the node j in the first 5 seconds, then separated away for 3 seconds and contacts the node j for 2 seconds again, the weight $w_{i,j}$ of two nodes is $10/(3+2+1) = 5/3$ [7].

If the weight $w_{i,j}$ is larger than the predefined threshold T_h, an edge between the nodes i and j is created between the two nodes in the node i's social network SN_i,. And then, the node i also includes the node j's social network SN_j, (Friends of Friends) to its SN_i,. The SN_j, is transferred to the node i through the node j's hello message. The algorithm 1 represents the social network construction procedure for a node i where V_i and E_i means set of nodes and edges in SN_i, and N_i means set of 1-hop neighbor nodes of node i. Note that the algorithm is performed every unit time.

Algorithm 1 Social network construction procedure for a node i

```
Slide its contact window by a unit time
for a node j which the node i has at least one contact with do
    if w_{i,j} > T_h then
        V_i ∪ {j}, E_i ∪ {(i, j)}
        if N_j ≠ ∅ then
            V_i = V_i ∪ N_j
            E_i = E_i ∪ {(j, k)} for k ∈ N_j
        end if
    else
        if j ∈ V_i then
            V_i = V_i − {j} − (N_j − N_i)
            E_i = E_i − {(i, j)}, E_i = E_i − {(j, k)} for k ∈ N_j
        end if
    end if
end for
```

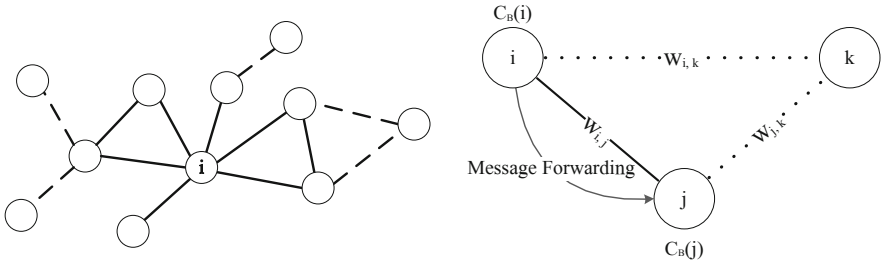
2.2 Expanded Ego Betweenness Centrality

The betweenness centrality is used as important measure to examines the extent to which a node is between all other nodes within the network [2]. When the message is forwarded to nodes with high betweenness centrality, that message can be disseminated to entire network in fast way. In this paper, we use the betweenness centrality to increase overall routing efficiency. For an arbitrary node i in network, the equation of betweenness centrality $C_{B(i)}$ is defined as follows:

$$C_{B(i)} = \sum_{s \neq i \neq t \in V, s<t} \frac{\rho_{st}(i)}{\rho_{st}} \tag{2}$$

where V is the set of nodes in the network and n is the total number of nodes, ρ_{st} the number of shortest paths between the node s and t, and $\rho_{st}(i)$ the number of those shortest paths that include the node i. The betweenness centrality requires the entire network information but a node cannot know it due to lack of the whole network-wide end-to-end connectivity. Therefore, it is hard to utilize the betweenness centrality in DTNs.

As we saw in previous section, a node constructs its social network with its friends and friends of friends information. Hence, we use the expanded ego betweenness which is calculated only with the local information [8].



(a) Expanded ego network (b) $w_{j,k} > w_{i,k}$ or $C_B(j) > C_B(i)$

Fig. 1. Expanded ego network and message forwarding strategy

Fig. 1(a) illustrates an expanded ego network of a node i. The network is constituted by the ego (the node i), its 1-hop neighbors, and its 2-hop neighbors. While the solid links presents the *ego network* introduced in literature [5,6], the solid and dashed links represents the *expanded ego network* which is constructed by the algorithm 1. In short, the expanded ego betweenness centrality of the node i is equal to the betweenness centrality of the node i in its expanded ego network. In our previous work [8], we verified that the expanded ego betweenness centrality is highly correlated with the betweenness centrality in the complete network. For the details of the expanded ego betweenness centrality, refer to [8].

3 Routing Strategy

In our algorithm, a node 1) calculates the weight based on the contact history recorded for other nodes it has encountered, 2) constructs its own social network and 3) calculates the expanded ego betweenness centrality. With these information, a node makes a decision on a relay node when it tries to send a message to the destination node.

3.1 Link Weight Based Strategy

Fig. 1 (b) depicts a situation where a node i tries to send a message to the remote destination node k and it just contacts to a node j. The node i has to make a decision that it should forward (i.e., copy) the message to the node j. Basically, at this time, the node i considers the two link weight values $w_{i,k}$ and $w_{j,k}$. A high link weight between a pair of two nodes represents that they are friendly and the future contact opportunity comes high. In our scheme, the node i forwards the message to the node j if the following condition is met.

$$\textit{Condition}\ \text{I:}\ \ w_{j,k} > w_{i,k}$$

This strategy is similar to the one proposed by [7].

3.2 Expanded Ego Betweenness Centrality Based Strategy

On the other hand, we use the expanded ego betweenness centrality to increase message delivery efficiency. In DTNs, some nodes hardly meet other nodes and they have very low link weights for the previous encounters. If one of these isolated nodes is set to be the destination node, source node might not find a proper relay node to deliver a message, when it would forward the message only by using the link weight based strategy. In such situation, the message destined to the isolated node could be removed from the sender's buffer by TTL (time-to-live) expiration before delivered to the destination.

To prevent this situation, a node i forwards a message to a node j if $C_B(j)$ is larger than $C_B(i)$ even though *Condition* 1 is not met. That is, the following is the second condition for the message forwarding.

$$\textit{Condition}\ \text{II:}\ \ C_B(j) > C_B(i)$$

A high expanded ego betweenness centrality of a node j represents that it is active and socially related with many other nodes. Hence, the message forwarding to it makes high the opportunity that the message will reach the destination node, and thus such forwarding strategy can increase the overall efficiency of message delivery.

3.3 Message Delivery Cost Reduction

In this section, we propose a message management scheme in a node's buffer to decrease the overall delivery cost. If a node i has a message in its buffer destined to the node k and encounters a node j which satisfies $w_{j,k} > w_{i,k}$ (Condition I), the node i will forward the message to the node j. In this case, if the link weight $w_{j,k}$ is the largest among the link weights between other nodes in SN_i and the node k, the node i deletes the message from its buffer after forwarding the message to the node j to prevent further dissemination. Algorithm 2 represents our overall routing strategy.

Algorithm 2 Message forwarding procedure when a node i tries to deliver the message to the destination node k

```
upon reception of a Hello message from a node j do
if j ∈ SN_i then
    if w_{j,k} > w_{i,k} then
        forward the message destined for the node k to the node j
        if w_{j,k} > w_{m,k} for all m ∈ SN_i then
            delete the message destined for the node k from the node i's buffer
        end if
    else if C_B(j) > C_B(i) then
        forward the message destined for the node k to the node j
    end if
end if
```

4 Performance Evaluation

In this chapter, we demonstrate our simulation results and compare the proposed scheme with the epidemic [4] and the friendship [7] routing schemes. We use three metrics to evaluate our scheme as follows: 1) message delivery ratio, 2) message delivery cost and 3) message delivery efficiency. The delivery ratio is the proportion of messages delivered to their destinations among the total messages generated. The delivery cost is the average number of forwards done during the simulation. Finally, delivery efficiency is defined as the ratio of delivery ratio to the delivery cost.

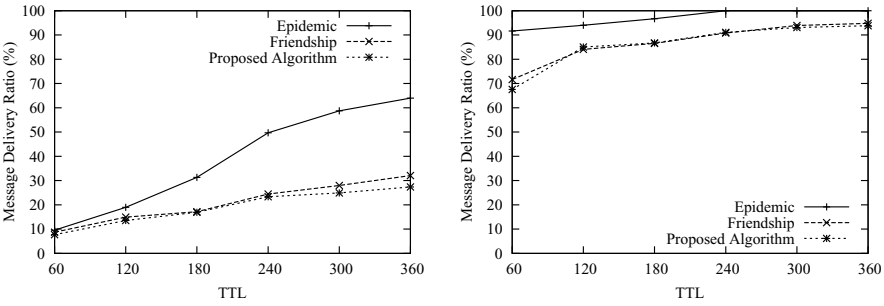
To evaluate our scheme, we used real trace-driven simulations based on pre-defined node mobility data. From the mobility data, we generated the contact information logged during the simulation time. After W_s period of time, we generated 1000 messages, each from a node to a random destination node. Each message has a certain TTL value and is removed after the TTL expiration. The simulation ended when the 1000 messages are delivered to the destination or expired. All results are averaged over 10 runs. Table 1 summarizes the simulation parameters.

Table 1. Simulation Environment

Area Size (m)	1000 × 1500
Number of Nodes	25, 75
Communication Range (m)	3
Moving Speed (m/s)	0.5, 1.0, 1.25, 1.5
Contact Window Size (W_s, second)	600
TTL (second)	60, 120, 180, 240, 300, 360
Threshold (Th)	0.01
Number of Messages	1000

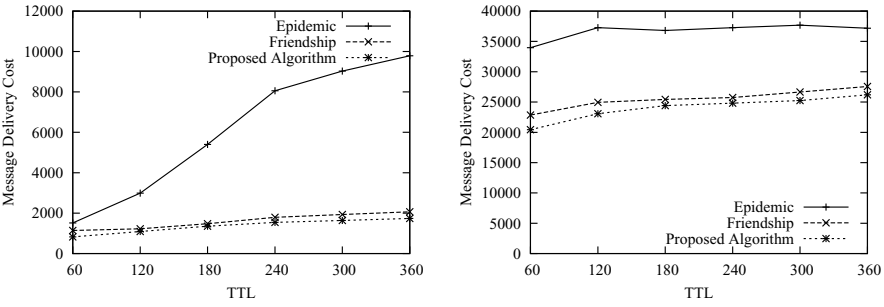
Fig. 2 shows the delivery ratio achieved by each schemes. As the TTL increases, all schemes deliver more messages to the destinations. As expected, the epidemic routing scheme has the high packet delivery ratio. It is noted that the performance of proposed scheme is almost similar to the friendship routing scheme. Fig. 3 shows the

message delivery cost of each scheme. We can see that the epidemic scheme has the worst performance for the delivery cost because it uses the flooding strategy basically. It should be noted that the proposed scheme is more cost effective than the friendship



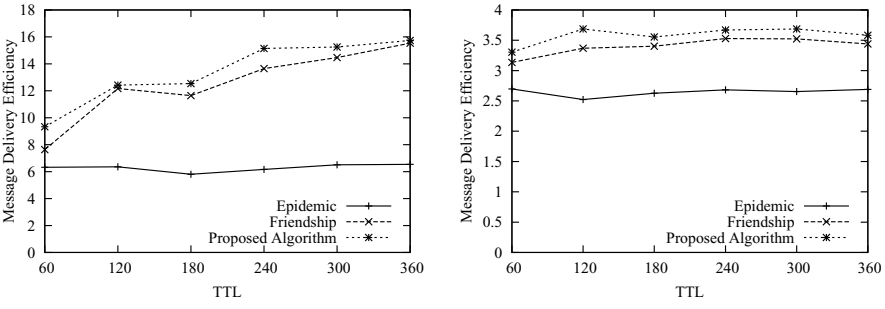
(a) N=25 (b) N=75

Fig. 2. Delivery Ratio (%)



(a) N=25 (b) N=75

Fig. 3. Delivery Cost



(a) N=25 (b) N=75

Fig. 4. Delivery Efficiency

scheme. It is because our routing scheme uses the expanded ego betweenness centrality based strategy as well as the basic weight based strategy, and it also removes the message from a node's buffer to prevent imprudent dissemination.

Finally, Fig. 4 shows the message delivery efficiency achieved by each schemes. As can be seen by the figure, the routing efficiency achieved by the proposed scheme is higher than the ones of other schemes. It means that our scheme has the benefit of cost effective routing with a little performance degradation of delivery ratio.

5 Conclusion

In this paper, we introduced a cost effective routing scheme that uses the contact information and the expanded ego betweenness centrality information. We simulated our scheme and compared the its performance with the existing epidemic and friendship schemes's ones. We have shown our scheme achieves higher delivery efficiency than the existing schemes. Since most of nodes in DTNs are energy constrained, we plan to further examine energy-efficient routing in DTNs.

Acknowledgments. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2013R1A1A2010050), and also financially supported by the Ministry of Knowledge Economy (MKE) and Korea Institute for Advancement of Technology (KIAT) through the Workforce Development Program in Strategic Technology.

References

1. Warthman, F.: Delay-Tolerant Networks(DTNs): A tutorial, Warthman Associates (March 2003)
2. Freeman, L.: Centrality in social networks: Conceptual clarification. Social Networks 1, 215–239 (1979)
3. Schurgot, M., Comaniciu, C., Jaffres-Runser, K.: Beyond Traditional DTN Routing: Social Networks for Opportunistic Communication. Accepted for Publication in IEEE Communications Magazine, 155–162 (2012)
4. Vahdat, A., Becker, D.: Epidemic Routing for Partially Connected Ad Hoc Networks. Technical Report CS-200006, Duke University (2000)
5. Everett, M., Borgatti, S.P.: Ego Network Betweenness. Social Networks 27(1), 31–38 (2005)
6. Odella, F.: Using Ego-networks in Surveys: Methodological and Research Issues. In: Proceedings of International Conference on Network Science (2006)
7. Bulut, E., Szymanski, B.K.: Friendship based routing in delay tolerant mobile social networks. In: Proceedings of IEEE Global Telecommunications Conference (GLOBECOM) (2010)
8. Kim, Y.-H., Kim, C.-M., Han, Y.-H., Jeong, Y.-S., Park, D.-S.: Betweenness of Expanded Ego Networks in Sociality-Aware Delay Tolerant Networks. In: Han, Y.-H., Park, D.-S., Jia, W., Yeo, S.-S. (eds.) Ubiquitous Information Technologies and Applications. LNEE, vol. 214, pp. 499–505. Springer, Heidelberg (2012)

A Data Aggregation Based Efficient Clustering Scheme in Underwater Wireless Sensor Networks

Khoa Thi-Minh Tran and Seung-Hyun Oh*

Department of Computer Science, Dongguk University, South Korea
{ttmk84,shoh}@dongguk.ac.kr

Abstract. The restricted underwater wireless sensor networks (UWSNs) such as large propagation delay, low bandwidth capacity, high bit error rates, mobility, limited memory as well as battery and so on pose many challenges for scientists doing UWSN construction. In this paper, we take into account of proposing a promised clustering scheme that can overcome the UWSN's confined. Our proposed data aggregation based clustering scheme involves 4 phases: initial phase, cluster head election phase, clustering phase, and data aggregation phase. The main goals of our proposed include reducing the energy consumed of the overall network, increasing the throughput, and minimizing data redundancy while still guarantying data accuracy.

Keywords: Cluster, Data Aggregation with Similarity Function, Underwater Wireless Sensor Networks.

1 Introduction

Recently, aquatic environment has received much attention of many scientists because of its potential information and resources. The interests aim to monitor underwater environment for various applications such as oceanographic data collection, disaster prevention, undersea exploration, surveillance applications, etc. [1, 2]. However, communication in underwater wireless sensor networks (UWSNs) faces many challenges due to the differences from the communication in terrestrial wireless sensor networks. For example, the large propagation delays is about 1.5×10^3 m/s, the sensor nodes in UWSNs move with water current, the low bandwidth capacity of the order of KHz results in high bit error rates, limited and difficulty in recharge of battery power, and so on [3, 4, 5]. Thus, how to design an UWSN that saves the energy consumption and prolong the network lifetime has become a major concern.

Cluster structure makes a network look smaller and more stable in the view of each mobile terminal [6]. Moreover, cluster structure is a promising method to reduce the network consumption that has recently received much attention while deploying a network in aquatic environment. Cluster-base concept divides the network into groups of nodes (or clusters), defines mechanism by which all clusters connect to each

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*, Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_69,

other's [4, 7, 8]. There are various ways doing research on network clustering such as research on how to optimize the cluster size [9], how to select a cluster head [10], how to communicate among nodes and among clusters [7, 9], how to aggregate data in cluster [11, 12, 13], and so on.

Data aggregation is the technique which attempts to collect the sensed data from the sensors and transmit to the base station (BS) or the sink. The main role of data aggregation is not only to eliminate the redundancy data receive from sensor nodes, but also reduce the number of transmissions to the BS/sink. The aggregation of data using similarity functions can minimize data redundancy and reduce data packet size to be sent to the BS/sink [11]. Subsequently, it reduces the traffic load and prolongs the network lifetime.

In this paper, we consider all layers in forming a network. Also, we give an idea of a clustering scheme in which the cluster head can be reelected, and clusters can be reconstructed due to the change of conditions such as energy consumption, networks movement, etc. Re-clustering not only retains the best cluster structure, but also prolongs the network life time. The combination between clustering and data aggregation with similarity function is the best way to reduce the overall network consumption, increase network throughput, and achieve data accuracy. Through the simulation results, we show that our proposed data aggregation base clustering scheme can achieve a better throughput and energy consumption then clustering without data aggregation.

The paper is structured as follows. Section 2 reviews related research about network clustering, cluster head selection method, and data aggregation in cluster based networks. Section 3 describes in details the proposed clustering scheme. Section 4 shows our simulations and results. Finally, Section 5 concludes the paper.

2 Related Research

Paper [7] proposed a dependable clustering protocol to provide a survivable cluster hierarchy against cluster-head failures in such networks. In the paper, the clustering protocol tries to select a primary cluster head and a backup cluster head for each cluster member during clustering. Authors believe that the cluster member can quickly switch over to the backup cluster head when its primary cluster head is not working by any reason.

The objective of paper [8] is to propose a data aggregation algorithm that achieves the energy saving, increasing network lifetime, and reducing the amount of bandwidth utilized. In this paper, authors considered on forming clusters, electing cluster heads, and applying averaging technique in data aggregation. Also, they did the comparison between network with aggregation and without aggregation.

Authors in paper [10] proposed an energy efficient cluster head selection scheme by considering the node's energy and distributed position. Under analyzing the energy consumption of LEACH protocol in underwater channel: the energy of distant nodes will be exhausted early, the energy efficiency will decrease due to the cluster head is

randomly selected. However, the sensor network is assumed static network which rarely happens in underwater environments.

In paper [11], authors gave an idea to use similarity function for data aggregation in cluster-based UWSNs. They proved through the results that similarity function, especially Euclidean distance and cosine distance, can achieve an efficiency underwater network by reducing the packet size and minimizing the data redundancy. However; in this paper, the authors assumed the network is already clustered and focus only on applying the similarity functions to the cluster heads or aggregators.

3 A Data Aggregation Based Efficient Clustering Scheme in UWSNs

A cluster based network is a network which partitioned into non-overlapping clusters. Each cluster consists of one cluster head and several cluster members. Cluster members eventually sense surrounding environment, then transmit information to its cluster head. The main role of cluster heads are to collect sensed data from its member nodes, aggregate collected data, and transmit aggregated data to sink node or BS.

For easy reference, all notations are described in Table 1.

Table 1. Description of notations

Notation	Description
t_{round}	The time when a round starts until it finishes
N	Number of sensor nodes in the network
tx_sink	Transmission range of sink nodes/BS
ps	Propagation speed of signal in shallow underwater environment (1500 m/s)
$t_{clustering}$	The time of clustering process
tx_max	The maximum transmission range
$E_{residual}$	The residual energy of the node
$d_{(nodeId1,nodeId2)}$	The distance from between two sensor nodes
$t_{receive}$	The time when a node receive a message

The proposed clustering scheme works with rounds – a round finishes when t_{round} expires. The t_{round} is the time need for clustering, and transmitting data of sensor nodes to sink/BS nodes, defined as (1).

$$t_{round} = N \times \frac{tx\_sink}{ps} + t_{clustering} \quad (1)$$

Each round consists of 4 phases: initial phase, cluster head selection phase, clustering phase, and data aggregation phase.

- *Initial phase*: The sink nodes broadcast request messages with a timestamp, tx_max, and t_{round}. The tx_max is a random value between 100 m and 200 m.

Sensor nodes which receive the request message use the timestamp and t_{round} to know when a round will be finished. From the time when a new round start, the sink node will re-broadcast request messages with different tx_max. In the initial phase, only sink nodes are allowed to work for avoiding waste of energy.

— *Cluster head selection phase:* All sensor nodes have received request message from the sink nodes will set up the tx_max. Then, they broadcast hello message. The hello message contains Eresidual, d(nodeId,sinkId), and a timestamp. With the hello message, sensor nodes they can elect themself to become a cluster head by knowing of others' information. The conditions to become cluster head are based on the highest energy and the closest distance to its sink, showed in (2).

$$\begin{cases} Max(E_{residual}) \\ Min(d_{(nodeId,sinkId)}) \end{cases} \tag{2}$$

The below pseudo code explains how our cluster head selection works:

```
A node receives a msg
{
    If the received msg is Hello msg
    {
       Create a priority table to store Hello msg's
       information.
       Compare among nodes' information.
       If a node with highest residual energy and
       closest to the sink node
       {
          Node elects itself to become CH by
          broadcasting Invite msg to others.
       }
       Else
       {
          Wait for Invite msg from CH node.
       }
    }
}
```

One disadvantage of broadcasting technique is the collision caused by the travelling of many messages in the network. Hence, we apply a random timer to each sensor node for delay the broadcasting.

— *Clustering phase*: The cluster head sends invite message to all its neighbor nodes within the transmission range. The invite message contents of the cluster head identification, a timestamp, and neighbor node identification. Note that a cluster head knows information of its neighbor nodes through the cluster head selection phase. A sensor node may receive more than one invite message from different cluster head. Hence, the sensor node will measure the distance to each

cluster head. Then, it will accept the invitation from the cluster head with a closer distance. The distance from a node to the cluster head is calculated by (3).

$$d_{(nodeId,CH_{Id})} = ps \times (t_{receive} - timestamp) \quad (3)$$

- *Data aggregation phase*: Since all clusters are formed within a round. Communication types such as intra-communication (cluster head – cluster members, cluster member – cluster member within a cluster), inter-communications (cluster head – sink, cluster member – cluster members of two different clusters) are set up. The cluster heads invoke data aggregation mechanism to aggregate and transmit data to their own sink. Additionally, sensor nodes can sense similar data and transmit redundant information to the cluster head because they monitor the environment most of the time. This increases the number of transmissions to the sink as well as increases the energy consumption, reduces the network life time. In this phase, hence, we apply data aggregation technique with similarity function to cluster heads [11] in order to eliminate transmitting of redundant data to the sink/BS. Means, a cluster head collects the raw data from its cluster members and performs similarity comparison for each pair of data from different member nodes.

As soon as an old round finishes and a new round starts, cluster heads are re-selected and clusters are re-constructed according to the state of the network such as energy residual of all sensor nodes, and the network movement. Also, the communication among nodes, clusters are set up again due to the new construct of the network. We consider applying TDMA for intra-cluster communication and CDMA for inter-cluster communication. The communication between cluster head and sink/BS is considered using handshake approach for the reliable data transmitting.

4 Simulations and Results

The performance of the proposed scheme is evaluated using simulation. We have run our simulations with QualNet5 simulator. Dimension of the scenario is 1500 m × 1500 m. The scenario consists of 50 sensor nodes and 4 sink nodes deployed 200 m below the sea level. The sensor nodes are deployed randomly while the sink nodes are equidistant to another. In order to facsimile the real shallow underwater environment, the channel frequency and propagation speed are set 35 KHz and 1500 m/s, respectively. The energy consumption parameters are set according to the UWM100 LinkQuest Underwater Acoustic Modem [14]. The communications between cluster head and sink/BS using cooperative MAC scheduling scheme [15] – a handshake approach have been proved for the reliable data transmitting. Besides, intra-cluster communications and inter-cluster communication are set with TDMA and CDMA, respectively. All sensor nodes operated with the data rate of LinkQuest UWM100 equals 7 Kbps. The transmission range power of sink nodes are 30 dBm (around 544 m) while the transmission range of common sensor node is depended on the random value between 100 m and 200 m.

The throughput of the network is shown in Fig. 1, the red lines indicated the throughput per node with data aggregation, and the blue line indicates the results without data aggregation. The help of phases in our clustering scheme and the use of random timer generated in each sensor nodes eliminate the message collisions. Hence, the throughput per node with data aggregation is higher than non-data aggregation. However, at 2.4 bps of offered load, the throughput of non-aggregation is better than aggregation. The first reason for this bad result can be explained that we have used a random timer to each sensor node for delay broadcasting. However, collision will happen if the random values of nodes are similar. Another reason is the collisions happen at the CH/aggregator because it received many messages from neighbor nodes.

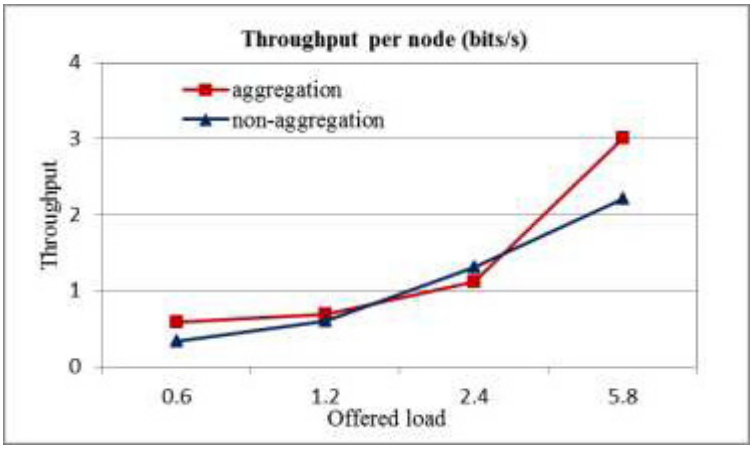


Fig. 1. Throughput (bits/s)

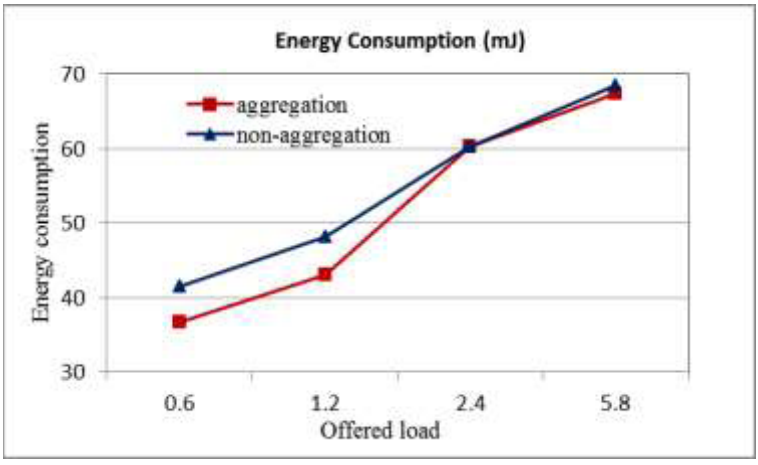


Fig. 2. Energy Consumption of the network (mJ)

Fig. 2 shows the energy consumed of the network with and without data aggregation. The red line indicates the energy consumption of the overall network with data aggregation, and the blue line indicates the results without data aggregation. In this proposed scheme, the energy is saved at each phase of the clustering scheme. For example, only sink nodes are allowed to work during the initial phase. Also, data aggregation with similarity function can save energy by reducing the number of transmissions from cluster heads to the sinks/BS. Hence, the clustered network with data aggregation consumes less energy than clustered network without data aggregation.

5 Conclusions

In this paper, we have proposed a new clustering scheme in UWSNs based on data aggregation with similarity function (Euclidean distance). The data aggregation with similarity function is applied to cluster heads in order to reduce data redundancy. Besides, our new clustering scheme based on data aggregation also achieves a better network throughput and energy consumption than without data aggregation. In the future works, we intend to work deeply on the similarity ratio for more ensuring the accuracy of the data.

References

1. Akyildiz, I.F., Pompili, D., Melodia, T.: Underwater acoustic sensor networks: research challenges. Ad Hoc Networks (Elsevier) 3, 257–279 (2005)
2. Akyildiz, I.F., Pompili, D., Melodia, T.: State-of-the-art in protocol research for underwater acoustic sensor networks. In: Proceedings of the 1st ACM International Workshop on Underwater Networks (WUWNet 2006). ACM, New York (2006)
3. Akyildiz, I.F., Pompili, D., Melodia, T.: Challenges for Efficient Communication in Underwater Acoustic Sensor Networks. ACM Sigbed Review 1(2) (July 2004)
4. Domingo, M.C., Prior, R.: A Distributed Clustering Scheme for Underwater Wireless Sensor Networks. In: IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2007), pp. 1–5 (September 2007)
5. Manvi, S.S., Manjula, B.: Issues in underwater acoustic sensor networks. International Journal on Computer and Electrical Engineering 3(1), 101–111 (2011)
6. Yu, J.Y., Chong, P.H.J.: A survey of clustering schemes for mobile ad hoc networks. IEEE Communications Surveys & Tutorials 7(1), 32–48 (First Qtr. 2005)
7. Pu, W., Cheng, L., Jun, Z., Mouftah, H.T.: A Dependable Clustering Protocol for Survivable Underwater Sensor Networks. In: IEEE International Conference on Communications, pp. 3263–3268 (May 2008)
8. Manjula, R.B., Manvi, S.S.: Cluster based data aggregation in underwater acoustic sensor networks. 2012 Annual IEEE India Conference (INDICON), 104–109 (December 2012)
9. Salva-Garau, F., Stojanovic, M.: Multi-cluster protocol for ad hoc mobile underwater acoustic networks. In: Proceedings of the OCEANS 2003, vol. 1, pp. 91–98 (2003)
10. Yang, G., Xiao, M., Cheng, E., Zhang, J.: A Cluster-Head Selection Scheme for Underwater Acoustic Sensor Networks. In: International Conference on Communications and Mobile Computing (CMC), vol. 3, pp. 188–191 (April 2010)

11. Tran, K.T.-M., Oh, S.-H., Byun, J.-Y.: Well-Suited Similarity Functions for Data Aggregation in Cluster-Based Underwater Wireless Sensor Networks. International Journal of Distributed Sensor Networks 2013, Article ID 645243, 7 pages (2013)
12. Virmani, D., Sharma, T., Sharma, R.: Adaptive Energy Aware Data Aggregation Tree for Wireless Sensor Networks. International Journal of Hybrid Information Technology (IJHIT) 6(1), 25–36 (2013)
13. Maraiya, K., Kant, K., Gupta, N.: Wireless sensor network: A review on data aggregation. International Journal of Scientific and Engineering Research 2(4), 269–274 (2011)
14. LinkQuest, http://link-quest.com/html/uwm1000.htm
15. Tran, K.T.M., Oh, S.H.: A cooperative MAC scheduling scheme for underwater sensor networks. Applied Mechanics and Materials 295-298, 903–908 (2013)

Efficient Voice Communications over Wireless Sensor Networks

Hyunchul Yoon[1] and JaeHyung Lee[2]

[1] LIG Nex1, Seongnam-City, Gyeonggi-do, Korea
hyunchul.yoon@lignex1.com
[2] Hanyang University, Seoul 133-791, Korea
jhlee@ewc.hanyang.ac.kr

Abstract. Wireless Sensor Networks(WSN) are used to communicate between Sensor nodes. The ZigBee telecom applications profile specification is announced for a voice communications in the ZigBee networks by the ZigBee Alliance since 2010[1,2], but it has not been considered to interface with the long distance communications. We designed Voice over Sensor Networks(VoSN) considered on long distance communications with each other nodes.

Keywords: Wireless Personal Area Networks(WPAN), IEE 802.15.4 Standard, ZigBee, Voice over ZigBee(VoZ), Voice Over Sensor Networks(VoSN).

1 Introduction

Applications of WSN in ISM band are being increased continuously, and IEEE standardization committee recommend IEEE 802.15.4 standards for 2.4GHz wireless personal area network [1]. Standard of IEEE 802.15.4 using 2.4GHz frequency is marked to support such as low power, low cost, and low data rate communication. Telecomm Applications Profile Specification based ZigBee announced the standard for application, which is used in telecom markets; "peer-to-peer small data sharing," and "chatting,""etc."[2,3].

According to the voice communication device based on sensor network using DSP Codec chip by Hu[4], it is demonstrated a voice communication of the TDD method for dividing uplink and downlink on ZigBee, but this method is not suitable for multichannel, because it uses ADPCM that support only low compressed voice codecs. On the other hand, different codecs are proposed to compress voice data, it has raw audio codec such as 32, 62.5, 30.7, and 122.9 kbps in [5], 32 kbps, ADPCM-1 16 kbps, GSM-1 13 kbps, ADPCM-3 8 k bps, GSM-2 7kbps in [6], and G.729A 8 kbps in [7]; moreover, TDMA-based protocol with scheduled and contention slots is proposed in [6]. But those methods are used in G.729 codec and TDMA protocol concurrently. Eunchang in [8] consider multi nodes based the IEEE 802.15.4 Low Rate Wireless Personal Area Networks(LR-WPAN) using codec such as 16 kbps, 32 kbps, and 64 kbps, but these studies are not fit long distance like telecommunication and multi-user

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_70,

communications. So we suggest a network configuration method for multi users, long distance telecommunications in WSN. In the rest of paper, we present the architecture of VoSN in Section 2. We evaluate the performance of VoSN in Section 3. Finally, we conclude this paper in Section 4.

Mobile Nodes(MN) can communicate only through the coordinator when it initiates to communicate other MNs, It supports only singe hop. To improve this problem, the VoSN system is suggested by the method connected with WiFi networks used in general. VoSN consists of MN, Base Stations(BS), and Proxy servers(PS). Figure 1 shows the reference model of VoSN for voice communications.

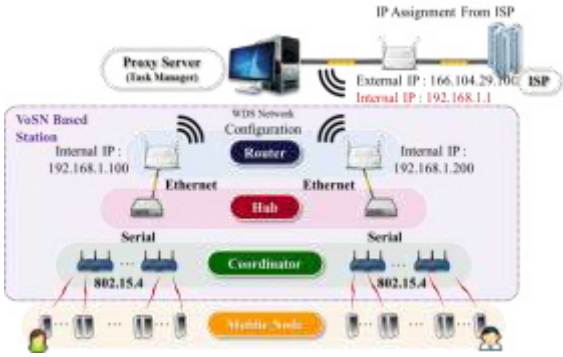


Fig. 1. Reference model of VoSN for voice communication

The VoSN of wireless communication system is designed in the following basic processing method. It is shown in Figure 2.

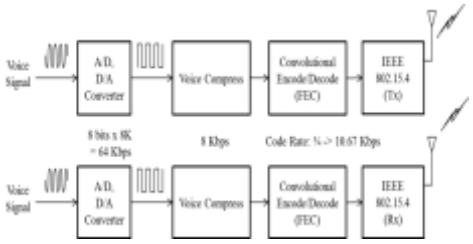


Fig. 2. VoSN system block

The BS consists of the coordinator, HUB and router. The coordinator connects between MNs or MN and router. Router has function to control the MN to manage multi channels, and to support voice communication multiplexing between MNs. The Proxy Server(PX) has Simple Session Initiation Protocol (S-SIP) that has several functions such as handoff, subscriber management, location registration, and authentication.

2 Implementation

The VoSN has various protocol layers such as PHY, MAC, Network, Application: A/D conversion, G.729A voice compression, then transmit the data to the network through the RF modulation based IEEE 802.15.4 finally; received data carry on the demodulation, G.729A voice decompression, D/A conversion and through low pass filter(LPF) in order to get original analog voice signal. In addition, VoSN supports Wi-Fi networks for the long distance communications.

2.1 Mobile Nodes(MN)

MN consists in Figure 3. It shows MN structure and components.

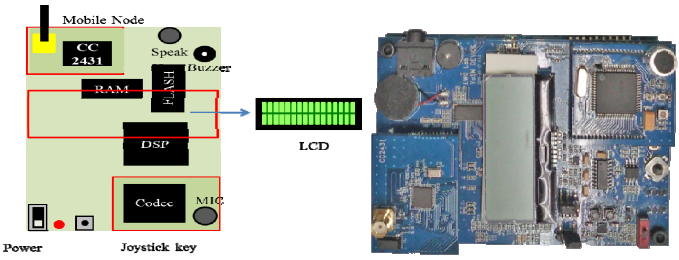


Fig. 3. Mobile Node Structure and Components

A/D(8bit x 8kHz sampling), D/A, G.729 codec, echo cancellation, Band Pass Filter (BPF), 1/2 convolution coding, interleaving function are designed in the VoSN.

2.2 VoSN Base Station (BS)

The BS consists of a coordinator, a HUB, and a router which has following features.

- The coordinator supports the function of the star network topology for multiuser, to transmit beacon messages to MN, to manage the time slot and the call processing and frequency channels in cell.
- The HUB which is connected the router is used to connect with parallel coordinators, and convert the mutual protocol between serial and TCP/IP.
- The router is used to connect with each other routers for routing a voice frame generated from voice codec by 20ms.

2.3 SIP Server

SIP Server has a feature managing the network topology of MNs, coordinators, routers and HUBs in the PS. Figure 4 shows the SIP routing operation.

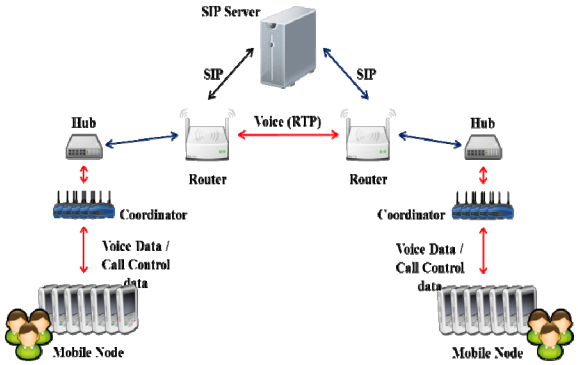


Fig. 4. SIP Routing operations

2.4 Protocol Stack

Figure 5 shows the VoSN protocol stack. It is designed for flexible communication for each MNs, coordinators, and gateways. MN protocol stack has the IEEE 802.15.4 PHY, MAC, Simple Real Time Protocol(S-RTP), S-SIP, and application layer. The coordinator protocol stack is similar to the MN, but it adds MN to the RS-232 for a serial communications with the gateway, do not need application layer. The gateway stack has the RS-232 PHY to serial communications with the coordinator, the IEEE802.11n PHY to communicate with other router, UDP/RTP, SIP, and Application layers. The gateway stack has the full function of SIP to support the VoIP service in the future; therefore, the S-SIP is designed to get a compatible ability with the SIP. Figure 6 shows the VoSN protocol stack.

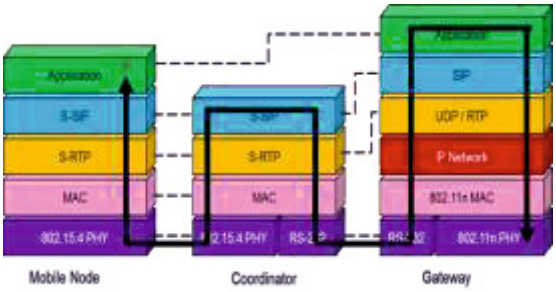


Fig. 5. VoSN Protocol Stack

2.5 MAC Design

ZigBee networks do not support the beacon mode in IEEE 802.15.4. Therefore, we design time slots for VoSN using IEEE 802.15.4 MAC.

In order to operate the beacon mode, and to support multi-channel, we design the MAC structure of VoSN. Figure 6 shows a super-frame structure for the VoSN. The Beacon Interval(BI) and the Super-frame Duration(SD) are designed to have the same slot duration size, and to maximize channel capacity.

$$SD = aBaseSuperframeDuration \times 2^{SO} \quad (1)$$

$$BI = aBaseSuperframeDuration \times 2^{BO} \quad (2)$$

Where

aBaseSuperframeDuration

$$= aBaseSlotDuration \times aNumSuperframeSlots \quad (3)$$

SO is the parameter of MAC Super-frame Order. BO is the parameter of MAC Beacon Order. Value of each parameter has to connect in one cycle of BI in 20ms which is voice codec cycle. *aBaseSuperframeDuration* is 960 symbols because aBaseSlotDuration has the constant value of 60[1,2]. aNumSuperframeSlots has a constant value of 16 in IEEE 802.15.4 specification[1]. Therefore, 60 symbols are created during one slot. It is possible to transfer 30 octets' data. In here, two symbols are one octet. In addition, voice data generated from voice codec is 20 octets during 20ms. Moreover, one octet of the sequence number used to count voice packet is added in 20 octets. Figure 6 shows the Super Frame Structure.

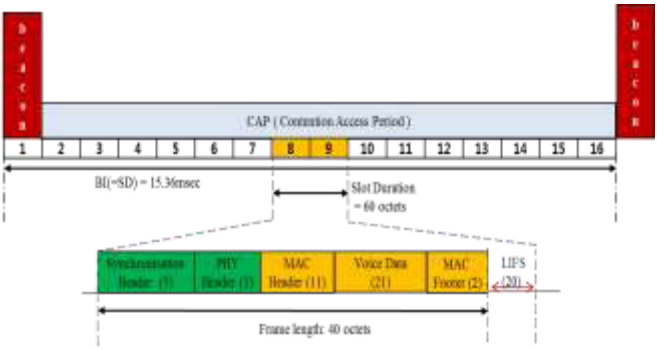


Fig. 6. Super Frame Structure

In here, BI is 15.36ms and a voice data are generated by 20ms, BI and voice data generation cycle is mismatched each other; nevertheless, voice data can transmit because the BI cycle is faster than the voice sampling cycle.

3 Evaluation

Figure 7 shows the test configuration to evaluate two modes of VoSN such as peer to peer mode, and mode through routers. We select 15 channel numbers for minimizing interference among IEEE 802.15.4 frequency channels, because this channel is not used in Wi-Fi.

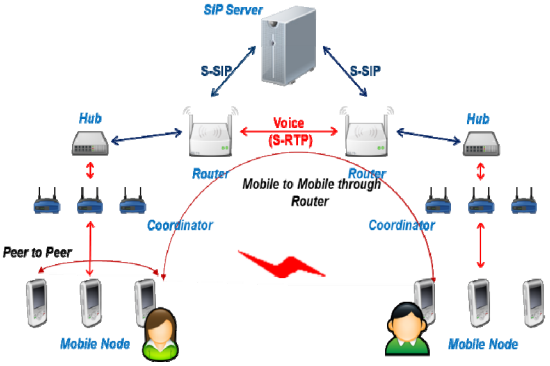


Fig. 7. Test Configurations of VoSN

Figure 8 shows the voice transmission delay of about 120ms in the peer to peer mode of VoSN and voice transmission delay of about 130ms in the mode through router.

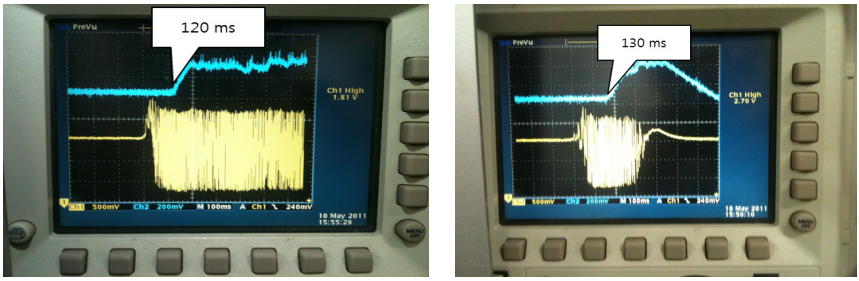


Fig. 8. Voice transfer delay using peer to peer and through router

The first method is satisfied by VoIP specification to allow120ms delay for voice commutation; however, the second method does not. Nevertheless, time difference between first and second method is very small, so it can be ignored. Table 1 shows performance of peer to peer and mobile to mobile through a router.

Table 1. Performance between Peer to Peer delay and Mobile To Mobile through router

Distance(m)	Peer to Peer		Mobile to Mobile through router	
	Packet Loss(%)	MOS	Packet Loss(%)	MOS
50	0.1	5	0	5
100	1.2	5	0	5
160	4.9	5	0	5
200	8.4	4	0	5
240	30.5	3	0	5
260	68.2	1	0	5
280	98.4	disconnected	0	5

Table 2 shows the MOS and the packet loss rate according to distance between two MNs.

Table 2. MOS and packet loss measurement results according to distance between MNs

Type	Frame Delay(ms)	MOS
Peer to Peer	120	5
Mobile to Mobile through Router	130	5

4 Conclusion

This paper shows the possibility of voice communications between MNs to use the method proposed in the VoSN. It shows the communication available maximum distance using the first method. Therefore, we look forward to it used as the communications system connected between WSN and telecommunication in future.

References

1. IEEE Std 802.15.4-2006, wireless medium accesses control (MAC) and physical layer (PHY) specifications for low-rate wireless personal area networks (WPANs)
2. ZigBee Applications Profile Specification, ZigBee (April 1, 2010)
3. ZigBee Alliance, http://www.zigbee.org
4. Hu, R.L., Yin, J.R., Gu, X.J., Gu, X.P., Chen, L.Q.: The Research and Design on TDD Voice WSN. In: International Conference on Multimedia Technology (ICMT), pp. 1–4 (October 2010)
5. Lee, J.U., Choi, E.H., Huh, J.D.: Voice over ZigBee Networks with Environmental Monitoring. In: IEEE International Conference on Consumer Electronics (ICCE), January 9-12, pp. 633–634 (2011)
6. Rahul, M.H., Anthony, R.W., Raj, R.K., Ryhei, S.Z.K.: Voice over Sensor Net-works. In: 2006 27th IEEE International on Real-Time Systems Symposium, December 5-8, pp. 291–302. RTSS (2006)
7. Li, W.F., Cao, C.G., Han, F.: Short-distance Wireless Voice Communication. In: International Conference on Management and Service Science (MASS 2009), September 20-22, pp. 1–4 (2009)
8. Choi, E.C., Hur, Y.K., Huh, J.D., Nam, Y.S., Yoo, D.H., Choi, W.C., Choi, W.C.: Simulation and Implementation of Voice-Over-IEEE 802.15.4 LR-WPAN. In: International Conference on Consumer Electronics (ICCE 2008), January 9-13. Digest of Technical Papers, pp. 1–2 (2008)

Development of SWF Based Virtual Prototyping Framework for Simulating Ubiquitous Systems

Soo Young Jang, Jihun Kim, and Woo Jin Lee

School of CSE, Kyungpook National University, Republic of Korea
ssf321@naver.com, hjkwlgns@gmail.com, woojin@knu.ac.kr

Abstract. Currently, ubiquitous systems, whose mission is to control peripheral hardware and to interact with peripheral environment, are widely used in various fields such as home appliance, medical device, vehicle, and so on. In developer's viewpoint, it is difficult to set up the hardware and operational environment required for the system during software development. Virtual prototyping approach can be used for simulating hardware and operational environment of embedded software. However, cost and effort required for developing virtual prototyping are very high. Furthermore, related supporting tools have some limitations such as detailed representation, dynamic performance, modifiability, and so on. In this paper, we propose an SWF based virtual prototyping framework for implementing virtual prototype in embedded systems. And our experiment conducted for showing the applicability of our approach in real world. Our approach also supports advanced simulation and multi-dimensional analysis during software development process for embedded systems.

Keywords: Virtual Prototyping, Embedded Software, State Machine Diagram, SWF.

1 Introduction

Ubiquitous systems are widely used in various fields such as home appliance, medical device, vehicle, and so on. Their mission is to control peripheral hardware and to interact with peripheral environment. Since embedded software development that makes up the ubiquitous system should consider hardware couplings, environmental variables, and limited resources, it is nontrivial to develop ubiquitous embedded software. In addition, due to the requirement of actual hardware for operational and verification process, the development complexity is more increased to make the development time longer. Long development cycle is particularly not good for IT industry where the market changes rapidly because it may cause the loss of competitive market advantage [1]. Large part of the problem can be solved by using virtual prototype approach which is applied for simulating hardware and operational environment of embedded software. However, cost and effort of developing virtual prototype are very high.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_71,

In this paper, we propose the SWF (Shockwave Flash) based virtual prototyping framework for simulating the ubiquitous systems. The framework can simulate the ubiquitous system environment by connecting independent software simulator and the SWF based virtual prototype with other software simulator. The SWF based virtual prototype is composed of internal components and those internal components are operated by the state message delivered from the software simulator. Conversely, when a user provides input events to the SWF based virtual prototype using standard I/O devices, the events are converted into message types and then transmitted to the simulator for determining the next decision and the internal state of the target software.

The rest of the paper is organized as follows. Section 2 introduces the embedded software virtual prototype approach and related tools. Section 3 presents the SWF based VP framework, a design method for the VP components, and a technique of composing the SWF based entities. Section 4 shows a case study for applying this framework with the Flash component editor that is developed in this work. Section 5 concludes the paper.

2 Related Work

2.1 Virtual Prototyping of the Embedded Software

Generally, virtual prototyping is a modeling method for assessing the validity and performance of the software and hardware systems prior to full-scale production. The prototype methods are divided into virtual prototype and physical prototype depending on the type of the result. The virtual prototype is created by using graphical computer technique to simulate the behaviors of functions and the real hardware components. In the developers' viewpoint, this approach can be used to execute the software model such as the state machine diagram for analyzing the system requirements as well as it can be used to execute the source code developed in an environment without hardware for checking the execution results. These advantages not only reduce the gap of the common goal for the target system but also support to implement the connecting system that meets the specifications and multidimensional analysis of design. In order to extract these benefits of virtual prototyping, the research on more easily and efficiently producing the virtual prototype has been actively conducted [2] [3]. However, there are still some of problems such as cost of visual effect, degree of difficulty in development, modifiability, and so on.

2.2 Related Tools for the Virtual Prototyping

Virtual prototypes can be developed using several programming techniques. However, cost and effort required for developing virtual prototyping are too high so they are often considered as another software project especially in the case of distributed ubiquitous environments, which increase the complexity of the VP application. The RapidPLUS[4] and Rhapsody[5][6] are the representative tools for resolving the aforementioned problems. However, RapidPLUS has limited support for

components required for the embedded systems. Furthermore, visual effects of supported components can not represent dynamical behavior because of its static characteristic. Therefore, in order to express special and complex features of components similar to the actual system dynamically, professional graphics technologies are required. Rhapsody, similar to RapidPLUS, produces all the visual objects for simulation by programming technique and then it links those visual objects to virtual prototype application after creating the objects library. However, when the requirements are changed, it is quite difficult to modify the application since the VP application has complex structure. The VP application produced through this process has as complex structure and also it is quite difficult to modify the application when the requirements are changed.

3 Design of the SWF Based VP Framework

3.1 Structure of the SWF Based VP Framework

The SWF is an application format representing the vector graphic objects. Using Adobe's Flash for creating the SWF formatted entity does not require visual expertise in VP developers and also dynamic and detailed behavior of hardware can be easily expressed. Therefore, representing the operational environment and virtual hardware dynamically by using SWF based VP image is much easier than to develop a VP application through general programming technique. The SWF VP is operated by exchanging the state message with software simulator. To develop a state model or the source code on the target system, such as the software element that is being done on the software simulator, environmental factors and the surrounding hardware attributes are implemented as SWF based VP.

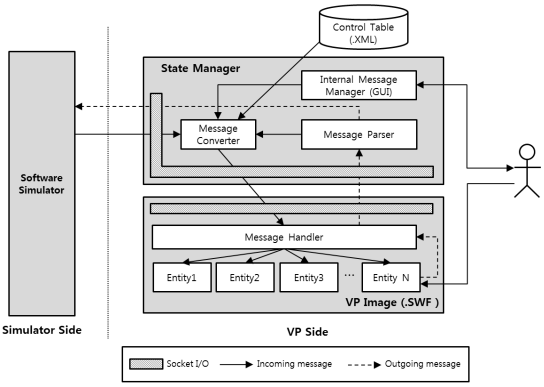


Fig. 1. Structure of the SWF based VP Framework

Internally, the SWF based VP internally consists of VP images in SWF format, a control table, and a state manager as shown in Figure 1. A VP image also consists of a message handler and a set of entities. The message handler, which directly controls

the entities located in the VP image, receives control messages from the state manager. Entities can be made using Adobe Flash to visualize the components of hardware in VP. In designer's viewpoint, a VP component can have multiple entities and also a VP image can have multiple VP components. The dependencies between a VP component and its entities do not appear in the VP image file because the entities are stored in sequential format. Therefore it is not possible to visualize the components of the hardware. The control table contains information on how to group the entities, how to represent and visualize the state of VP components, and how to operate the VP image according to the state messages. The state manager manages communication between the simulator and the SWF based VP image and it also controls the VP image according to the state message delivered by the simulator. Once a state message is received, the state manager converts the state message into control messages with reference to the control table and then transmits the control messages to the VP image. Conversely, when a user enter an event to the VP image, the state manager sends it to the simulator after converting the event into the type of message, and manages the internal message. Figure 1 shows the structure of the SWF based VP framework.

3.2 Steps for the Development of SWF Based VP Images

As mentioned previously, the VP image is created through combining the entities into a single file with the message handler dynamically. Following are the steps for the development of SWF VP:

- *Step 1: Describing a control state diagram of VP component.*

The target system is first divided into component units. The VP component should be defined as the unit retaining physical control state in the target terminal of ubiquitous system. The terminal is made up of one or more components. The behavior of a component is defined as the interface featuring hardware viewpoint and the control state is defined based on what behavior should operate in which state. This makes possible for VP image to operate by calling the predefined interface when the state of component is changed. Control state diagram notation is based on UML state machine diagram [7] and the describing method for the diagram is described in the Section 3.4.

- *Step 2: Creating entity images to configure components.*

An entity is the smallest unit in the VP image and represents a part of the VP component. The entity is in SWF format and created by using Adobe Flash tool. When the component diagram is placed in a particular state, the entities properties are dynamically changed to represent the specific behavior of the component. After completion of creating the entities, the VP component is designed visually by setting the layer and the positions at the entities.

- *Step 3: Composing entities and generating a VP image.*

An entity is located visually and mapped into particular state of the component state diagram and those mapping relationships are saved in both the control table file and the component property file. Control table information allows the compatibility between the software simulator and the SWF VP images. To generate the VP image, entities are combined into a single SWF file with reference to the property file. Through the de-compilation of SWF file, internal crashes such as identifier crash are found and corrected. And then the visual properties are applied before recompiling into a single SWF image file.

3.3 Definition of the Control State Diagram

In order to represent the control flow of the hardware components, we define a control state diagram by adapting UML state machine diagram. The difference between two diagrams is that the state machine diagram is used to model the software whereas the control state diagram is used to model the hardware components. In the control state diagram, the status of a VP component is defined into the state and switching between these states is expressed into the transition. Figure 2 shows the notation for defining the control state diagram using the BNF [8]. The notation is designed to operate between the VP component and software simulator based on the message.

```
<transition> ::= <trigger> [‘/’ <activity >[‘,’ <activity >]*]
<trigger> ::= <incoming-message> | <user-event>
<incoming-message> ::= <incoming-spec>[‘(’<parameter-spec>‘)’]‘?’
<user-event> ::= ‘[’<entity-name>‘]?’
<activity > ::= <internal-message> | <outgoing-message>
<internal-message> ::= <internal-spec>‘&’
<outgoing-message> ::= <outgoing-spec>‘!’
```

Fig. 2. The Notation of the Control State Diagram

In this notation, the activity can be composed of the internal message and outgoing message. The internal message has no effect on the software simulator and it is used when the messages generated in the VP is being processed within the VP. On the other hand, outgoing message is generated by the VP image and is delivered to the software simulator. This message has effects on the software simulator and it is generated when the operation on the VP component is changed such as button or sensor operation.

3.4 Composing the SWF Based Entities

To combine the entities encoded into the binary code into a single VP image file, we should remove the header of each entity files because the header is optimized separately for each entity file. And then properties generated while designing VP

component are applied to the body of the entity files. Finally, the body of all the entities is appended to the end of header that is newly defined. The header of the SWF file has such properties as the length of the file, signature type, frame size, and frame rate [9]. Table 1 shows the hexadecimal representation of encoded entity file. The first 3 bytes indicate the signature type of the file, and it is defined in the CWS (Compressed Work Schedules) or FWS (Flexible Work Schedules) accordingly. As shown in the front part of the Table 1, ASCII code values 0x43, 0x57 and 0x53 means that the signature type is CWS. From Table 1, we can also see that the 2 bytes of the field 0x11 (0x00, 0x1E) indicates that the frame rate was set to 30 frames per second. By applying this principle, we can set the customized header for all entities.

Table 1. Hexadecimal Representation of the Encoded SWF File

Off	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
00	43	57	53	09	42	13	00	00	78	00	05	0F	00	00	0F	A0
10	00	00	1E	01	00	44	11	00	00	00	00	43	02	FF	FF	FF
20	FF	14	4A	00	00	00	01	00	7F	FF	63	99	3F	FD	93	AA
⋮								⋮								⋮

After completing the header configuration, we should resolve the identifier conflict among SWF entities by modifying the properties with reference to the information which was created in the previous step. The movie clip symbol presented in flash tool has data fields that are composed of type, length, extended length, identifier, frame count, and objects. According to the identifier, the SWF player can access to all objects that are located in the SWF file. While combining the entities which were created separately, an error can occur due to duplicate identifier because the value of identifier is assigned in increasing order starting from 1. In order to prevent this error, the identifier should be reassigned to each entity before combining them. Since an object field in the movie clip symbol has a set of visual objects that make up the entity, the object field should be modified to complete the composition of the VP component.

4 Development of a VP Composing Tool

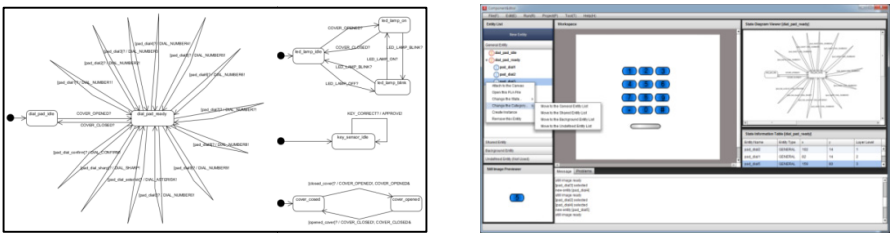
4.1 The Flash Component Editor

The Flash component editor is a tool for supporting this framework. This section explains the process to develop the VP image by using the Flash component editor. At first, the Flash component editor reads the control state diagram file to display the diagram to the screen. Then, a user selects a particular state in the control state diagram and creates entities by using the Adobe's Flash tool. Those entities created by the user represent selected state of the VP component. Next, the Flash component editor extracts a PNG format still image from an entity created by user using SWF Render [10]. And the user places the still image on the canvas of the Flash component

editor visually. The whole process is repeated for all of the control state diagrams. The Flash component editor creates the control table file and the message handler with reference to design information from aforementioned process automatically and then it finally combines the entities into a single SWF file to create the VP image. For automatically combining the entities, we use the transtfor-3.0.2 library [11].

4.2 VP Example for Applying to This Framework

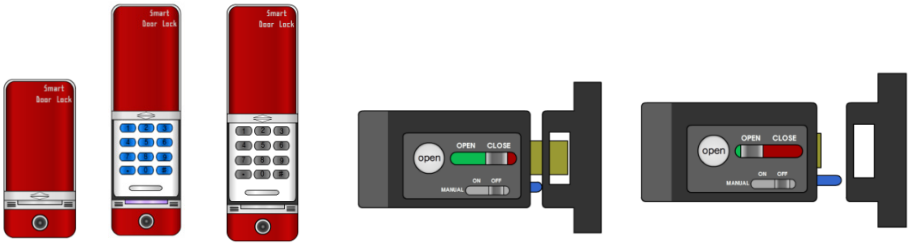
The digital door lock is one of the devices commonly used for home automation system. In this section, as an example, we create a VP image of the digital door lock to illustrate our framework. To create a VP image, we divide the device of the digital door lock into VP components and modeled control state diagram for each VP component. For each state of the control state diagram, the entities are created to visualize VP component by using the Flash component editor.



(a) Control State Diagram of the Digital Door Lock Components

(b) Composition Process of the VP Component of the Digital Door Lock

Fig. 3. The Control State Diagram of the Digital Door Lock Components



(a) Front side Device

(b) Backside Device

Fig. 4. VP Image of the Door Lock Devices

Figure 3 (a) shows the control state diagram for digital door lock device composed of button set, dial cover, LED lamp, and key recognizer. This diagram is modeled to change the locking status and the LED lamp status through user interaction with the digital door lock system. The snapshot of Flash component editor is shown in

Figure 3 (b). Figure 4 represents snapshots of VP images which interact with simulator drivers. At the front side device, a user can control the VP image by using mouse click as if controlling the real device, for example, opening the device cover and pushing the dial buttons. As shown in Figure 4 (b), the user can control not only the lock and unlock status but also the locking mode composed of automatic and manual in the backside device. In our framework, the transition of the VP image takes place dynamically and smoothly, so it is possible to visualize the behaviors of the devices.

5 Conclusion and Future Work

In this paper we have proposed a framework of SWF based virtual prototyping which overcomes the limitations of existing virtual prototyping methods and tools. The Flash component editor supports the automation of most of the processes including creating entities, and generating VP image. Furthermore, we created a VP example for the digital door lock in order to show that our framework and the Flash component editor can easily develop a VP environment of an embedded system. The SWF based virtual prototype supports advanced simulation and multi-dimensional analysis in the ubiquitous system. Besides, VP application can be easily developed and modified.

In future we intend to analyze the software errors through this VP framework and to define the context diagram which can represent both control state diagram and the state machine diagram.

Acknowledgments. This work was supported by the IT R&D program of MISP (Ministry of Science, ICT & Future Planning)/KEIT. [10041145, Self-Organized Software platform (SoSp) for Welfare Devices] and the MSIP, Korea, under the C-ITRC (Convergence Information Technology Research Center) support program (NIPA-2013-H0401-13-1005) supervised by the NIPA (National IT Industry Promotion Agency).

References

1. Kamat, S.P.: Time to Market: Handheld Consumer Electronics Devices and Its Impact on Software Quality. Consumer Electronics Magazine IEEE (2012)
2. Sukmana, H.T.: User-level Virtual Prototyping for Television Simulation using SystemC and Java GUI (2005)
3. Kim, D.-E., Lee, J.-B., Rim, K.-W., Hwang, Y.-S., Ahn, S.-S.: Design and Implementation for Cross Development Environment based on Virtual Proto-typing Development Tools. Journal of Korea Contents Association 9(5), 40–49 (2009)
4. RapidPLUS, https://sites.google.com/site/rapidpluscommunity/
5. Rational Rhapsody, http://www-142.ibm.com/software/products/kr/ko/ratirhap/
6. Liu, X.-H., Cao, Y.-F.: Design of UA V Flight Control System Virtual Prototype using Rhapsody andSimulink. In: 2010 International Conference on Computer Design and Applications, ICCDA (2010)

7. OMG.: OMG Unified Modeling Language (OMG UML) Version 2.4.1
8. Backus–Naur Form, http://en.wikipedia.org/wiki/Backus%E2%80%93Naur_Form/
9. Adobe systems corp.: SWF File Format Specification Version 10
10. SWFRender, http://wiki.swftools.org/wiki/Main_Page/
11. Flagstone Software Ltd., http://www.flagstonesoftware.com/transform/index.html/

Study on Relation between Social Circles and Communities in Facebook Ego Networks

Soo-jin Shin[1], Yong-jin Jeong[1], Chan-Myung Kim[1], Youn-Hee Han[1,*], and Chan Yeol Park[2]

[1] Advanced Technology Research Center,
Korea University of Technology and Education, Korea
{soojin1116,jyjin989,cmdr,yhhan}@koreatech.ac.kr
[2] Supercomputing Center,
Korea Institute of Science and Technology Information,
Daejeon, Republic of Korea, 305-806
chan@kisti.re.kr

Abstract. Community detection is a core problem in social network analysis. Strictly speaking, however, the communities does not exactly correspond to the real group, well-known as social circles. In this paper, we study on 1) how close relation between the ground-truth social circles and communities exists and 2) whether the social circles can be detected by the classical community detection algorithm or not. We use the SNAP facebook dataset to reveal the correlation between the social circles and the detected communities. We listed up the community's modularity values and the balanced accuracy values with the ground-truth circles per each level in the iterative process of divisive clustering. We analyzed the Spearman's rank correlation between the paired data. The experimental results show that there is a strong correlation between the ground-truth social circles and the communities detected by classical method.

1 Introduction

Nowadays, there are numerous online social networks services, e.g., Facebook, Twitter, Google+, etc. They form human social networks by generating new connections between nodes. In a social network, each node usually denotes a user and edge denotes a connection with a friend or an acquaintance [1].

Most social networks services also allow users to categorize their friends into social circles, e.g., 'circles' on Google+, and 'lists' on Facebook and Twitter. Currently, users usually manage their circles either by manually identifying friends sharing a common attribute. This method is time consuming and does not update automatically as a user adds more friends into his/her social network. In addition to that, it cannot manage users social circles finely when a user's profile information is missing [2]. Therefore, the problem of automatically discovering users social circles has become an important issue.

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_72,

One of other important issues in social network analysis is the detection of community structure in social networks [3]. The most widely used and accepted definition of a community is follows: “a group of nodes of a graph which are more strongly connected to each other than with other nodes in the same graph.” Many approaches to detect communities in a social network have been proposed in the past [1]. Among them, we use the hierarchy-centric divisive clustering method that is to build a hierarchical structure of communities based on network topology [4].

In this paper, we study the relation between communities and social circles. As mentioned above, the communities are detected by the given algorithm, particularly hierarchy-centric divisive clustering algorithm, while social circles are manually managed by users identifying friends sharing a common profile attribute. If there is a strong correlation between them, we may be able to solve the social circle detection problem by using the classical community detection methods.

The rest of this paper is organized as follows. Section 2 presents preliminaries and general approach for analyzing the correlation between social circles and communities. Section 3 shows an experimental analysis, and Section 4 finally concludes this paper.

2 Relation between Social Circles and Communities

2.1 Preliminaries

In this subsection, we present the used notions for analyzing the correlation between social circles and communities.

To compare the social circle with identified community, we first use the hierarchy-centric divisive clustering algorithm that is to build a hierarchical structure of communities based on network topology [4]. More specifically, it first partitions the nodes into several disjoint sets. Then each set is further divided into smaller ones until each set contains only one node. The key here is how to split a network into several parts, and one particular divisive clustering algorithm is to recursively remove the “weakest” tie in a network. that is, The weakest tie, the higher value of edge betweenness. Newman and Girvan [5] proposed to find the weak ties based on “edge betweenness”. Edge betweenness is a measure to count how many shortest paths between pair of nodes pass along the edge, and this number is expected to be large for those between-group edges.

For calculating the correlation between social circles and identified communities by the divisive clustering algorithm, we use the community modularity, the balanced accuracy and Spearmans rank correlation coefficient. Modularity [3] is usually used to measure the strength of a community partition by taking into account the degree distribution of nodes. Accuracy [6] considers all the possible pairs of nodes and checks whether they reside in the same community and social circles. It is considered error if two nodes of the same community are assigned to different social circles, or two nodes of different communities are assigned to the same social circle. Balanced

Accuracy [6] can be defined as the average accuracy assigning equal importance to false positives and false negatives. Spearmans rank correlation coefficient assesses how well the relationship between two variables can be described using a monotonic function ((i.e., that when one number increases, so does the other, or vice-versa). A perfect Spearman correlation of +1 or 1 occurs when each of the variables is a perfect monotone function of the other.

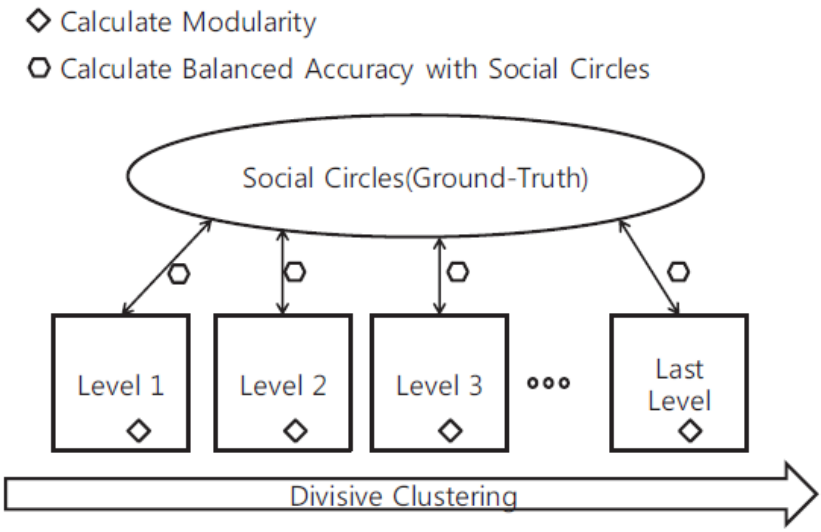


Fig. 1. Experimental Process

2.2 Social Circle and Community

It would be nice if there were a direct link connecting social circle and community. The general approach is simple. If there is a strong correlation between them, we may be able to solve the social circle detection problem by using the classical community detection methods. For finding this correlation, the social circle for each user can be known. Unfortunately, this is not possible in general. It is not easy to see a scenario that hardly presents itself in real-world large-scale networks. For the study, so, we use real Facebook ego network datasets provided by SNAP (Stanford Network Analysis Project) [4].

The figure 1 shows the experiment process for study. In advance of analysis, First we generate an ego network for each node based on Facebook ego network datasets. An ego network consists of a local node (“ego”) and the nodes (“alters”) to whom the ego is directly connected to plus the edges, if any, among the altars [7]. For each ego network dataset, we first 1) executed the community detection by using the hierarchy-centric divisive clustering algorithm, 2) listed up the communitys modularity and the balanced accuracy between community and the social circles (i.e., ground truth) per

each level in the iterative process, and 3) analyzed the Spearmans rank correlation between the two datasets, i.e., the modularity and the balanced accuracy.

3 Experimental Results

We conducted the above experiment on a Facebook ego networks. We randomly choose the seven ego networks with 0, 348, 414, 686, 690, 3437 and 3980 node. For each ego network we increase the level of divisive clustering from 0 to 2500.

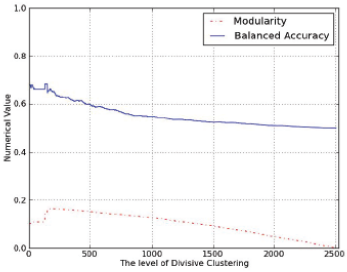
Figure2 depicts the two values, the modularity and the balanced accuracy, per the modularity and the balanced accuracy. In figure2, the level of Divisive clustering is plotted along the X axis, and the numeric value is plotted along the Y axis. As can be seen from the figure2, in general, the two values are highly correlated, meaning that there is high strength of a linear relationship between the paired data. For getting exact value of correlation coefficient we also obtained the value of Spearman correlation. The results are shown in Table 1 and the average of Spearman's Rank is 0.852.

In case of 348 and 3980 node, Spearman's Rank is relatively low compared with other node's cases. The most widely used and accepted definition of a community is follows: "a group of nodes of a graph which are more strongly connected to each other than with other nodes in the same graph". However, there are many nodes weakly connected to each other than with other nodes unlike above definition in those cases. Those weak ties make Spearman's Rank low. Actors in real worlds tend to form closely-knit social circles. That is, individuals interact more frequently with members within social circles than those outside the social circles. Therefore, it is very unusual cases in a real worlds.

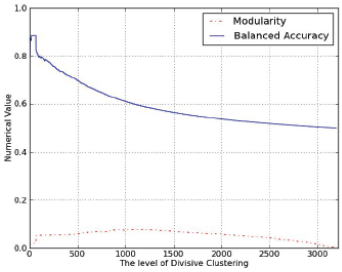
These results seem to confirm our suspicions. Clearly, we need to do more extensive testing of this hypothesis but at this stage it seems reasonable to conclude that these is evidence that classical community detection method can be used to detect social circles.

Table 1. Results of Spearman's Rank obtained in our experiment

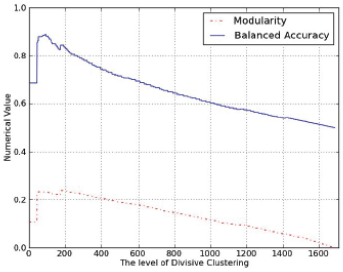
Ego network	Spearman's Rank
0	0.922
348	0.570
414	0.986
686	0.950
690	0.987
3437	0.981
3980	0.571
Average	0.852



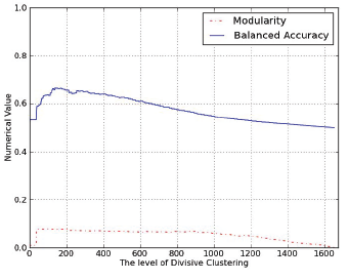
(a) 0.Ego Network Result



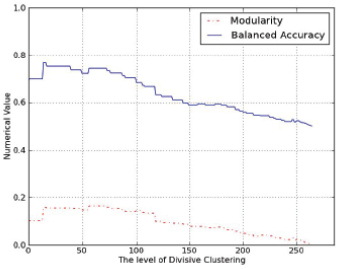
(b) 348.Ego Network Result



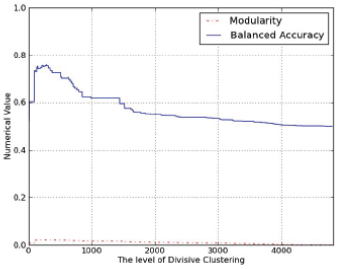
(c) 414.Ego Network Result



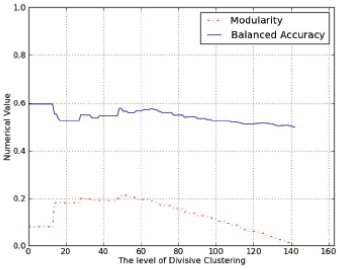
(d) 686.Ego Network Result



(e) 698.Ego Network Result



(f) 3437.Ego Network Result



(g) 3980.Ego Network Result

Fig. 2. Results of our experimental analysis on Facebook ego networks

4 Conclusions

In this paper, we study the relation between the ground-truth social circles and the communities are detected by the classical community detection Title Suppressed Due to Excessive Length 5 algorithm. From the experimental results, we obtained an average value of 0.852 of Spearman's rank correlation between them. While social circles are manually managed by users, we can conclude that classical community detection method can be used to detect social circles. However, the details on how to apply the algorithms to detect social circles are still open issue. We just disclosed the possibility of it. In future works, we will conduct experiments on various networks by using many community detection 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 (2013R1A1A2010050), and also financially supported by the Ministry of Knowledge Economy (MKE) and Korea Institute for Advancement of Technology (KIAT) through the Workforce Development Program in Strategic Technology.

References

1. Tang, L., Liu, H.: Community Detection and Mining in Social Media. Morgan & Claypool Publishers (2010)
2. McAuley, J., Leskovec, J.: Learning to Discover Social Circles in Ego Networks. NIPS (2012)
3. Newman, M.: Modularity and community structure in networks. PNAS 103(23), 8577–8582 (2006)
4. Karypis, G., Han, E., Kumar, V.: Chameleon: A hierarchical clustering algorithm using dynamic modeling. IEEE Computer 32(8), 68–75 (1999)
5. Newman, M., Girvan, M.: Finding and evaluating community structure in networks. Physical Review E (2004)
6. Brodersen, K.H., Ong, C.S., Stephan, K.E., Buhmann, J.M.: The balanced accuracy and its posterior distribution. In: Proc. of the 20th International Conference on Pattern Recognition (August 2010)
7. Marsden, P.V.: Egocentric and sociocentric measures of network centrality. Social Networks 24(4), 407–422 (2002)

Comparative Analysis of Graphic Contents Rendering Techniques in a Multi-view System through Agent-Mediator Based Communication

Fahad, Muhammad Azhar, Muhammad Sajjad, Irfan Mehmood, Soon Il Kwon, Jong-Weon Lee, and Sung-Wook Baik*

Digital Contents Research Institute, Sejong University, Seoul, Republic of Korea
{fahad,azhar3797,sajjad,irfanmehmood}@sju.ac.kr,
{sikwon,jwlee,sbaik}@sejong.ac.kr

Abstract. One of the major issues in mixed reality multi-agent systems is synchronization of display, which can adversely affect system performance and hinder user interaction. Real time response from the system cannot be achieved because of the aforementioned issues. If the content displayed on agents is complex and cannot be feasibly rendered on a single agent, then a better strategy is to divide the contents among multiple agents. In this way, only a fraction of the entire contents is rendered by each agent. In this paper, two alternative techniques for multi-agent based content management are proposed, namely, full contents on agents (FCOA) and partial contents on agents (PCOA). In FCOA, each agent in the multi-agent system renders all the contents and only a specific part of the contents is displayed depending upon the usage scenario. In PCOA the agents receive partial contents from the mediator. A comparative study has been presented in this paper to identify the pros and cons of each method.

Keywords: Multi-agent system, Adaptive Mixed Reality space, Agent-Mediator based Communication, Mixed Reality.

1 Introduction

In recent years, the rendering power of computer has increased [3], but despite this exceptional growth in rendering power, some applications have complex graphic details that cannot be feasibly rendered on a single computer node [3, 4]. Multi-agent rendering is a useful technique to handle such an issue. In this technique, complex graphical contents are divided intelligently among various agents. Each agent renders its own part, making rendering on multi-agent systems relatively easier with less time complexity [5].

Research has shown that large displays elevate user productivity and engagement with the application, moreover large displays increases user focus and user satisfaction is immensely increased by these displays [7, 8, 9, 10]. Single large display can be expensive and have a limited field of view (FOV). In addition to this, multiple displays are more appropriate in displaying multimedia contents at high resolution

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_73,

and a larger field of view [6]. Therefore, mixed reality systems adopt a multiple displays mechanism to create an interactive virtual environment [2]. Adaptive mixed reality (AMR) space, which is built on an agent-mediator based framework, is an advance application of mixed reality [1].

AMR space uses multiple displays to create an immersive interactive virtual environment. Each agent in AMR space is connected to a smart media wall (SMW) upon which graphic contents are displayed. Agents also collect user interaction data from SMW and send it to the mediator. The Mediator plays a key role in the synchronization of contents on SMW and also helps in the smooth working of the system. The Mediator processes the received interaction data and broadcasts processed data to all agents. The broadcasted data contains an updated interaction model and instructions to display part of the graphical content on SMW. After receiving data from the mediator, each agent displays the frame of the rendered scene specified by the mediator.

In AMR space, complete graphic contents are individually rendered by each agent before being displayed on SMW. The rendering of graphic contents as discussed above, is a computationally expensive task that can slow the agent's performance. Moreover, if the graphic contents are complex, serious performance issues for the system can result, also affecting the display seamlessness.

In this paper, two techniques on content management for the AMR space are proposed: (1) graphic contents rendering via full contents on agent (FCOA) and (2) via partial contents on agents (PCOA). In the case of FCOA, each agent has full contents locally available and the contents are rendered by each agent. This technique is currently used in AMR space, but it suffers from performance issues while dealing with large and complex contents. On the other hand, PCOA divides the contents among agents with the help of a mediator and sends the partial contents to corresponding agent. This rendering mechanism is computationally feasible for agents and makes sure that the system runs smoothly. In the case of network traffic, the FCOA technique outperforms the PCOA technique, as the network bandwidth fir transmitting graphical content can become very high [3]. However, PCOA is less computationally expensive as compared to FCOA. There is a trade-off between computational and communication cost among agents.

The rest of this paper is organized as follows. Section 2 presents system design. Section 3 provides a detailed comparative analysis of the rendering techniques, and finally, section 4 concludes the paper.

2 System Design

In AMR space, the contents displayed are divided into multiple screens, and each screen is controlled by the corresponding agent. AMR space uses the virtual reality (VR) concept for authoring contents. The 3D graphic contents have been composed using content authoring tools. These contents are loaded into the system for display purposes. Multiple agents are required to display the composed 3D scene.

In order to display contents on multiple agents, inter-agent synchronization is important for real time visualization. If each agent works independently without coordination then synchronization cannot be achieved. To resolve this issue, an agent-mediator based framework has been introduced [1]. In this system, the purpose of the mediator is to provide synchronization among various display nodes. The task of the

mediator is to receive the interaction data from all the agents and convert that raw interaction data into useful information. After that, the Broadcaster module of the mediator broadcasts information to the agents. Agents receive the corresponding information and act accordingly.

To achieve a real-time and quality inter-agent synchronization, two architectures have been proposed for content management, i.e. FCOA and PCOA.

2.1 Graphic Contents Rendering via Full Contents on the Agent

In this case, each agent has complete contents of the scene. This assists in rendering the whole scene locally without dividing the contents. The main steps of the FCOA are shown in Fig. 1. Each agent has its own 3D depth sensor (for finding the user's location), camera (for capturing user's gesture, Pose and face) and audio sensor (for user's spoken words) in the SMW. Agent waits to capture data from its input devices (sensors, camera etc.) and with the help of Dispatcher, sends the received data to Mediator for further processing.

The Mediator captures the interaction data from all the agents. The Receiver passes that data to processing unit (PU), which operates on the data captured and converts the data into useful information. PU passes that information to the Broadcaster module, which broadcasts only the relevant and useful information to each agent. The information sent by Mediator contains instructions which tells each agent which part of the rendered content to be displayed on SMW.

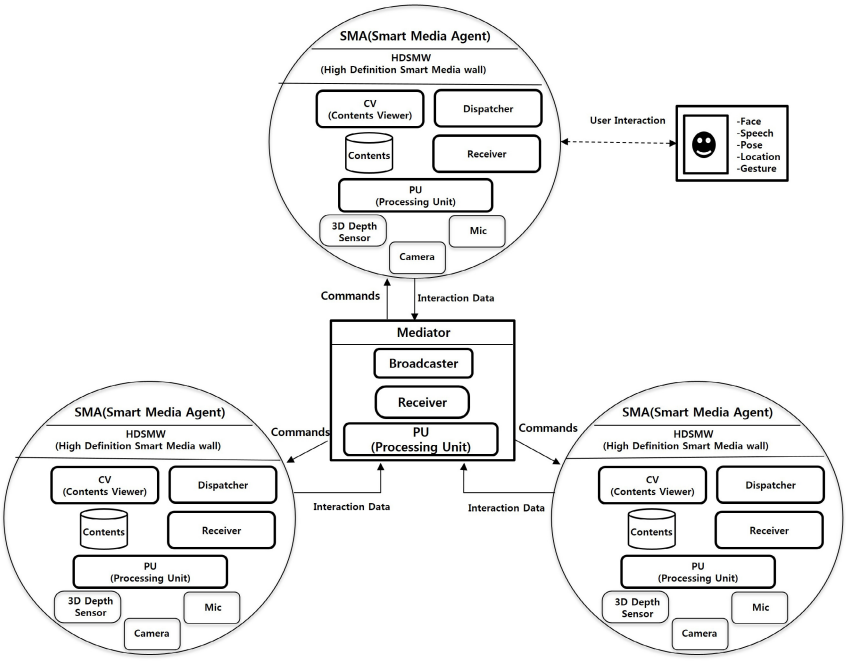


Fig. 1. Graphic Contents Rendering Via Full Contents on the Agents

The Receiver of each agent receives information sent by the Broadcaster. After reception, the Receiver module passes that information to the Content Viewer (CV). As each agent has the entire rendered scene, the CV module of each agent therefore only changes the part of the scene displayed on the SMW in accordance to instructions send by the mediator. In this case, the Agents play a key role as they have to decide which part of the whole scene to display depending upon the command received from the Mediator. To comply with the commands sent by the Mediator, agents display the corresponding part of the scene on their respective SMW in such a way that the whole scene looks synchronized.

2.2 Graphic Contents Rendering via Partial Contents on the Agents

In the case of partial contents, agents work similar to thin clients, i.e. the agent's responsibility is to show received data that is sent by the Mediator. The Framework of the proposed PCOA is shown in Fig. 2. Each agent has its own 3D Depth sensor (for finding the user's location), Camera (for capturing user's gesture, Pose and face) and Audio Sensor (for user's spoken words). The Agent waits for the sensor data from its input devices (Sensors, Camera etc.), while the Dispatcher of corresponding agent sends that unprocessed sensor data to the Mediator. The Mediator receives the interaction data sent by the agents and analyzes it for further processing. The Mediator does not broadcast data to agents until it finds some useful information, which can be in the form of gesture/pose/speech/face information.

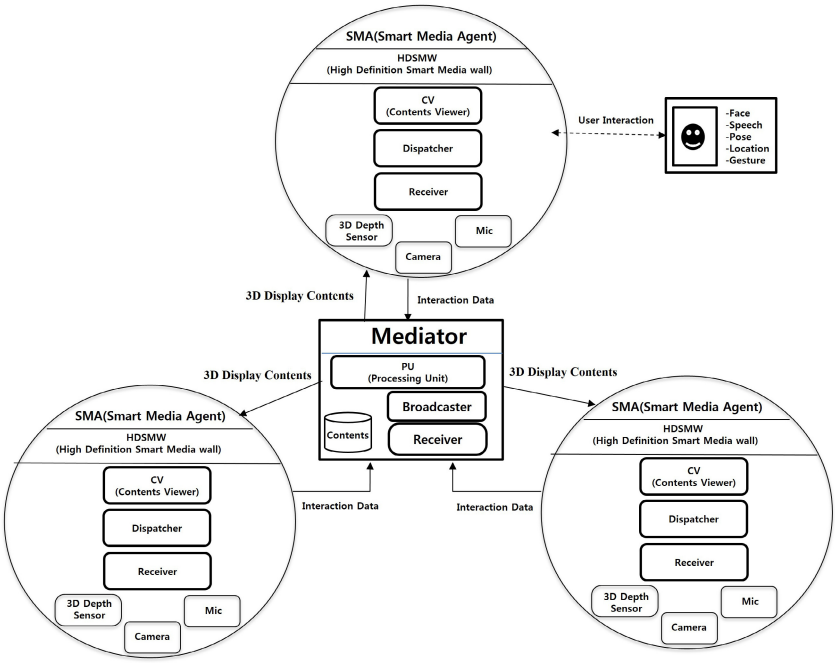


Fig. 2. Graphic Contents Rendering via Partial Contents on the Agents

When it finds required the information, PU of the mediator analyzes the whole interaction data. After analyzing the interaction data, PU splits the scene in such a way that it is synchronized on each agent. After splitting, PU passes partial scenes to Broadcaster, whose task is to broadcast contents to the agents. The Receiver module of the agents receive partial scene. After receiving the information, the agents send acknowledgement to the mediator that the have been received. If the mediator does not receive acknowledgment from some the agent, it sends the corresponding missing data again for consistency. After sending acknowledgment to the mediator, the Receiver module passes received data to CV, whose task is to display the partial scene on SMW. There is 1-1 relationship between agent and SMW i.e. there is only one SMW attach with each agent.

The Agent's task in the case of partial content is reduced as it has to send the interaction data to mediator and display scene on the screen after receiving. Hence, the responsibility of PU is reduced in the case of PCOA.

3 Discussion

In the case of large and complex scenes, rendering becomes a challenging task. In such scenarios, PCOA is the most feasible option as it divides the scene contents intelligently and easily transfers them to corresponding agents. This makes the whole rendering process more efficient, which helps to provide a better experience from the human-computer interaction point of view as shown in Figure 3.

A summary of the comparative analysis of the graphic contents rendering via FCOA and PCOA is shown in Table 1. It shows that FCOA is more feasible by placing the whole scene on each agent and allowing each agent to render the whole scene locally and make them intelligent enough to decide which part of the whole scene to be displayed according to the instructions received from the mediator. In the case of Network traffic comparison, FCOA is more appropriate as agents only receive commands from the mediator. Latency is therefore minimized due to less data transfer. It also minimizes the responsibilities of the mediator. FCOA maintains synchronization by dividing the contents among various agents. In this way, the system behaves like distributed system by incorporating the property of fault tolerance. In case of PCOA, the scene looks synchronized as the contents are pre rendered on each agent. This is why, when the input devices of the agent capture user's input, only part of the scene displayed change is comparatively faster than rendering the scene at runtime.

Another advantage of PCOA is the reduction of processing time on each agent. In PCOA, storage is reduced considerably as compared to FCOA in which complete scene contents have to be placed on each agent. This can cause serious storage issues if the scene under consideration is large and complex.

In the case of FCOA, all the agents have the scene data locally on their machines, which puts an extra burden on agents as they have to collect interaction data and send it to the mediator. On receiving the processed global interaction content from the Mediator, agents have to analyze that data and decide which part of the whole scene to display. This workload on agents can cause serious performance issues.

Table 1. A Comparative analysis of Full Contents and Partial Contents on Agents

	Less Network traffic	Storage Efficiency	Scalability	Extent of Processing
FCOA	YES	NO	NO	NO
PCOA	YES	YES	YES	YES

Fig. 3. Contents shown in physical AMR space

4 Conclusion

In this paper, two contents management techniques (i.e. FCOA and PCOA) have been proposed. After describing the architectures of FCOA and PCOA in detail, a comparative analysis has been made with respect to synchronization and content management. In PCOA, agents behave like thin clients, but in FCOA agents have to render the whole scene by themselves and display the specific part according to the commands sent by the mediator. The contents have to be displayed on SMW. In FCOA, network latency is reduced due to less data transfer as compared to PCOA. But in the case of processing power, PCOA performs better than FCOA as extensive processing for rendering and distributing contents shifted from agents to the Mediator. In AMR space, depending on the context, both architectures can be used alternatively. It has been concluded that if the system is simple and contents are limited, then FCOA is the better choice. In the case of complex and huge contents, PCOA provides a more feasible solution to achieve better performance.

Acknowledgements. This work was supported by the Industrial Strategic technology development program, 10041772, The Development of an Adaptive Mixed-Reality space based on Interactive Architecture, funded by the Ministry of Science, ICT & Future Planning (MSIP).

References

[1] Sajjad, M., Lee, J.J., Gu, B.W., Mehmood, I., Fahad, L.M., Jang, Y., Baik, S.W.: Synchronized and Efficient Communication System based on Agent-Mediator Framework for Adaptive Mixed Reality Space. In: International Conference on ICT for Smart Society ICISS (2013)

[2] Cremer, J., Severson, J., Gelo, S., Kearney, J., Mcdermott, M., Riccio, R.: "This Old Digital City": Virtual Historical Cedar Rapids, Iowa circa 1900. In: Proc. VSMM 2000, pp. 27–34 (2000)

[3] Repplinger, M., Löffler, A., Rubinstein, D., Slusallek, P.: URay: A flexible framework for distributed rendering and display. Technical Report 2008-01, Universität des Saarlandes, Saarbrücken (2008)

[4] Stone, P., Veloso, M.: Multiagent systems: A survey from a machine learning perspective. Autonomous Robots 8, 345–383 (2000)

[5] Gonzalez-Morcillo, C., Weiss, G., Jimenez, L., Vallejo, D.: A Multi-Agent Approach To Distributed Rendering Optimization. In: Proceedings Of The National Conference On Artificial Intelligence, p. 1775 (1999, 2007)

[6] Staadt, O.G., Walker, J., Nuber, C., Hamann, B.: A survey and performance analysis of software platforms for interactive cluster-based multi-screen rendering. In: Proceedings of the Workshop on Virtual Environments 2003, pp. 261–270. ACM (2003)

[7] Czerwinski, M., Smith, G., Regan, T., Meyers, B., Robertson, G., Starkweather, G.: Toward characterizing the productivity benefits of very large displays. In: Proc. Interact 2003, vol. 3, pp. 9–16 (2003)

[8] Czerwinski, M., Tan, D.S., Robertson, G.G.: Women take a wider view. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2002, pp. 195–202. ACM (2002)

[9] Shupp, L., Andrews, C., Dickey-Kurdziolek, M., Yost, B., North, C.: Shaping the display of the future: The effects of display size and curvature on user performance and insights. Human–Computer Interaction 24(1-2), 230–272 (2009)

[10] Navrátil, P.A., Westing, B., Johnson, G.P., Athalye, A., Carreno, J., Rojas, F.: A practical guide to large tiled displays. In: Bebis, G., et al. (eds.) ISVC 2009, Part II. LNCS, vol. 5876, pp. 970–981. Springer, Heidelberg (2009)

A Digital Forensic Model Based on the Generated Fuzzy Number Using FCM Clustering

Seokhwan Yang, Youngjun Son, and Mokdong Chung

Dept. of Computer Engineering, Pukyong National University, Korea
{seokhwan,say4ever,mdchung}@pknu.ac.kr

Abstract. Due to the wide spread use of smartphones, the quantity of information is increasing at an exponential rate. This means that a great deal of time and effort should be dedicated to digital forensic analysis. Therefore, we need faster methods. FCM clustering algorithm is the most popular clustering algorithm, but it has some problems. This paper proposes a digital forensic model which links the digital media with a criminal profiling system by classifying the data using an extended FCM clustering algorithm. The extended FCM clustering algorithm provides generated fuzzy numbers by the result of FCM clustering algorithm.

Keywords: Digital Forensics, FCM Clustering Algorithm, Fuzzy Number, Context-Awareness.

1 Introduction

Digital Forensics is a concept which refers to a set of procedures and methods to find digital evidence associated with the criminal case by collecting and analyzing the data of digital media [1]. Recently, due to the increase of digital information, a great deal of time and effort should be dedicated to digital forensic analysis. Since digital data is extremely susceptible to modulation, great scale and high volatility, the rapidity of forensic analysis is critical.

Many researchers are using FCM clustering algorithm which is the most popular for classifying data flexibly. However, the classical FCM clustering algorithm has some problems such as high sensitivity to noise and input data, the different result from the intuitive grasp, and the setting of initial round and the number of clusters.

This paper proposes a digital forensic model which links the digital media with a criminal profiling system by classifying the data from digital media using an extended FCM clustering algorithm. The extended FCM clustering algorithm provides generated fuzzy numbers by the result of FCM clustering algorithm.

Section 2 reviews related work. In section 3, we introduce a model of fuzzy number generation using result of FCM clustering. In section 4, we introduce a digital forensic model using the proposed clustering techniques and criminal profiling systems. Section 5 draws conclusions and discusses the directions of our future work.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_74,

2 Related Work

2.1 FCM (Fuzzy C-Means) Clustering Algorithm

FCM [2] is the data clustering algorithm that uses Fuzzy division technique and the membership function U which may have value between 0 and 1. The sum of the value of membership function for the data set is always 1 [3]. The objective function for FCM clustering may have the following formula.

$$J(u_{ik}, v_i) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m}(d_{ik})^2, \qquad d_{ik} = d(x_k - v_i) = \left[\sum_{j=1}^{1}(x_{kj} - v_{ij})^2\right]^{\frac{1}{2}}$$

$$v_{ij} = \frac{\sum_{k=1}^{n}(u_{ik})^m x_{kj}}{\sum_{k=1}^{n}(u_{ik})^m}, u_{ik} = \frac{1}{\sum_{j=1}^{c}(\frac{d_{ik}}{d_{jk}})^{\frac{2}{m}-1}}$$

u_{ik}: Degree of membership of the k_{th} data of the x_k, that belongs to the i_{th} cluster

v_i: Centroid vector of i_{th} cluster

m: Parameter for control the volume of the Fuzzy characteristics. m=2 in general.

d_{ik}: Distance between k_{th} data of the x_k and i_{th} cluster's centroid v_i

$j_{(u_{ik},v_i)}$: Objective function for the FCM clustering

2.2 Digital Forensics

The Digital Forensics Research Workshop (DFRWS) defines the digital forensic as an act of preserving, collecting, validating, identifying, analyzing, documenting, and presenting digital evidence in a scientific and provable way. The general procedure of the digital forensic consists of collection, examination, analysis and reporting on digital evidence [4].

2.3 Criminal Profiling

Criminal profiling is a kind of investigation techniques applied to infer the characteristics of criminals through various actions of offenders, crime scenes or the patterns of daily life that contribute to arresting the criminals [5]. It offers criminal tendencies, behavioral patterns, types of crime, and overall understanding of the case, thus providing useful information to crimes which are becoming more intelligent.

3 Generation of Fuzzy Number Using FCM Clustering Algorithm

FCM clustering algorithm is a method to find a desired membership function by performing repetitive operations to minimize the object function. The classical FCM

clustering algorithm is based on the degree of membership and the Euclidean distance between input vector and centroid vector of cluster. Therefore, it has some problems, such as high sensitivity to noise and local data and the different clustering results from the intuitive grasp.

Our model generates fuzzy number using the range of cluster from the result of FCM clustering algorithm. The generated fuzzy number is based on the generation of fuzzy rules. Figure 1 shows generation process of fuzzy rules in the suggested model.

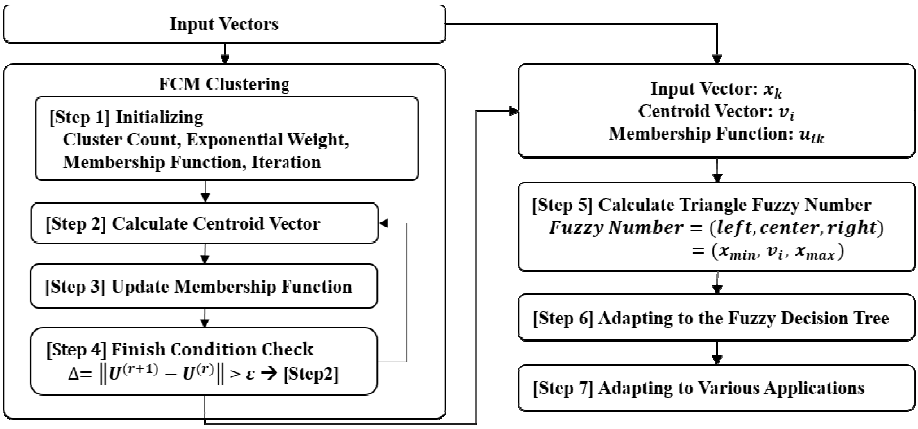


Fig. 1. Generation process of fuzzy rules in the suggested model

The results of FCM clustering are centroid vector of the cluster and membership function for input data. The suggested model generates triangle fuzzy numbers using the centroid vector of each cluster by the Eq. (1). Algorithm of the proposed model is shown in Table 1.

$$\text{Fuzzy Number} = (x_{min}, v_i, x_{max}),\ \ x \in cluster(i) \tag{1}$$

Table 1. Algorithm of the Proposed Clustering Model

```
function SystemClustering(){
  res[x] =FCM(data[x]);
  fuzzy_number[]=GenFuzzyNumber(res[x], data[x]);
  lv[x]=FuzzyRule(fuzzy_number[c], data[x]);
  executeInfoService(lv);
}

function FCM(data [x]){
  initialize();
  while(i<x){
    res[i]=getCenterOfCluster(data[x]);
```

```
    membership_ft=updateMembershipFt(res[x]);
    while(delta>e) repeat;
  }
  return res[x];
}
function GenFuzzyNumber(res[x], data[x]){
  while(i<clusters){
    while(j<dimension){
      fuzzy_number[i][j]=(data[x][j].min,
      c_vector[j], data[x][j].max);
    }
  }
}
```

4 Improved Digital Forensic Model

4.1 Architecture and Algorithm of the Proposed Model

In this paper, we propose an effective digital forensic model which is linked to the criminal profiling system using extended FCM clustering algorithm. The digital forensic procedure consists of four major stages of collection, examination, analysis, and reporting [4]. Before extracting data from the digital forensic examination phase, the proposed digital forensic model classifies the data that exists in the digital media by using a proposed clustering algorithm. Then the secondary data obtained from clustering is applied to the system of criminal profiling. The secondary data and the clues from profiling analysis are used as information for the investigation, and they give feedback to each stage of digital forensics to improve its efficiency. Figure 2 shows the architecture of an improved digital forensic model using proposed clustering model.

Specific steps are as follows: Firstly, the digital media and images were obtained in accordance with legal procedures of the general digital forensic procedures. The goal of forensic work was set to extract the information associated with the criminal case.

When a particular data to extract is set as target data such as specific document files, photos, videos, etc., the target data is extracted and analyzed using the file search, keywords, and hash value.

In contrast, when the entire digital data has to be checked because of the absence of cues or deny of the crimes, we apply the proposed model which provides the automatic clustering of the data that exists in the digital media.

4.2 Clustering and Variables for Criminal Profiling

We select the input data from the data saved on the smartphone for clustering because smartphone is very likely to have many different kinds of data formats such as the user's patterns of behavior, tendency, hobbies, etc. Three criteria are applied for

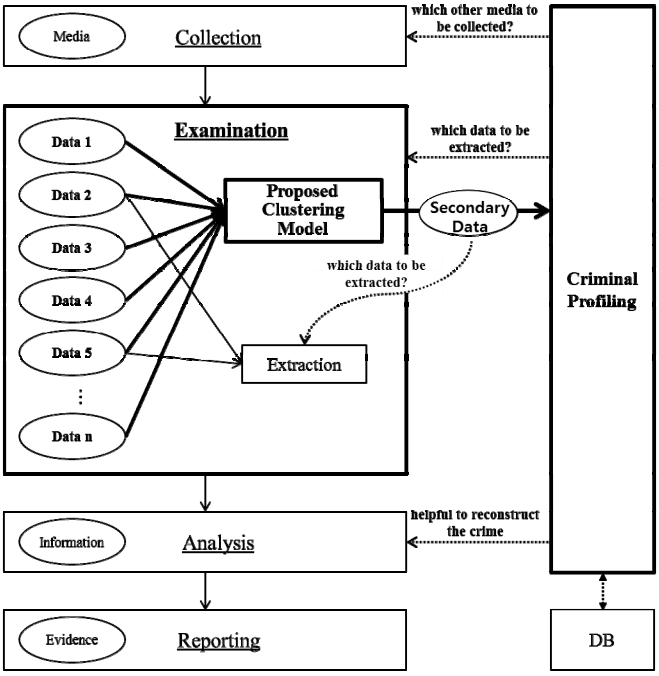


Fig. 2. Architecture of the Improved Digital Forensic Model using Proposed Clustering Model

Table 2. Algorithm of the Improved Digital Forensic Model

```
Function examination(data){
   target = check(data);
   if(target == true)
      information = extract(target);
   else {
      secondary = Clustering(data);
      clue = profiling(secondary);
      meaningfuldata = secondary + clue;
      info = extract(discover(meaningfuldata));
   }
   return information;
}
Function Clustering(data){
   produce secondary data using proposed FCM clustering
   algorithm;
}
Function profiling(data){
   find information and clues about a criminal case;
}
```

selecting input data: The first criterion is the relationship between the input data and predictable behavior pattern of the user. The second one is the range of the input data since a piece of data may yield different results depending on how it is approached. The third one is the variables used in criminal profiling to apply that system.

5 Conclusions

After integrating the techniques developed in the fields of digital forensics, clustering and criminal profiling, we proposed an improved digital forensic model which is linked to the criminal profiling system by using extended FCM clustering technique. In the proposed model, we classified the data in the smartphone using the proposed clustering algorithm to gather clues for investigation. This improved the efficiency of the process of digital forensic analysis. And by offering useful information for a criminal case, the proposed model might accelerate the speed of investigation by systematizing criminal profiling data. Because the proposed clustering algorithm automatically classified the data using the similarities of each data in the smartphone, we expect to obtain sufficient valid results. This means that the proposed model might contribute to resolving some of the limitations of temporal resources for digital forensics by providing the necessary data for criminal profiling, such as finding clues through profiling, understanding of crime, and accumulation of criminal data.

In the future research, we are planning to develop the system into the general system that can be applied to the general domains.

Acknowledgement. Following are results of a study on the "Leaders INdustry-university Cooperation" Project, supported by the Ministry of Education, Science & Technology (MEST).

References

1. A Road Map for Digital Forensic Research. DFRWS Technical Report, p. 16 (November 2001)
2. Bezdek, J.C., Ehrlich, R., Full, W.: FCM: The Fuzzy c-Means Clustering Algorithm. Computers & Geosciences 10(2-3), 191–203 (1984)
3. Oh, S.-K.: Computational Intelligence by Programming focused on Fuzzy, Neural Networks, and Genetic Algorithms. Naeha Publishing Co. (2002)
4. Kent, K., et al.: Guide to Integrating Forensic Techniques into Incident Response. NIST Special Publication 800-86, pp. 1–7 (September 2006)
5. Park, J., Choi, N.: Perception of Offender Profiling and its Development Strategies. Thesis Journal of Korea Contents Association 12(6), 413–423 (2012)

Development of a PC-Based Code Simulator for Verifying Ubiquitous Embedded Software

Sooyong Jeong, Sunghee Lee, and Woo Jin Lee

School of CSE, Kyungpook National University, Daegu, Republic of Korea
kyo1363@naver.com, lee3229910@gmail.com, woojin@knu.ac.kr

Abstract. Embedded system is playing a significant role in the ubiquitous computing. Therefore, verification of embedded software is necessary and important. In order to verify embedded software, a software developer has to prepare a running environment since embedded software is designed for the particular target system and it needs compatible hardware and its running environment. The development of its running environent may be costly and time-consuming job. Therefore, it is not easy to run and verify embedded software in the ealier development stage. To solve this problem, we develop a code simulator running on PC for embedded software. The simulator analyzes the embedded software code and makes it compatible with PC to run and verify by simulating the target system.

Keywords: Code simulation, virtual prototyping, embedded system.

1 Introduction

Nowadays ubiquitous computing is applied to various industrial fields and embedded system is playing a leading role in ubiquitous computing industry. As embedded systems are being used in human life more and more, reliability and safety issues of the systems are more concerned. In the case of safety-critical systems, it's very important to verify the working of those systems properly because even a small error or failure can cause devastating effect on environment or human life. The majority of accidents in embedded system arises from software error or failure, for example the Ariane 501 satellite failure on June 1996 [1].

However, developers often face many problems in verifying the embedded software mainly due to platform and hardware dependency of the embedded software. For instance, the embedded software may be using different processors with PC, using particular hardware dependent code, and so on. When some of the embedded platform provides its own simulator such as Android Emulator [2], a developer can verify the software on PC by using the given simulator. But most of embedded platforms don't provide any kind of simulator, so the developer needs to prepare the target system, run the code on the system directly, and then verify the software, which is costly and time-consuming job. The verification through executing the software in real environment takes significant amount of time and cost. Therefore, the verification of embedded software is still difficult mission for the developers.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_75,

In this paper, we propose and develop a code simulator for executing embedded software on PC to verify them. The proposed approach identifies hardware dependent parts by analyzing source code and converts those parts into the stubs without changing the original source code. After the conversion into virtual drivers, we get a code simulator runnable on PC. By executing the simulator, we can run and check the embedded software code on PC instead of its real target system. Our code simulation approach makes the verification process begin earlier and also reduces the verification time and cost without developing target system and environment.

The rest of the paper is organized as follows. Section 2 introduces the existing techniques about executing embedded software on PC. Section 3 presents the developing processes of the code simulator in detail. A case study of our approach is provided in section 4. Section 5 concludes this paper.

2 Related Work

2.1 Emulation

Emulation is a technique that imitates a specific device in a different running environment. An emulator emulates the target hardware architecture and provides a running environment for user-installed software. For example, Android Emulator [2] is a famous one used in the software verification. It emulates Android [3] operating system, ARM CPU, displayer, their I/O device. Using the emulator, developers can run Android applications in their own PC. And this emulator can be used in software verification of embedded system. But many embedded platforms don't provide their own emulators because developing and providing the emulator like Android Emulator is difficult and may cost much higher than providing target system in the developers' viewpoint. Due to this reason, emulators are used in limited areas to verify the embedded software.

2.2 Code Simulation

Code simulation can be used to show the behavior of the software in host computers. It runs target code by using host system's API rather than imitating target hardware. So code simulation has less hardware dependency than emulation approach, which makes it easier to develop a code simulator than an emulator. For example, FreeRTOS [4] provides its code simulator [5], which uses Windows' multithreading to implement FreeRTOS tasks. However, there are some weaknesses in using only code simulation in the perspective of user interface. Generally emulation covers the most of target hardware including input and output. So emulator users can utilize the interface like real hardware, while code simulation mainly focuses on the logic of software rather than hardware. Some code simulator makers doesn't add the efficient interface like graphics because it is not mandatory for them. All of these make difficult for a user to simulate with the code simulator.

2.3 Virtual Prototyping

Virtual prototyping simulates a physical product in host computer by making digital mock-up. The digital mock-up in host computer is called virtual prototype which is used to analyze and test the product without real physical model [6]. In practice, it's not so easy to implement the virtual prototyping since the development process and the tools of virtual prototyping are few. Many developers prefer physical prototyping to virtual prototyping but the trend is slowly moving to virtual prototyping.

3 Development of a Code Simulator for Embedded Software

We propose and develop a code simulator which support to simulate embedded software with virtual drivers on PC in the earlier developement phase. An overall developement process of the code simulator is illustrated in Figure 1. At first, we analyze the embedded source code to gather hardware dependent parts, then convert those parts to PC based simulation stubs called virtual drivers in order to run them on PC without real hardware devices. In order to efficiently handle the inputs and outputs on PC, the converted simulation code is combined with I/O virtual prototype in the form of GUI. Finally, we can get a code simulator which is runnable on PC.

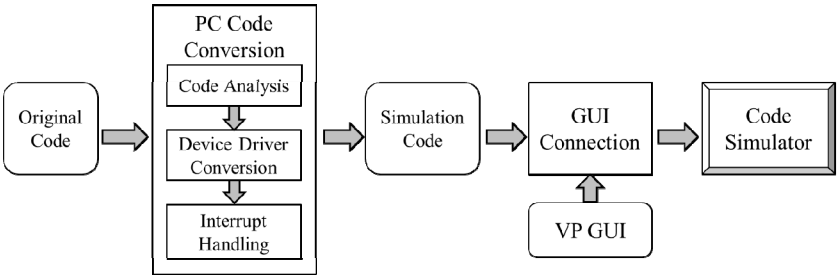


Fig. 1. Overall Development Process of the Code Simulator

In the embedded system, it is not easy to support a common code simulator, since there are various kinds of hardware specification such as the system's main chipset, I/O device connection approach, programming style, and so on. Therefore, we assume that the target system should have following constraints. At first, the target embedded software should be layered architecture that shields lower layer from direct access by higher layers [7]. All the layers interact with each other through the function calls. Figure 2 shows an example of how the embedded software runs in layered architecture. For example, the application in top layer calls CtrlFanRPM() to decide the fan speed and the middleware consecutively calls DriveFan() to run the fan in the device. In the layered architecture, There are some restrictions in using low-level registers. We allow register access only when the register is used only for data input/output or buffer. At second, the target embedded software should only use interrupt handling mechanism to process input and output data. In the other word, machine check interrupt and internal interrupt in the software is not allowed.

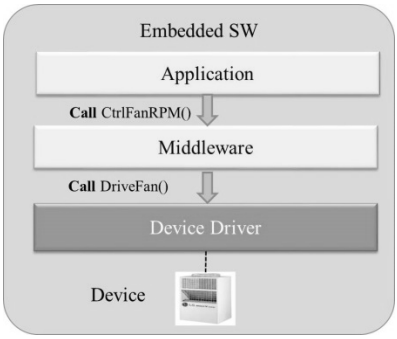


Fig. 2. An Example of Interaction on Layered Embedded Software

3.1 Virtual Conversion of Device Drivers

To simulate embedded software, we analyze the code to identify hardware dependent parts like device drivers. Since running the embedded software with the real device drivers intactly on PC is nearly impossible, through code analysis, we identify and convert the device drivers to the virtual drivers which are code stubs running on PC as shown in Figure 3. Since most of embedded software are written in C language, C parser is needed for analyzing the source code. We choose the C parser generated by ANTLR [8] since it is the open-source parser generator and there is ANSI C grammar can used in them.

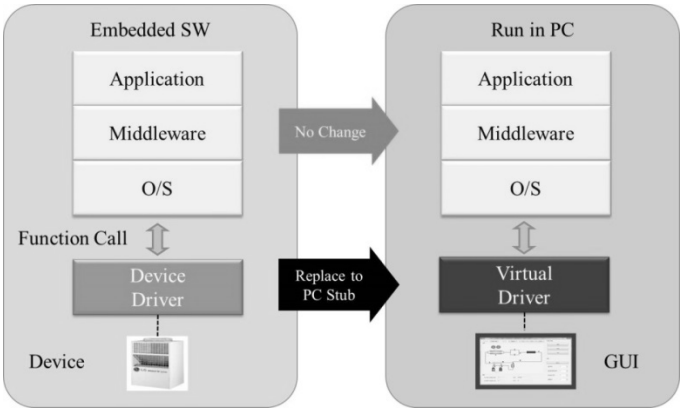


Fig. 3. The Virtual Conversion of Device Drivers

The first step of code analysis is to identify and remove its device drivers. We can remove device drivers only because we supposed that the target software has layered architecture. After parsing the remaining code, we identify which functions and variables are referred but missed. The missing ones imply hardware dependent parts since the device driver parts are removed from the original code before the parsing.

Then the missed parts are converted to the virtual drivers which have a role of the device drivers to run the software on PC and have a mission to connect the simulated code to VP GUI. The stub code of virtual drivers are generated according to the software's specification and I/O features.

3.2 JNI Connection between VP GUI and Simulation Code

In order to change the embedded software to work without its own hardware, it's necessary for a user to control the code simulator, to inject the stimulus to the software through GUI. Through the virtual prototype GUI, users can handle I/O interaction and environment interaction during simulation in their own hands. In designing the structure of the simulator, we consider making the interface in the C language since the most of embedded software is developed in C. It is convenient to run together C based user interface and C based simulation code. However, since making graphic user interface in C is not easy and inefficient and two parts are so tightly coupled, it is difficult to manage further changes or porting request. Alternatively, it is possible to provide a loosely-connected structure by separating the simulation code and GUI part, which provides a user preference in choosing GUI implementation language. We choose Java as the GUI implementation language and Java Native Interface(JNI) [9] for connceting two separating parts as shown in Figure 4. JNI allows running native C code at Java program and sharing the memory between C simulation code and Java user interface.

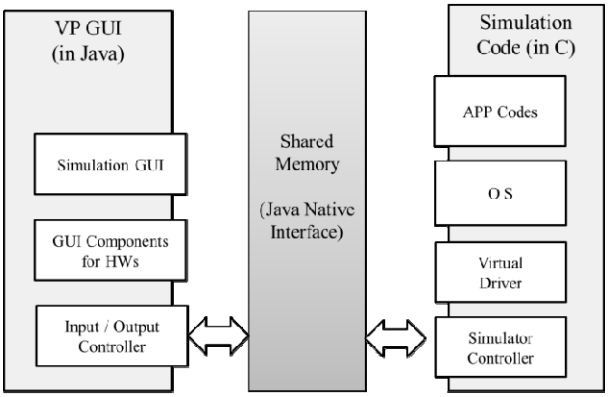


Fig. 4. Interaction of Java Graphic User Interface and Converted C Code

At first, a simulation user makes the interface as a virtual prototype of hardware in Java. And the simulator gets the variables to share with virtual prototype from simulation code. The shared variables are managed in the shared memory by JNI. The library is included to both virtual prototype and simulation code, so Java based virtual prototype and C based simulation code get access to the shared memory and have connection to each other.

3.3 Interrupt Handling of Simulation Code

A number of embedded software make frequent use of interrupts to indicate the events from hardware or software such as input/output, and the program called interrupt handler is used to deal with the interrupt. So interrupt and interrupt handling are considered to essential to embedded software. However, a new interrupt handling approach is needed to simulate embedded software on PC since they are dependent to hardware or operating system. Our approach utilize multithreading in the host PC's API. We attach a module called launcher and a new interrupt handler for PC to the simulation code. At first, in the target code, the definitions and prototypes of interrupt should be changed to the function of host PC's API. It will be able to perform interrupt of target code as like an interrupt of host PC. This kind of interrupt conversion can be automated because the definition form of interrupt is fixed. For example, an external interrupt is defined to 'SIGNAL (SIG_INTERRUPT1)' in Atmel AVR, it will be changed to 'unsigned int WINAPI SIG_INTERRUPT1' in simulation code by using Windows API.

The launcher is used to start the main code and the interrupt handler as concurrent threads. Then they run repeatedly in loops until the software is terminated. The interrupt handler monitors the occurence of interrupts and deal with them in simulation. When input/output or events are occured, it suspends the main logic thread and create a new thread for the interrupt. The interrupt handler waits until the interrupt thread ends, then resume the main logic to continue the program. In Figure 5 a sequence diagram shows how an interrupt is handled in the simulation.

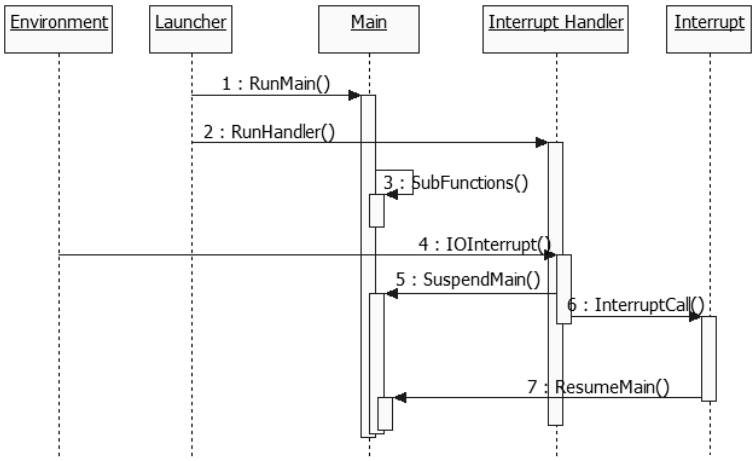


Fig. 5. A Sequence Diagram for Interrupt Handling

4 Case Study of Air Conditioners

In this section we have implemented a code simulator for an air conditioning system as case study. When the target code of embedded software is given, we make a Java

based virtual prototype for substituting the target hardware instead of preparing the real hardware. The user interface includes the stimuli to hardware and input/output values. Figure 6 shows a snapshot of the user interface to connect the simulation code.

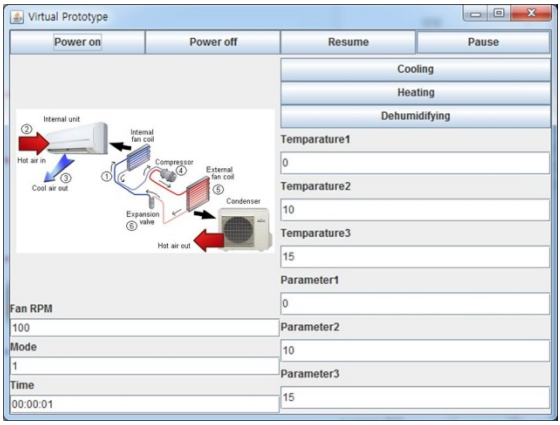


Fig. 6. An Example of Code Simulator for Air Conditioner

When the virtual prototype is prepared, we analyze the target source code to convert the device driver and interrupt to virtual ones. In this phase, we get the simulation code where its device driver and interrupt is replaced. Finally we connect the simulation code to the virtual prototype with JNI and run it on PC.

5 Conclusion and Future Work

We proposed a method of code simulation in embedded software written in C and developed a PC based code simulator. The developed code simulator can be used for showing how the embedded software in layered archiecture works on PC for simulating the software in the earlier phase. Since the code simulator has loosely connected structure of GUI part and simulation code by JNI, it is relatively easy to perform changes and porting to other platforms or to enhance its funtionality. In order to show the availability of our code simulation approach, we performed a case study to air conditioning system.

In future, we intend to reduce the constraints of target code in our simulation and to fully automate the code conversion of hardware dependent parts. We also measure and minimize the overhead in order to improve speed of simulation.

Acknowlegement. This work was supported by the IT R&D program of MISP (Ministry of Science, ICT & Future Planning)/KEIT. [10041145, Self-Organized Software platform (SoSp) for Welfare Devices] and the MSIP, Korea, under the C-ITRC (Convergence Information Technology Research Center) support program (NIPA-2013-H0401-13-1005) supervised by the NIPA (National IT Industry Promotion Agency).

References

1. Dowson, M.: The Ariane 5 software failure. In: ACM SIGSOFT Software Engineering Notes, vol. 22(2), p. 84. ACM, New York (1997)
2. Android Emulator, http://developer.android.com/tools/help/emulator.html
3. Android, http://www.android.com/
4. FreeRTOS, http://www.freertos.org/RTOS.html
5. FreeRTOS Windows Simulator, http://www.freertos.org/FreeRTOS-Windows-Simulator-Emulator-for-Visual-Studio-and-Eclipse-MingW.html
6. Wang, G.: Definition and Review of Virtual Prototyping. Journal of Computing and Information Science in Engineering 2(3), 235 (2003)
7. Raut, A.: Pattern oriented software architecture:a system of patterns, p. 35. Wiley, New York (1999)
8. ANTLR, http://www.antlr.org/
9. Liang, S.: The Java Native Interface:Programmer's Guide and Specification, p. 5. Addison-Wesley Professional, Boston (1999)

Distinguishing Attack on SDDO-Based Block Cipher BMD-128

Jinkeon Kang[1], Kitae Jeong[1], Changhoon Lee[2,*], and Seokhie Hong[1]

[1] Center for Information Security Technologies, Korea University,
Anam-dong, Seongbuk-gu, Seoul, 136-713, South Korea
{jinkeon.kang,kite.jeong}@gmail.com, shhong@korea.ac.kr
[2] Department of Computer Science and Engineering, Seoul National University of Science and Technology, Gongneung-ro, Nowon-gu, Seoul, 139-743, South Korea
chlee@seoultech.ac.kr

Abstract. BMD-128 is a 128-bit block cipher with a 256-bit secret key and is based on switchable data-dependent operations. By using these operators, this algorithm was designed to ensure the high applicability in the transaction needing the change of session keys with high frequency. In this paper, we show that it is possible to distinguish between a 7-round reduced BMD-128 and a 128-bit random permutation under a related-key amplified boomerang attack scenario. This result is the first known cryptanalytic result on BMD-128.

Keywords: BMD-128, Switchable data-dependent operation, Block cipher, Cryptanalysis, Distinguishing attack.

1 Introduction

Recently, the design of block ciphers based on data-dependent permutation(DDP) has attracted much attention to their efficiency. As concrete examples, there are SPECTRH64 [2], CIKS-family(CIKS-1 [16], CIKS-128 [1], CIKS-128H [19]) and Cobra-family(Cobra-S128 [3], Cobra-F64a [3], Cobra-F64b [3], Cobra-H64 [20], Cobra-H128 [20]). Since all of them use very simple key schedules in order to avoid time-consuming key preprocessing, they are suitable for use in networks which require high-speed encryption and frequent key changes. However, most of them have been cryptanalyzed because of the linearity of DDP and simple key scheduling algorithms [7-13].

For enhancing the security of DDP-based ciphers, several ciphers have been proposed such as Eagle-64 [18], Eagle-128 [17], MD-64 [14], which are based on data-dependent operations(DDO), and SCO-family [15], which is based on controlled operational substitution(COS). However, they were shown to be still vulnerable to related-key attacks [4-6].

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_76,

BMD-128 [21] is a 128-bit block cipher with a 256-bit secret key and is based on switchable data-dependent operations(SDDO). The number of rounds is 8. The authors of BMD-128 evaluated that this algorithm has the efficient security against statistical analysis and differential cryptanalysis, and the good performance in fast and highly efficient telecommunication systems.

In this paper, we show that it is possible to distinguish between a 7-round reduced BMD-128 and a 128-bit random permutation under the related-key amplified boomerang attack. In detail, we can construct a related-key amplified boomerang distinguisher on a 7-round reduced BMD-128 with a probability of 2^{-152}. Thus, if we can get 2^{79} related-key chosen plaintexts, our attack algorithm can distinguish between a 7-round reduced BMD-128 and a 128-bit random permutation. This result is the first known cryptanalytic result on BMD-128.

This paper is organized as follows; In Section 2, we describe BMD-128 briefly. In Section 3, we propose a distinguishing attack on a 7-round reduced BMD-128. Finally, Section 4 concludes this paper.

2 Description of BMD-128

To begin with, we introduce some notations used throughout the paper. Bits are indexed from left to right. If $P = (p_1, p_2, \cdots, p_n)$, p_1 is the most significant bit and p_n is the least significant bit.

- $e_{i,j}$: 32-bit binary string in which the i-th bit and j-th bit are one and the others are zeroes, e.g., $e_{1,3} = (1,0,1,0,\cdots,0)$.
- ΔI_r: the input difference in round r.
- ΔQ_r, ΔU_r: the round key difference in round r.

BMD-128 is 128-bit block ciphers with 256-bit secret key and the number of rounds is 8. Since our attack works in the chosen plaintext attack scenario, we consider only the encryption procedure. The following is the encryption procedure of BMD-128.

1. 128-bit plaintext is divided into two 64-bit subblocks L and R.
2. For r = 1 to 7 do:
   (a) $(L, R) = Crypt^{(0)}(L, R, U_r, Q_r)$,
   (b) $(L, R) = (R, L)$.
3. Perform the transformation:
   (a) $(L, R) = Crypt^{(0)}(L, R, U_8, Q_8)$.
4. Perform the final transformation:
   (a) $(L, R) = (L \oplus U_9, R \oplus Q_9)$.

In the encryption process, the round function $Crypt^{(0)}$ of BMD-128 is described in Figure 1-(a). For the more detailed description of BMD-128, see [21].

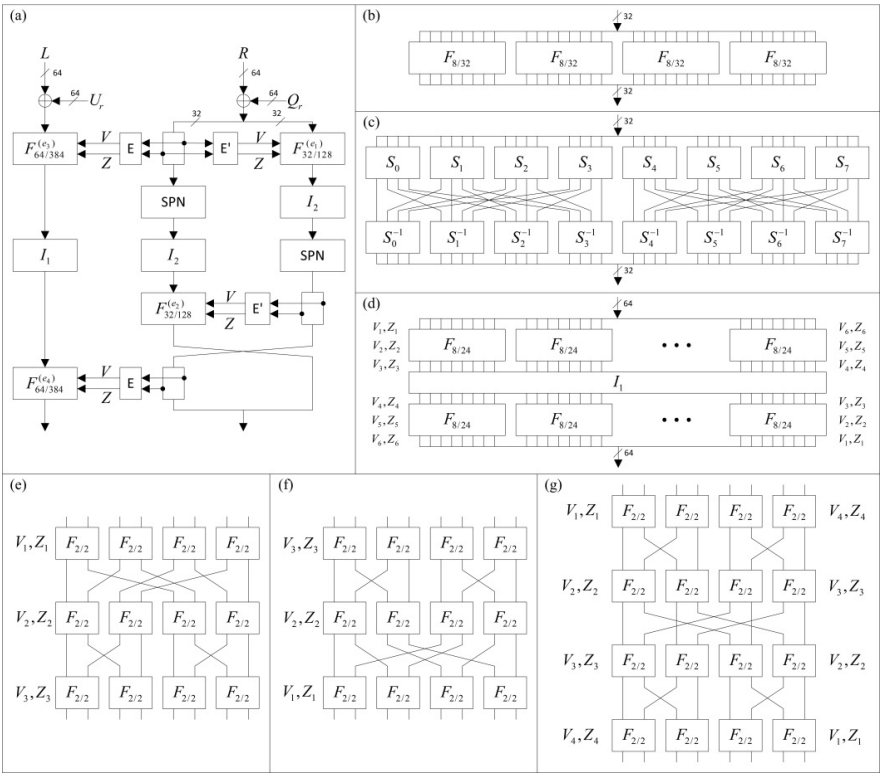


Fig. 1. (a) Round function $Crypt^{(0)}$, (b) $F_{32/128}$, (c) SPN, (d) $F_{64/384}$ and $F_{64/384}^{-1}$, (e) $F_{8/24}$, (f) $F_{8/24}^{-1}$, (g) $F_{8/32}$ and $F_{8/32}^{-1}$

First, the fixed permutation I_1 and I_2 are defined as follows.

$$I_1 = (1)(2,9)(3,17)(4,25)(5,33)(6,41)(7,49)(8,57)(10)(11,18)(12,26)(13,34)(14,42)(15,50)(16,58)(19)(20,27)(21,35)(22,43)(23,51)(24,59)(28)(29,36)(30,44)(31,52)(32,60)(37)(38,45)(39,53)(40,61)(46)(47,54)(48,62)(55)(56,63)(64)$$

$$I_2 = (1)(2,9)(3,17)(4,25)(5)(6,13)(7,21)(8,29)(10)(11,18)(12,26)(14)(15,22)(16,30)(19)(20,27)(23)(24,31)(28)(32)$$

As shown in Fig. 1-(a), BMD-128 uses two extension boxes E and E'. An extension box E' outputs a 128-bit controlling vector $(V, Z) = (V_1, V_2, V_3, V_4, Z_1, Z_2, Z_3, Z_4)$ with a 32-bit input value (A, B) by using the following equations.

$$V_1 = A, \quad V_2 = A^{\lll 8}, \quad V_3 = B^{\lll 8}, \quad V_4 = B,$$
$$Z_1 = A^{\lll 6}, \quad Z_2 = A^{\lll 12}, \quad Z_3 = B^{\lll 12}, \quad Z_4 = B^{\lll 6}.$$

Similarly, an extension box E outputs a 384-bit controlling vector $(V, Z) = (V_1, V_2, V_3, V_4, V_5, V_6, Z_1, Z_2, Z_3, Z_4, Z_5, Z_6)$ with a 32-bit input value (A, B) by using the following equations.

$$V_1 = A \parallel B^{\lll 2}, \quad V_2 = A^{\lll 4} \parallel B^{\lll 6}, \quad V_3 = A^{\lll 8} \parallel B^{\lll 10},$$
$$V_4 = B^{\lll 8} \parallel A^{\lll 10}, \quad V_5 = B^{\lll 4} \parallel A^{\lll 6}, \quad V_6 = B \parallel A^{\lll 2},$$
$$Z_1 = A^{\lll 6} \parallel B^{\lll 4}, \quad Z_2 = A^{\lll 12} \parallel B^{\lll 8}, \quad Z_3 = A \parallel B^{\lll 12},$$
$$Z_4 = B \parallel A^{\lll 12}, \quad Z_5 = B^{\lll 12} \parallel A^{\lll 8}, \quad Z_6 = B^{\lll 6} \parallel A^{\lll 4}.$$

As shown in Fig. 1, $F_{32/128}$ and $F_{64/384}$ are constructed by using an elementary box $F_{2/2}$. $F_{2/2}$ is defined as follows.

– $F_{2/2}((x_1, x_2), v, z) = (y_1, y_2)$.
- $y_1 = vzx_1 \oplus v\boldsymbol{x} \oplus zx_1 \oplus zx_2 \oplus v \oplus x_1 \oplus 1$.
- $y_2 = vzx_2 \oplus vz \oplus vx_1 \oplus zx_1 \oplus zx_2 \oplus v \oplus z \oplus x_2 \oplus 1$.

We checked that, from the above equations, x is not defined in [21] and these equations do not satisfy differential probabilities as shown Table 1 in [21]. So, we decided to modify these equations reasonably. The modified equations are defined as follows. In detail, x was modified to z.

– $F_{2/2}((x_1, x_2), v, z) = (y_1, y_2)$.
- $y_1 = vzx_1 \oplus v\boldsymbol{z} \oplus zx_1 \oplus zx_2 \oplus v \oplus x_1 \oplus 1$.
- $y_2 = vzx_2 \oplus vz \oplus vx_1 \oplus zx_1 \oplus zx_2 \oplus v \oplus z \oplus x_2 \oplus 1$.

A function SPN consists of eight different 4×4 S-boxes $S_0, S_1, \cdots, S_7$. However, they were not defined in [21]. So, we reasonably selected eight 4×4 S-boxes used in [17]. They are defined in Table 1.

Table 1. 4×4 S-boxes S_0, S_1,$\cdots$, S_7

S_0	14	4	13	1	2	15	11	8	3	10	6	12	5	9	0	7
S_1	3	13	4	7	15	2	8	14	12	0	1	10	6	9	11	5
S_2	10	0	9	14	6	3	15	5	1	13	12	7	11	4	2	8
S_3	1	4	11	13	12	3	7	14	10	15	6	8	0	5	9	2
S_4	10	6	9	0	12	11	7	13	15	1	3	14	5	2	8	4
S_5	11	8	12	7	1	14	2	13	6	15	0	9	10	4	5	3
S_6	10	15	4	2	7	12	9	5	6	1	13	14	0	11	3	8
S_7	1	15	13	8	10	3	7	4	12	5	6	11	0	14	9	2

The key schedule of BMD-128 is quite simple. The 256-bit secret key K is divided into four 64-bit subkeys K_1, K_2, K_3, K_4 and they entered into round function $Crypt^{(0)}$ as the specified order in Table 2.

Table 2. Key schedule of BMD-128

Round r	1	2	3	4	5	6	7	8	FT
Q_r	K_1	K_3	K_1	K_2	K_3	K_4	K_3	K_4	K_1
U_r	K_2	K_4	K_3	K_4	K_3	K_2	K_1	K_3	K_2
e_1	1	0	1	1	0	1	0	0	–
e_2	0	0	1	0	1	1	0	1	–
e_3	0	0	0	1	0	1	1	0	–
e_4	0	1	1	0	1	0	0	0	–

3 Distinguishing Attack on a 7-Round Reduced BMD-128

In this section, we introduce a distinguishing attack on a 7-round reduced BMD-128. Recall that the key schedule of BMD-128 is quite simple. Due to the simplicity of it, there are many useful properties which allow us to construct a related-key amplified boomerang distinguisher with a high probability.

To construct a related-key amplified boomerang distinguisher on a 7-round reduced BMD-128, we use the following properties. Here, X is a 64-bit input value of $F^{(\cdot)}_{64/384}$ and (V, Z) is a controlling vector of $F^{(\cdot)}_{64/384}$. These properties can be easily proved by using the definition of $F_{2/2}$.

- $\Pr\left[F_{2/2}(x_1, x_2, v, z) \oplus F_{2/2}(x_1 \oplus 1, x_2, v, z) = (1,0)\right] = 2^{-1}$.
- $\Pr\left[F^{(0)}_{64/384}(X, (V, Z)) \oplus F_{64/384}(X \oplus e_{64}, (V, Z)) = e_{64}\right] = 2^{-6}$.
- $\Pr\left[F^{(1)}_{64/384}(X, (V, Z)) \oplus F_{64/384}(X \oplus e_{64}, (V, Z)) = e_{64}\right] = 2^{-6}$.

Table 3. Two related-key differential characteristics of a 7-round reduced BMD-128

Round i	ΔI_i	$(\Delta U_i, \Delta Q_i)$	Probability
1	$(0, e_{64}) = \alpha$	$(0, e_{64})$	1
2	$(0, 0)$	$(0, 0)$	1
Output	$(0, 0) = \beta$	·	·
3	$(0, e_{64}) = \gamma$	$(0, e_{64})$	1
4	$(0, 0)$	$(0, 0)$	1
5	$(0, 0)$	$(0, 0)$	1
6	$(0, 0)$	$(0, 0)$	1
7	$(0, 0)$	$(e_{64}, 0)$	2^{-12}
FT	$(e_{64}, 0)$	$(0, e_{64})$	1
Output	$(e_{64}, e_{64}) = \delta$	·	·

We consider the situation that we encrypt plaintexts $P = (P_L, P_R)$, $P^* = (P^*_L, P^*_R)$, $P' = (P'_L, P'_R)$ and $P'^* = (P'^*_L, P'^*_R)$ and under keys K, K^*, K' and K'^* such that $\alpha = P \oplus P^* = P' \oplus P'^* = (0, e_{64})$, $\Delta K = K \oplus K^* = K' \oplus K'^* = (e_{64}, 0,0,0)$,

respectively. Then, as shown in Table 3, we can construct the first 2-round related-key differential characteristic $\alpha \rightarrow \beta$ for round 1 ~ 2 with probability 1, where $\beta = (0,0)$. Fig. 2-(a) present the propagation of the difference in round 1.

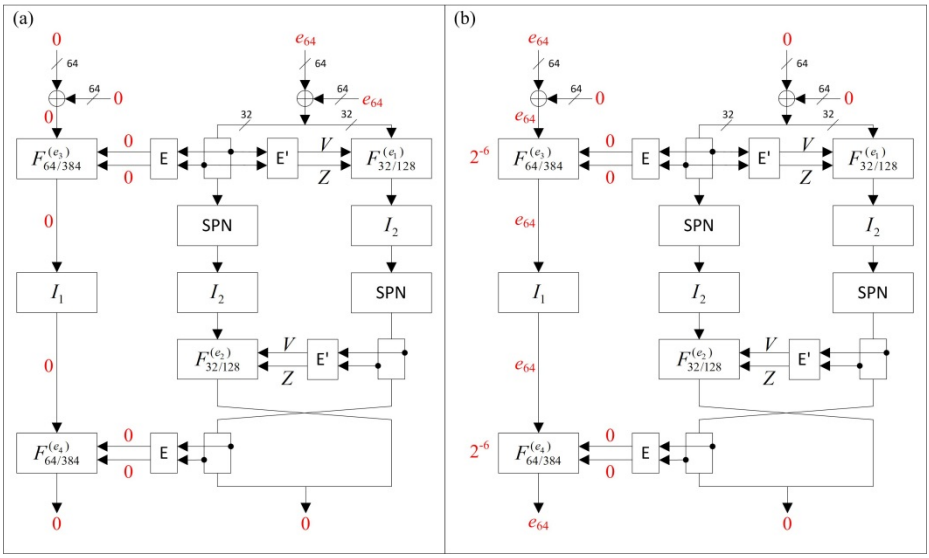


Fig. 2. Propagation of the difference in (a) round 1 and 3, (b) round 7

Similarly, we can construct a 5-round related-key differential characteristic for round 3 ~ 7. We encrypt intermediate values $I = (I_L, I_R)$, $I^* = (I_L^*, I_R^*)$, $I' = (I_L', I_R')$ and $I'^* = (I_L'^*, I_R'^*)$ and under keys K, K^*, K' and K'^* such that $\gamma = I \oplus I' = I^* \oplus I'^* = (0, e_{64})$, $\Delta K' = K \oplus K' = K^* \oplus K'^* = (e_{64}, 0,0,0)$, respectively. Then we can construct a 5-round related-key differential characteristic $\gamma \rightarrow \delta$ for round 3 ~ 7 with probability 2^{-12}, where $\delta = (e_{64}, e_{64})$. Fig. 2 present the propagation of the difference in round 1 and 7, respectively. The differential propagation of round 7 can be deduced by using the properties of $F_{64/384}^{(\cdot)}$.

We are now ready to present our distinguishing attack algorithm on a 7-round reduced BMD-128. We assume that a 7-round reduced BMD-128 cipher uses the secret key K and the related key K^* with difference $\Delta = K \oplus K^* = (e_{64}, 0,0,0)$. Our attack procedure is as follows.

1. Choose a pool of 2^{78} plaintext pairs (P_j, P_j^*) with the difference $\alpha = (0, e_{64})$ (j=1,… , 2^{78}) and construct 2^{155} plaintext quartets $(P_i, P_i^*, P_i', P_i'^*)$ $(i = 1, \cdots, 2^{155})$. With a chosen plaintext attack, $(P_i, P_i^*, P_i', P_i'^*)$ are encrypted using the keys K, K^*, K^*, K, respectively, to get the corresponding ciphertext quartets $(C_i, C_i^*, C_i', C_i'^*)$. We keep all these ciphertexts in a table.
2. For each ciphertext quartet $(C_i, C_i^*, C_i', C_i'^*)$, check that $C_i \oplus C_i' = C_i^* \oplus C_i'^* = (e_{64}, e_{64})$.

3. If the number of quartets passing Step 2 is greater than or equal to 6, output that the given ciphertexts were generated by using a 7-round reduced BMD-128. Otherwise output that the given ciphertexts were generated by using a 128-bit random permutation.

This attack requires a pool of 2^{78} plaintext pairs and thus the data complexity of this attack is 2^{79} related-key chosen plaintexts. If the given ciphertexts were generated by using a 7-round reduced BMD-128, our attack algorithm can distinguish between a 7-round reduced BMD-128 and a 128-bit random permutation with probability 1. Otherwise, the probability that our attack algorithm outputs that the given ciphertexts were generated by using a 7-round reduced BMD-128 is very low.

4 Conclusion

In this paper, we explained that it is possible to distinguish between a 7-round reduced BMD-128 and a 128-bit random permutation under the related-key amplified boomerang attack scenario. It means that BMD-128 is still vulnerable to related-key attacks. Moreover, our distinguishing attack can be extended to the related-key attack on this algorithm. So, our further research is to propose the related-key attack on BMD-128.

Acknowledgments. This research was supported by the MSIP(Ministry of Science, ICT&Future Planning), Korea, under the C-ITRC(Convergence Information Technology Research Center) support program (NIPA-2013-H0301-13-3007) supervised by the NIPA(National IT Industry Promotion Agency).

References

1. Goots, N., Izotov, B., Moldovyan, A., Moldovyan, N.: Modern cryptography: Protect Your Data with Fast Block Ciphers. A-LIST Publish, Wayne (2003)
2. Goots, N., Moldovyan, A., Moldovyan, N.: Fast Encryption Algorithm Spectr-H64. In: Gorodetski, V.I., Skormin, V.A., Popyack, L.J. (eds.) MMM-ACNS 2001. LNCS, vol. 2052, pp. 275–286. Springer, Heidelberg (2001)
3. Goots, N., Moldovyan, N., Moldovyanu, P., Summerville, D.: Fast DDP-Based Ciphers: From Hardware to Software. In: 46th IEEE Midwest International Symposium on Circuits and Systems (2003)
4. Jeong, K., Lee, C., Kim, J., Hong, S.: Security analysis of the SCO-family using key schedules. Information Sciences 179, 4232–4242 (2009)
5. Jeong, K., Lee, C., Sung, J., Hong, S., Lim, J.: Related-Key Amplified Boomerang Attacks on the Full-Round Eagle-64 and Eagle-128. In: Pieprzyk, J., Ghodosi, H., Dawson, E. (eds.) ACISP 2007. LNCS, vol. 4586, pp. 143–157. Springer, Heidelberg (2007)
6. Kang, J., Jeong, K., Yeo, S., Lee, C.: Related-Key Attack on the MD-64 Block Cipher Suitable for Pervasive Computing Environments. In: Proceedings of WAINA 2012, Fukuoka, Japan, pp. 726–731 (2012)

7. Ko, Y., Hong, D., Hong, S., Lee, S., Lim, J.: Linear Cryptanalysis on SPECTR-H64 with Higher Order Differential Property. In: Gorodetsky, V., Popyack, L.J., Skormin, V.A. (eds.) MMM-ACNS 2003. LNCS, vol. 2776, pp. 298–307. Springer, Heidelberg (2003)
8. Ko, Y., Lee, C., Hong, S., Lee, S.: Related Key Differential Cryptanalysis of Full-Round SPECTR-H64 and CIKS-1. In: Wang, H., Pieprzyk, J., Varadharajan, V. (eds.) ACISP 2004. LNCS, vol. 3108, pp. 137–148. Springer, Heidelberg (2004)
9. Ko, Y., Lee, C., Hong, S., Sung, J., Lee, S.: Related-Key Attacks on DDP based Ciphers: CIKS-128 and CIKS-128H. In: Canteaut, A., Viswanathan, K. (eds.) INDOCRYPT 2004. LNCS, vol. 3348, pp. 191–205. Springer, Heidelberg (2004)
10. Lee, C., Hong, D., Lee, S., Lee, S., Yang, H., Lim, J.: A Chosen Plaintext Linear Attack on Block Cipher CIKS-1. In: Deng, R.H., Qing, S., Bao, F., Zhou, J. (eds.) ICICS 2002. LNCS, vol. 2513, pp. 456–468. Springer, Heidelberg (2002)
11. Lee, C.-H., Kim, J.-S., Hong, S.H., Sung, J., Lee, S.-J.: Related-Key Differential Attacks on Cobra-S128, Cobra-F64a, and Cobra-F64b. In: Dawson, E., Vaudenay, S. (eds.) Mycrypt 2005. LNCS, vol. 3715, pp. 244–262. Springer, Heidelberg (2005)
12. Lee, C.-H., Kim, J.-S., Sung, J., Hong, S.H., Lee, S.-J., Moon, D.: Related-Key Differential Attacks on Cobra-H64 and Cobra-H128. In: Smart, N.P. (ed.) Cryptography and Coding 2005. LNCS, vol. 3796, pp. 201–219. Springer, Heidelberg (2005)
13. Lu, J., Lee, C., Kim, J.: Related-Key Attacks on the Full-Round Cobra-F64a and Cobra-F64b. In: De Prisco, R., Yung, M. (eds.) SCN 2006. LNCS, vol. 4116, pp. 95–110. Springer, Heidelberg (2006)
14. Minh, N., Bac, D., Duy, H.: New SDDO-Based Block Cipher for Wireless Sensor Network Security. International Journal of Computer Science and Network Security 10(3), 54–60 (2010)
15. Moldovyan, N.: On Cipher Design Based on Switchable Controlled Operations. In: Gorodetsky, V., Popyack, L.J., Skormin, V.A. (eds.) MMM-ACNS 2003. LNCS, vol. 2776, pp. 316–327. Springer, Heidelberg (2003)
16. Moldovyan, A., Moldovyan, N.: A cipher Based on Data-Dependent Permutations. Journal of Cryptology 15(1), 61–72 (2002)
17. Moldovyan, N., Moldovyan, A., Eremeev, M., Sklavos, N.: New Class of Cryptographic Primitives and Cipher Design for Networks Security. International Journal of Network Security 2(2), 114–225 (2006)
18. Moldovyan, N., Moldovyan, A., Eremeev, M., Summerville, D.: Wireless Networks Security and Cipher Design Based on Data-Dependent Operations: Classification of the FPGA Suitable Controlled Elements. In: Proceedings of CCCT 2004, Texas, USA, vol. VII, pp. 123–128 (2004)
19. Sklavos, N., Moldovyan, N., Koufopavlou, O.: A New DDP-based Cipher CIKS-128H: Architecture, Design & VLSI Implementation Optimization of CBCEncryption & Hashing over 1 GBPS. In: Proceedings of The 46th IEEE Midwest Symposium on Circuits & Systems, Cairo, Egypt, December 27-30 (2003)
20. Sklavos, N., Moldovyan, N., Koufopavlou, O.: High Speed Networking Security: Design and Implementation of Two New DDP-Based Ciphers. In: Mobile Networks and Applications, MONET, vol. 25(1-2), pp. 219–231. Kluwer Academic Publishers (2005)
21. Bac, D., Minh, N., Duy, H.: An Effective and Secure Cipher Based on SDDO. International Journal of Computer Network and Information Security 11, 1–10 (2012)

A Receiver-Initiated MAC Protocol for Energy Harvesting Sensor Networks

Kien Nguyen[1], Vu-Hoang Nguyen[2], Duy-Dinh Le[1,2], Yusheng Ji[1], Duc Anh Duong[2], and Shigeki Yamada[1]

[1] National Institute of Informatics,
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
{kienng,ledduy,kei,shigeki}@nii.ac.jp
[2] Multimedia Communications Laboratory, University of Information Technology,
KP6, Linh Trung Ward, Thu Duc District, Ho Chi Minh City, Vietnam
{vunh,ducda}@uit.edu.vni

Abstract. Energy harvesting technology potentially solves the problem of energy efficiency, which is the biggest challenge in wireless sensor networks. The capability of harvesting energy from surrounding environment enables an achievement of infinitive lifetime at a sensor node. The technology promisingly changes the fundamental principle of communication protocols in wireless sensor networks. Instead of saving energy as much as possible, the protocols should keep the efficient operation and maximum performance of networks while guaranteeing the harvested energy is equal or bigger than the consumed energy. In this paper, we propose ERI-MAC a new receiver-initiated MAC protocol for energy harvesting sensor networks. ERI-MAC leverages the benefit of receiver-initiated and packet concatenation to achieve good performance both in latency and energy efficiency. Moreover, ERI-MAC employs a queuing mechanism to adjust the operation of a sensor node following the energy harvesting rate from the surrounding environment. The extensive simulation results in ns-2 show that ERI-MAC achieves good network performance, as well as, enables infinitive lifetime of sensor networks.

Keywords: MAC protocol, energy harvesting, WSNs.

1 Introduction

The developments of sensing, computing technologies and wireless communication drive the appearance of wireless sensor networks (WSNs) with various types of applications such as structure health [1], environmental monitoring [2,3] or healthcare [4]. A WSN usually contains numerous inexpensive sensor nodes, which are spatially distributed over a monitored commonplace. The sensor nodes sense the physical changes of its surrounding environment and wirelessly forward the sensing data to a base station, i.e., a sink. An individual sensor node normally has a small size; and it is powered with a limited capacity battery. Therefore, the operation and performance of WSNs largely depends on the finite capacity of power sources. Traditionally, most of

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_77,

research in WSNs pays attention on designing energy efficient communication protocols, especially Medium Access Control (MAC) protocols. That is because the MAC protocols control the operation of radio module, which is the biggest consumer of energy on a sensor nodes. In general, the MAC protocols save consumed energy by adopting the duty cycling mechanism, which periodically turns on and off the radio modules. There are a huge number of power-saving MAC protocols have been published, from low to ultra low duty cycle [5 - 7], or achieves good performance with different types of traffic [8]. However, if the WSN applications requires a long lifetime (months, or years), the capacity of battery is still not sufficient. On the other hand, re-cent advances in energy harvesting technology give a promising solution for the energy problem on WSNs.

Energy harvesting refers to the capability of extracting energy from ambient environment of a sensor node (e.g., from the solar energy, wind power, etc. [9, 10]). Moreover, the extracted or harvested energy can be used to recharge the node's battery. By doing so, the sensor node potentially maintains an infinite life-time of battery. The technology therefore will change the fundamental principle of designing MAC protocols for WSNs. Instead of focusing on the power-saving aspect, the objectives of new MAC protocol on energy-harvesting WSNs include increasing both the network performance and lifetime under a given condition of harvested energy. Different to the traditional MAC protocols, the one in energy harvested WSN achieves infinite lifetime by keeping the sensor node operate at a so-called energy neutral operation (ENO) state [9]. When a node is in the ENO state, its energy consumption is always less than or equal to the energy harvested from the environment. Besides that, WSNs with energy harvesting capability assume the correlation between the performance and energy harvesting. The more energy a sensor node is harvested the better performance it achieves. A sensor node is said to reach the state of ENO-Max when it operates at the maximum performance as well as remains the state of ENO [11]. Generally, the MAC protocols in energy harvested sensor networks are designed with new algorithms of dynamically adapting the duty cycle at a node in order to maximize both the lifetime and performance.

The remainder of the paper is organized as follows. Section 2 describes the design of ERI-MAC protocol. In Section 3 presents the evaluation results. Finally, we conclude the paper in Section 4.

2 ERI-MAC for Energy Harvesting Sensor Networks

In this section, we describe the operation of ERI-MAC protocol. We initially present the basic communication scheme in ERI-MAC, and then we discuss the use of a dynamic queuing mechanism in order to achieve the ENO state.

2.1 Basic Communication Scheme

Receiver-Initiated Mechanism. Receiver-initiated mechanism is always adopted by asynchronous duty cycling protocols, which do not require any clock synchronization between sensor nodes. The MAC protocols equipped the mechanism has been proven to outperform the state-of-the-art of traditional sender-initiated protocol and the

synchronous protocols [12]. Moreover, the mechanism is therefore taking part in main stream of designing MAC protocols on real sensor motes, and real deployments [13]. Figure 1 shows a basic operation of receiver-initiated communication. In the Figure 1, SIFS is abbreviated for short inter-frame space, the duration needed to process a packet and switching radio mode. The mechanism is always combined with duty cycling radio (i.e., sleep/wakeup) following operational cycles. In an operational cycle, after waking up each non-sender node immediately broadcasts a beacon packet. The beacon contains the node's address; it is used to announce that the node is ready for receiving a data packet. The node then samples the wireless channel for a short period (called dwell time) to determine there is any potential incoming packet. On the other hand, a sender that is holding a data packet keeps in the listening mode and waits for a beacon from its intended receiver. When the sender receives the expected beacon, it immediately sends the pending packet. A successful transmission is completed when a beacon with acknowledge (ACK) function arrives at the sender. This beacon however can serve not only as an ACK packet but also as a new receiver-initiated beacon. After the completed transmission if the sender has no queued packet, it becomes a non-sender. The node then broadcasts a beacon right after its next wakeup time. ERI-MAC also adopts the same collision detection and retransmission schemes from RI-MAC and AQ-MAC [8]. When a collision occurs at a receiver, it retransmits a new beacon, which includes a value of back-off window. Each contending sender utilizes a random backoff period before a retransmission to avoid collisions.

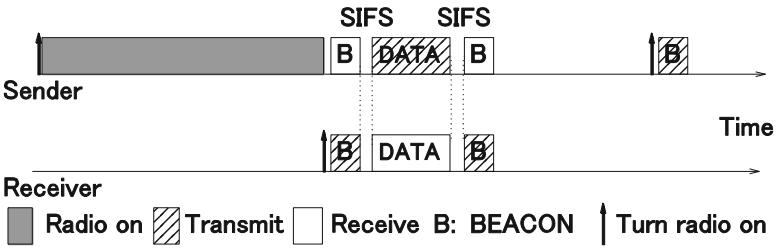


Fig. 1. ERI-MAC 's basic communication scheme

Packet Concatenation. Packet concatenation refers to the implementation of concatenating several small packets, which share a same characteristic in to a bigger one. In WSNs, this scheme is common and necessary since the nodes periodically sleep to save energy while the sensing activities are continuous. As a result, the sensing information has to be stored in queued packets, that are normally destined to one sink. Therefore, the scheme improves the network performance both in latency and energy efficiency by reducing control overhead and queuing time. Implementing packet concatenation at MAC layer is originally proposed in our previous work [14]. The original scheme is equipped to a synchronous multi-hop duty cycling MAC protocol. However, the scheme can be adopted by other duty cycling protocols whenever the protocols have to handle with queuing packets such as in [8]. In our

packet concatenation, we define the big packet as super packet. The size of super packet is limited by a threshold value depending on the radio's capability.

2.2 Queuing Mechanism to Achieve ENO

The queuing mechanism is first proposed to handle with Quality of Service (QoS) provision for low priority traffic in AQ-MAC. The packets are queued until a timeout value before sending out at a node. The original mechanism uses a fix and predetermined values of timeout and very efficient in terms of energy and latency efficiency. We found that that the mechanism can fit well in the context of energy harvesting sensor networks. Therefore, we extend the mechanism in order to achieve the ENO state at a ERI-MAC's node. The timeout value is now dynamic and controlled by a node. The node compares it energy consumption with harvest energy from it environment. If the amount of energy consumption is bigger than the amount of harvested one, the sensor nodes reduce its transmissions and waiting for the harvested energy. Therefore, a node can reach to the desired state.

In ERI-MAC, we assume that the node knows the energy harvesting rate, its capacity of battery and a safe duration, which is the maximum period of awake state of radio. If the radio keeps on over that duration, the battery can be exhausted. After each safe duration, the nodes compare its consumed energy to the harvested energy by investigating the proportion between them. If the value is less than one, the node immediately goes to sleep until it can guarantee the battery is sufficiently safe. In our evaluation, we use the operational cycle with the length of one ms, and the safe duration is determined following the appropriate energy harvesting rate and the consume energy rate in the real sensor nodes' specifications.

3 Evaluation

Table 1. Networking Parameters

Bandwidth	250 Kbps	Slot time	320 us
CCA Check Delay	128 us	Tx Range	250 m
Carrier Sensing Range	550 m	SIFS	191 us
Backoff Window	0-255	Beacon size	6-9 bytes
Retry Limit	5	Dwell Time	10 ms
Tx Power	31.2 mW	L_{TH}	112 bytes
Rx Power	22.2 mW	Sleep Power	3 uW

3.1 Experiment Settings

We evaluate the performance of ERI-MAC using the network simulator ns-2 [15]. We demonstrate correlations of energy consumption to the performance of wireless sensor networks by modifying the energy module of ns-2. Table 1 lists the network parameters of a sensor node. Those parameters are collected from in Micaz mote and

Radio CC2420's specifications except the Transmission range (Tx range) and Carrier Sensing range. L_{TH} is the maximum size of the super packet, which concatenates four original 28-byte packets. Different with other related work in energy-harvesting sensor networks, we investigate the performance of ERI-MAC in a 49-node grid scenario. The distance between two neighbors in the grid is 200 meters, and all the data packets are destined to the sink in the center as shown in Fig. 2. The Random Correlated Event (RCE) model is used in the evaluation. In the model, an event is occurred at a random location within the area. An event is characterized by a so-called sensing range of the event. All nodes, which are within the sensing range of event, are going to generate one packet to the sink. In the evaluation, the inter-event values are randomly within zero to five seconds. We generate a total of 100 events with 500 meter-sensing range. The length of a cycle is one seconds and a safe duration is ve seconds. We adopt the energy harvesting model from previous work [16]. The energy harvesting rates are constant at 0.3 mWatt an 0.6 mWatt.

3.2 Results

This section presents the results of ERI-MAC performance in the evaluation. The values of latency and energy efficiency are shown in Fig. 3 and Fig. 4, respectively. Note that, each value of latency shown in Fig. 3 is calculated between the generated and received time of a packet. In the case of super packet, the generated time of the packet is considered as the one of the first packet in the concatenated form. If the rate of energy harvesting is small, ERI-MAC 's nodes tend to excess the safe duration more frequently. Therefore, the packets in that scenario have to be queued, that leads to the higher values of latency. On the other hand, the value of 0.6 mW harvesting rate is sufficient to guarantee the safe duration of ERI-MAC, hence the network has a good latency performance.

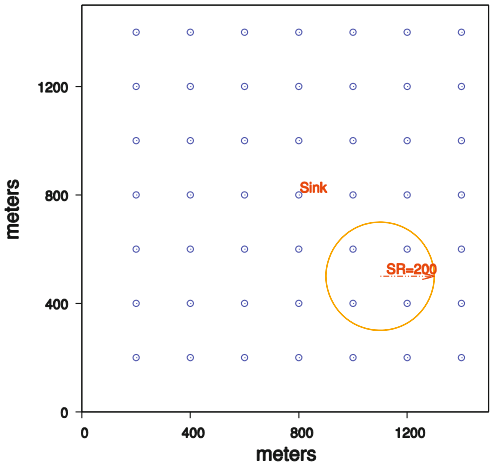


Fig. 2. 49-node Grid Scenario and an event with 200-meter sensing range

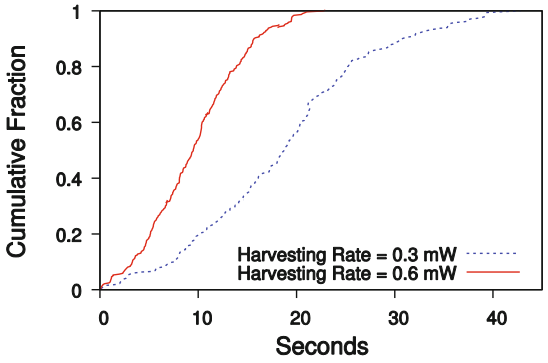


Fig. 3. Cumulative distribution of latency values

In order to investigate the energy efficiency, we use the ratio of consumed energy to harvested energy. If the ratio is smaller than one, the node is confirmed in an ENO state. Since when the harvesting rate is 0.6 mW, ERI-MAC's nodes do not excess the safe duration hence no appearance of adapting duty cycle. Therefore, we focus on the case of 0.3 mW harvesting rate. The results are in the Fig. 4. Since ERI-MAC has the queuing mechanism, each time the nodes consume more energy than the amount of harvested energy, they themselves keep their radio off in order to be in ENO sates. We can observe that all nodes in the grid have the ratio smaller than one. Hence, we can conclude that the network can achieve the infinitive lifetime.

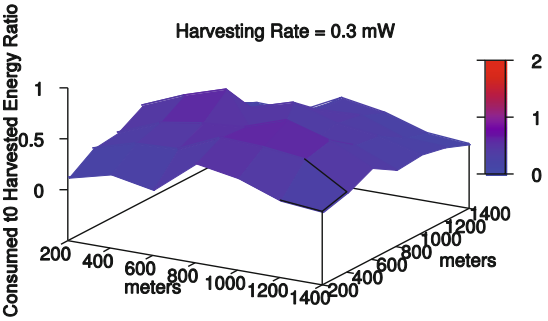


Fig. 4. The ratio of consumed to harvested energy

4 Conclusion

The energy harvesting technology, which lets a battery on a node be recharged by energy in its surrounding environments, is potentially a solution of overcoming the energy problem in WSNs. In this paper, we propose ERI-MAC a new receiver-initiated MAC protocol for WSNs with energy harvesting capabilities. ERI-MAC inherits the advantages of receiver-initiated communication, packet concatenation in order to achieve good network performances. Moreover, ERI-MAC 's nodes use the

queuing packet mechanism to adapt the operation of a sensor node to the rate of harvested energy. If the ratio between the consumed to harvested energy is larger than one, ERI-MAC's nodes switch to and stay in sleep modes until the batteries are safe. The simulation results show that the ERI-MAC's network achieves good network performances and keeps all nodes in ENO state, i.e., achieving infinitive lifetime.

In the future, we plan to theoretically analyze the operation and performance of ERI-MAC. The analysis will be added to the queuing mechanism in order to let sensor nodes reach the state of ENO-MAC. Moreover, we also plan to extend ERI-MAC for the other sensor networks with different energy harvesting models.

Acknowledgments. This research is funded by Vietnam National University HoChiMinh City (VNU-HCM) under grant number DHQG-C20132601.

References

1. Liu, X., Cao, J., Lai, S., Yang, C., Wu, H., Xu, Y.: Energy efficient clustering for wsn-based structural health monitoring. In: Proc. IEEE INFOCOM, pp. 2768–2776 (2011)
2. Mainwaring, A., Culler, D., Polastre, J., Szewczyk, R., Anderson, J.: Wireless sensor networks for habitat monitoring. In: Proc. 1st ACM International Workshop on Wireless Sensor Networks And Applications, WSNA 2002, pp. 88–97 (2002)
3. Tolle, G., Polastre, J., Szewczyk, R., Culler, D., Turner, N., Tu, K., Burgess, S., Dawson, T., Buonadonna, P., Gay, D., Hong, W.: A macroscope in the redwoods. In: Proc. 3rd ACM SenSys, pp. 51–63 (2005)
4. Shnayder, V., Chen, B.-R., Lorincz, K., Jones, T.R.F.F., Welsh, M.: Sensor networks for medical care. In: Proc. 3rd International Conference on Embedded Networked Sensor Systems, SenSys 2005, p. 314 (2005)
5. Ye, W., Heidemann, J., Estrin, D.: Medium Access Control with Coordinated Adaptive Sleeping for Wireless Sensor Networks. IEEE/ACM Transactions on Networking 12, 493–506 (2004)
6. Ye, W., Silva, F., Heidemann, J.: Ultra-low duty cycle MAC with scheduled channel polling. In: Proc. 4th ACM SenSys, pp. 321–334 (2006)
7. Sun, Y., Du, S., Gurewitz, O., Johnson, D.B.: DW-MAC: a Low Latency, Energy Efficient Demand-Wakeup MAC Protocol for Wireless Sensor Networks. In: Proc. 9th ACM MobiHoc, pp. 53–62 (2008)
8. Nguyen, K., Ji, Y.: Asynchronous mac protocol with qos awareness in wireless sensor networks. In: Proc. IEEE Global Communications Conference, GLOBECOM 2012, pp. 555–559 (2012)
9. Kansal, A., Hsu, J., Zahedi, S., Srivastava, M.B.: Power management in energy harvesting sensor networks. ACM Trans. Embed. Comput. Syst. 6(4) (September 2007)
10. Voigt, T., Ritter, H., Schiller, J.: Utilizing solar power in wire-less sensor networks. In: Proc. 28th IEEE Conference on Local Computer Networks LCN (2003)
11. Vigorito, C.M., Ganesan, D., Barto, A.G.: Adaptive control of duty cycling in energy-harvesting wireless sensor networks. In: Proc. 4th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, SECON 2007, pp. 21–30 (2007)

12. Sun, Y., Gurewitz, O., Johnson, D.B.: RI-MAC: a Receiver-Initiated Asynchronous Duty Cycle MAC Protocol for Dynamic Traffic Loads in Wireless Sensor Networks. In: Proc. 6th ACM SenSys, pp. 1–14 (2008)
13. Dutta, P., Dawson-Haggerty, S., Chen, Y., Liang, C.-J.M., Terzis, A.: Design and Evaluation of a Versatile and Efficient Receiver-Initiated Link Layer for Low-Power Wireless. In: Proc. 8th ACM SenSys, pp. 1–14 (2010)
14. Nguyen, K., Meis, U., Ji, Y.: MAC^2: a Multi-hop Adaptive Mac Protocol with Packet Concatenation. IEICE Transaction on Information and Systems E95-D(2), 480–489 (2012)
15. The network simulator ns2, `http://www.isi.edu/ns/index.html`
16. Fafoutis, X., Dragoni, N.: Odmac: an on-demand mac protocol for energy harvesting - wireless sensor networks. In: Proc. of the 8th ACM Symposium on Performance Evaluation of Wireless Ad hoc, Sensor, and Ubiquitous Networks, PE-WASUN 2011, pp. 49–56 (2011)

A Data Driven cSLAM of Multiple Exploration Robots

Sang-Hoon Ji, Phuc Truong Huu, Hong-Seok Kim, and Sang-Moo Lee

Korea Institute of Industrial Technology, Korea
{robot91,whitelion_pc,hskim,lsm}@kitech.re.kr

Abstract. This paper proposes a novel cSLAM method for multiple exploration robots with insufficient sensor ranges. For this purpose, we make the robots depend on a navigation map and sensor model driven from data obtained during the exploration instead of predefined sensor or navigation map model. And The proposed algorithm is implemented with OPRoS (Open SW Platform for Robotic Services) SW Components.

Keywords: cSLAM, Multiple Robots, Exploration, Data Driven Modeling.

1 Introduction

In recent, application of robots expand from tasks in known structured environments to tasks in various unknown environments such as surveillance and searching for landmines of unmanned military vehicles and resource exploration of unmanned submarines. For completion of the missions in dynamic environments, it is needed that the robots should be able to move independently without any prior information about environments.

SLAM (Simultaneous Localization and Mapping) technique is the heart of this ability, which enables the robot to gradually perceive and map its environment, while localizing its positing by utilizing the environment map. So, a lot of researches have been undertaken in regards to SLAM. At the earlier stage, they focused on SLAM methods which overcome sensor errors and control errors with filter algorithms such as EKF-SLAM, UKF-SLAM, Information Filter SLAM, and Particle Filter SLAM. While in recent, multi-robot SLAM methods are suggested. These methods perform SLAM using multiple robots which can communicate and cooperate with one another. The Multi-robot SLAM is often called Cooperative-SLAM (cSLAM).

Though cSLAM is more efficient and accurate than single-robot SLAM, cSLAM inherits the difficulties of single-robot SLAM. So, many types of cSLAM have been proposed. But these approaches have disadvantages of practical aspects in the unknown environments. So in this paper, we suggest a new data driven approach to cSLAM which is tolerant to sensor noises and can be implemented very easily.

This paper will be organized as follows. At first, the cSLAM is explained in the section 2. The section 3 describes the cSLAM scenario design based on data driven method. And then, finally, the Section 5 concludes the paper.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_78,

2 An Overview of cSLAM

SLAM is to simultaneously perform localization and mapping. While SLAM, robot localization is to estimate its own location (or path) using given map and robot mapping is to build the surrounding environments using given robot location. Moreover, when robot path-planning adds to SLAM, SLAM is converted to integration exploration.

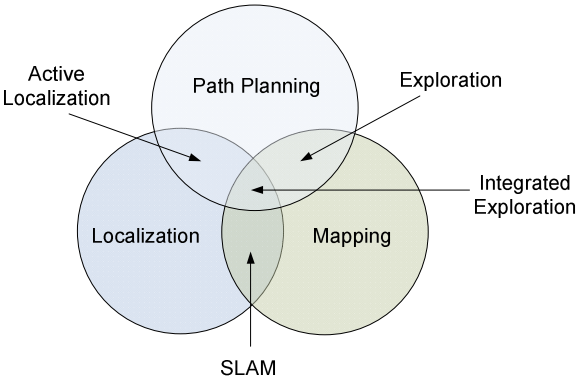


Fig. 1. Relation of SLAM and Exploration [2]

The SLAM problem is expressed as shown in Eq.(1).

$$\text{argmax}_{\{x,M\}} \{P(x_{1:t}, M \mid z_{1:t}, u_{1:t})\} \tag{1}$$

where, $x_{1:t}$ is the trajectory of robot at times 1, 2, .., t and M is the estimated map information. And $z_{1:t}$ is the sequence of observations from the robot at times 1, 2, …t and $u_{1:t}$ is the sequence of actions executed by the robot or trajectory command for the robot.

The SLAM problem can be converted to the cSLAM as shown Eq. (2).

$$\underset{x^1, x^2, M}{\text{argmax}} \{P(x^1_{1:t}, x^2_{s:t}, M \mid z^1_{1:t}, u^1_{1:t}, x^1_0, z^2_{s:t}, u^2_{s:t}, \Delta^{21}_s)\} \tag{2}$$

where, $x^i_{k:t}$ is the trajectory of robot i at times k, k+1, .., t, and M is the estimated map information. And $z^i_{k:t}$ is the sequence of observations from the robot i at times k, k+1, …t and $u^i_{k:t}$ is the sequence of actions executed by the robot or trajectory command for the robot i. And Δ^{21}_s is the relative pose between two robots at time s which is can be obtained by map merging.

Convenient map merging algorithms are summarized as shown on Table 1. Direct map merging (DMM) has the advantage for the computation time over indirect map merging (IMM) but is not robust against sensor noises. So, recent researches attempt at developing novel methods which take over the good features from DMM and IMM. Especially, if the robots observe specific landmarks or use relative position among robots, they can build up their common navigation map more accurately and faster.

Table 1. Overview of cSLAM algorithms

Category	Type	Description
Direct Map Merging	Robot-to-robot measurements	The visual and range measurements when the robots meet each other at a rendezvous
	Common regions/objects	The locations and orientations of common regions or objects in the multiple maps
Indirect Map Merging	Point feature matching	Finding MTM which maximally matches the point features by NNT or PFM
	Application of scan matching	Applying scan matchin(jg algorithms such as ICP, IDC and PSM to map matching
	Spectra-based map matching	Finding and matching spectral information on maps instead of geometric information

3 A Data Driven Approach to cSLAM (1.5)

We assume that there are two types of robots as followed.

- Sensor node robot: Local coordinated movement, limited ranging sensor, and limited range communication
- Mapping mobile robot: Global coordinated movement and limited range communication

[STEP 1] Deployment of Sensor Node Robots
The Sensor node robots are scattered according to a bio-inspired method such as potential force field until the all the area of environment is covered. The robots can distinguish any robots in their own sensing areas. And the sensing area is determined with the sensor ability, the environment where the sensor is deployment, and the degree of required robot localization. At the step, the robots have no idea of their own positions and poses, but they only have their local map.

$$F_i = \sum_j \frac{e_i(j)}{\|d_{ij}\|} \tag{3}$$

where F_i is the force acted on the sensor node robot i at the step and d_{ij} is the distance between robot i and robot j. And $e_i(j)$ is the unit vector which directs from robot i to robot j and the size of the vector is 1 when d_{ij} is smaller than the certain value. Other cases, it is the null vector.

The certain value of Sensor node robot is designed such that the robot can have confidence of locating any object which the distance between the robot and the sensor robot is smaller than it.

[STEP 2] Map the Environment Using Mapping Mobile Robots

The Mapping mobile robots navigate the environment. And when a Mapping mobile robot meets other Mapping mobile robots or Sensor node robots, it exchanges the location information with its friend robots with the rule of confidence of localization information.

The confidence function of localization information is defined as in Eq. (4).

$$Confidence(i) = \min\left(Confidence(i), Confidence\ (j) + \varepsilon \right) \tag{4}$$

Where i and j is the indexes of the robot I and robot j, and con, confidence (i) is the confidence function value of localization information for the robot i. And ε is the localization sensor error of the robots. We assume that all the robot have the same sized of localization sensor errors.

When there are three Mapping mobile robots, we assign a leader role to one robot and follower roles to the other robots as shown if Fig. 2.

A Leader robot has cooperation module with the robot plan the robots' paths globally and merge the map obtained by the robots. All the robots can move with their own local path planner and make their own map. And the robots register objects as natural land markers with model-based object detection modules.

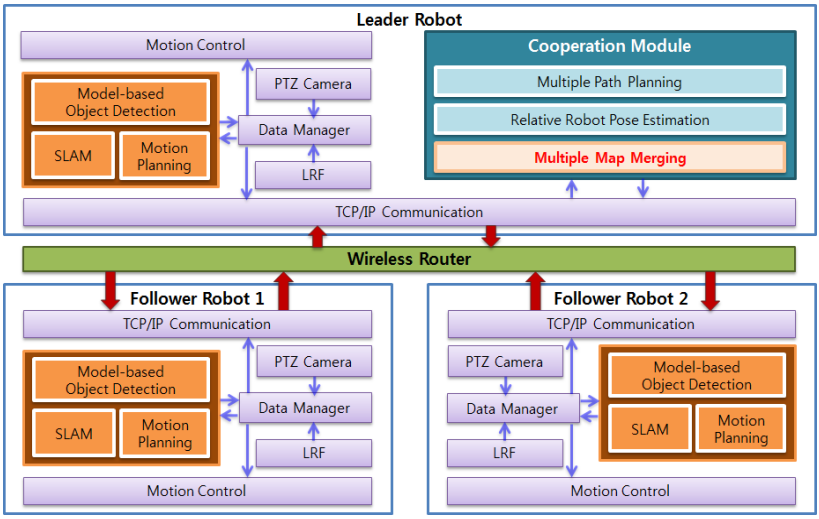


Fig. 2. Structure of cSLAM for three robots

[STEP 3] Finish the Map Building

If all the Sensor node robots have lower localization error than predetermined values, the robots finish mapping. If there are many sensor node robots enough that the union of their localization sensor ranges cover the entire exploration region, the sufficient confidence of any place can be obtained.

An OPRoS SLAM component consists of several function modules as shown in Fig. 3. The SLAM algorithm can use EKF, UKF and Particle Filter and obtain line feature information of the objects from LRF, CCD Camera, or ultra-sonar. In case there is a big error of map, the SW module can correct the map with Inverse EKF algorithm.

It is needed to abstract sensor models in order to provide sensor components with reconfiguration. For this reconfiguration, we divide algorithms with hardware characteristics. For example, if we design a component which utilizes an algorithm designed on range sensor models as shown in Fig. 3, it is not needed to change the component when the used sensor is exchanged with any other sensor.

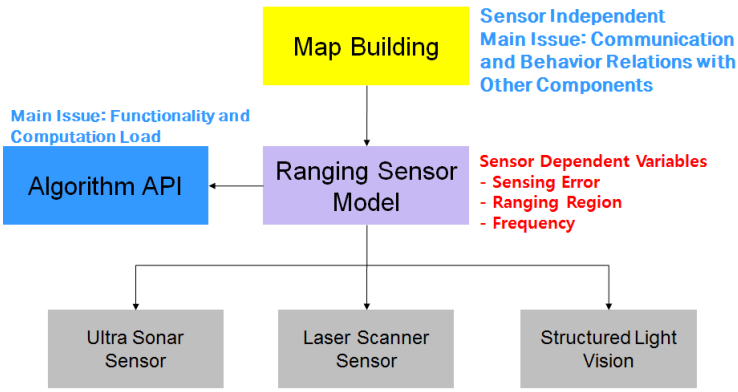


Fig. 3. Abstracion of OPRoS Mapping Component Modules

4 Conclusion

In this paper, we proposed a novel cSLAM method for multiple exploration robots with insufficient sensor ranges. The robots can build a navigation map and localize their own positions with guaranteed accuracy. And our proposed algorithm is implemented with OPRoS (Open SW Platform for Robotic Services) SW Components in order to help other researchers use the algorithm with less effort.

Acknowledgement. This work is supported by the Ministry of Trade, Industry and Energy (MOTIE), South Korea.

References

1. Thrun, S.: Probabilistic Robotics. MIT Press
2. Siciliano: Handbook of Robotics. Springer
3. Bailey, T.: Consistency of the EKF-SLAM Algorithm. In: IROS 2006 (2006)
4. Kurt-Yavuz: A Comparison of EKF, UKF, FastSLAM2.0, and UKF-based FastSLAM Algorithms. In: IEEE INES 2012 (2012)

5. Lee: Comparison and Analysis of Scan Matching Techniques for Cooperative-SLAM. In: URAI 2011 (2011)
6. Lee: Performance Evaluation of Scale Invariant Feature Detection Algorithms. In: KROS 2011 (2011)
7. Leung: Active SLAM in Structured Environments. In: ICRA 2008 (2008)
8. Henry: RGB-D Mapping. International Journal of Robotics Research (2012)
9. Lee: A Survey of Map Merging Techniques for Cooperative-SLAM. In: URAI 2012 (2012)

Efficient Purchase Pattern Clustering Based on SOM for Recommender System in u-Commerce

Young Sung Cho[1], Song Chul Moon[2], Seon-phil Jeong[3], In-Bae Oh[4], and Keun Ho Ryu[1]

[1] Department of Computer Science, Chungbuk National University, Cheongju, Korea
[2] Department of Computer Science, Namseoul University, Cheonan-City, Korea
[3] Computer Science and Technology, DST, BNU-HKBU United International College
[4] Chungbuk Health & Science University, Chungbuk, Korea
youngscho@empal.com, moon@nsu.ac.kr, spjeong@uic.edu.hk, iboh@chsu.ac.kr, khryu@dblab.chungbuk.ac.kr

Abstract. This paper proposes an efficient purchase pattern clustering method based on SOM(Self-Organizing Map) for Personal Ontology Recommender System in u-Commerce under ubiquitous computing environment which is required by real time accessibility and agility. In this paper, it is necessary for us to keep clustering the user's information to join the user's score based on RFM factors using SOM network and the analysis of RFM to be able to reflect the attributes of the user in order to reflect frequently changing trends of purchase pattern by emphasizing the important users and items, and to improve better performance of recommendation. The proposed makes the task of an efficient purchase pattern clustering based on SOM for preprocessing so as to be possible to recommend by the loyalty of RFM factors as considering user's propensity. To verify improved better performance of proposing system than the previous systems, we carry out the experiments in the same dataset collected in a cosmetic internet shopping mall.

Keywords: RFM, Collaborative Filtering, SOM(Self-Organizing Map).

1 Introduction

Customer segmentation is the process of grouping the customers based on their purchase habit. Data mining is useful in finding knowledge from huge amounts of data. Deboeck and Kohonen describe how SOM (Self-Organizing Map) can be used for effective clustering and segmentation of financial data[1]. Clustering algorithm is a kind of customer segmentation methods commonly used in data mining. In this paper, SOM network is applied to segment the purchase data to join user's information and finally forms clusters of the purchase data to join user's information with different features, demographic variable such as age, gender, occupation, skin type and RFM factors, we try to position the target customers of the u-commerce to promote properly, and then we can recommend the items with high purchasablility to target customer. The recommendation system helps customers to find easily items and helps the e-commerce companies to set easily their target customer by automated

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*, Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_79,

recommending process. Therefore, customers and companies can take some benefit from recommendation system. The possession of intelligent recommendation system is becoming the company's business strategy. A recommendation system using data mining technique based on RFM to meet the needs of customers has been actually processed the research[2-5]. We can make the solution for an efficient purchase pattern clustering based on SOM. Finally, we can improve the performance of personal ontology recommender system through SOM learning method based on the purchase data to show customer's buying patterns. The next section briefly reviews the literature related to studies. The section 3 is described a new method for personalized recommendation system in detail, such as system architecture with sub modules, the procedure of processing the recommendation, the algorithm for proposing system. The section 4 describes the evaluation of this system in order to prove the criteria of logicality and efficiency through the implementation and the experiment. In section 5, finally it is described the conclusion of paper and further research direction.

2 Related Works

2.1 RFM

RFM(Recency, Frequency, and Monetary) is composed of three measures and the definitions are described below [6]. RFM factors are the important attributes that determine the purchase behavior of the customer. For recency, the customer database is sorted by purchase dates by descending order. So, the top segment is given a value of 5 and the others are descendingly assigned of 4, 3, 2, and 1. For frequency and monetary, sorting customer visiting frequency data and the customer data related to the amount of the money spent in descending order, respectively. These three variables belong to behavioral variables and can be acted as the segmenting variables by observing customers' attitudes toward the product, brand, benefit, or even loyalty from the database. We can suggest that using average purchase amount instead of total accumulated purchase amount is better in order to reduce co-linearity of frequency and monetary. Finally, all customers are presented by 555, 554, 553, …, 111, which thus creates 125 (5×5×5) RFM cells. Moreover, the best customer segment is 555, while the worst customer segment is 111. Based on the assigned RFM behavior scores, customers can be classified into segments and their profitability can be further analyzed. The RFM score can be a basis factor how to determine purchasing behavior on the internet shopping mall, is helpful to buy the item which they really want by the personalized recommendation[5-6].

2.2 SOM

The SOM introduced by T. Kohonen, is an unsupervised learning algorithm for clustering[8]. Also SOM is called as a neural networks model based on competitive learning. SOM can convert a high dimensional input space into a simpler low dimensional discrete map. It has two layers which are input and feature layers. We

can cluster all elements by feature map with two dimensions. Firstly SOM performs clustering with input vector X and weight matrix W. The data point X_i is treated one at a time. Also the closest W_j to X_i is found by Euclidean distance, and then W_j is updated as the following [9].

$$W_k = W_j + \alpha\left(X_i - \mathrm{W}_j\right) \quad (1)$$

where W_j and W_k are current and new weights. So W_k moves to X_i. This learning is repeated until given conditions such as change rate of weights and the number of repeat. In this paper, we can use the SOM learning algorithm[8].

3 Our Proposal for a Personal Ontology Recommender System in u-Commerce

3.1 Clustering Method Using SOM Based on User's Information to Join User's Score

This approach used in this paper, clustering of the SOM is efficiently rather than clustering the data directly. First, a large set of prototypes (much larger than the expected number of clusters) is formed using the SOM or some vector quantization algorithm. The system can manage the user purchase patterns for recommending list according to the propensity of login user in u-commerce. In this section, we can describe the application for an efficient purchase pattern based on SOM. A SOM network is applied to segment user's information to join the user's score, different features, consisted of RFM factors as input vectors and finally forms clusters of user's information. An application of neural network using RFM factors of input vectors is studied, then there are two steps of the preparation for an efficient purchase pattern clustering. In the 1st step, we make the task of preprocessing so as to classify user's information using the code of classification, demographic variables such as age, gender, occupation, skin type, to recommend items according to the propensity of customer. In the 2nd step, we make the task of converting purchase data to join the user's information classified by user's propensity to SOM input data format of the bit pattern, for instance, the RFM factors needs 15 bits of pattern(each factor has 5 bits) in order to enter the SOM network. Finally, the prototyping is made, and then the prototyping result is classified in order to be possible to recommend by the loyalty of RFM factors as considering the propensity of customer. The clustering of the SOM can create an efficient purchase pattern by RFM factors as efficient way that would enhance customer knowledge, help to build brand loyalty, and increase customer satisfaction. The SOM learning algorithm for an efficient purchase pattern clustering is depicted as the following Table 1.

Table 1. The SOM learning algorithm for an efficient purchase pattern clustering

Step 1 : Initialize parameters of SOM model // Representative pattern of bits for RFM factors consists of 15 bits on patterns(each factor has 5 bits of patterns)
Step 2 : Set input value vector
Step 3 : Calculate Output value // Calculate distance between input pattern and connection weight vector of each neuron.

$$d_j = \sum_{i=0}^{n-1} (x_i - w_{ji})^2 \quad (2)$$

Step 4 : Select winner node
Step 5 : Readjust connection weights // Modify connection weight vector so that the neurons with short distance may be closer to input pattern.

$$w_{ij}(t+1) = w_{ij}(t) + \alpha(x_i(t) - w_{ij}(t)) \quad (3)$$

Step 6 : Completion of learning // *IF Reach the learning cycles then Make the result of SOM otherwise GO to Step 2*
Step 7 : Calculate output value
Step 8 : Calculate winner node
Step 9 : Result of pattern

3.2 The Procedural Algorithm for Recommendation

The system can search the information in the cluster selected by using the code of classification, then find an efficient purchase pattern for customer by the loyalty of RFM factors in the target group. It can scan the preference of brand item under the item category with the highest preference of target group, suggest the brand item with the highest score. This system can create the list of recommendation with TOP-N using brand item with the highest score based on RFM to be possible to measure the purchasability for the future, to recommend the item with purchasability. This system can recommend the items according to the basic the RFM factors through an efficient purchase pattern clustering based on SOM. The procedural algorithm for recommendation is depicted as the following Table 2.

3.3 The Analysis of Application by Procedural Algorithm for Recommendation

In this paragraph, we can describe the procedure of recommending in the result of SOM clustering. Of course, we can use the user group for the target login user, is deduced by clustering of the SOM using input data with input vectors as RFM factors. The system can suggest the list of recommending items with high purchasability to login user(customer) by Top-4. The login user's score is 96 points, so the system can use the user's propensity (age-gender, occupation, skin type, etc) and RFM factors. If login user(id = jery32, user's score = 96) had the condition of user's propensity, the code of classification(the range of age : 20~29, gender : female, occupation : student

Table 2. The procedural algorithm for recommendation

Step 1 : When the user joins the membership, user's information is created, managed the code of classification reflected demographic variable such as age, gender, occupation and skin type for customer.
Step 2 : The login user reads users' propensity by the code of classification and user's score.
Step 3 : The system can define the target of customers, the same of users' propensity, select the table of preference of item category for login user.
Step 4 : The system can read the preference of item category, then select item category by the highest probability in target group.
Step 5 : The system can select the brand item with the highest score in item category.
Step 6 : The system can create the list of recommendation with TOP-N so as to recommend the item with purchasability.
Step 7 : The system executes the cross comparison with purchased history data in order to avoid the duplicated re commendation which it has ever taken.

of university, skin type : dry) and user's score(score > = 90), the system could define the target group, extract user's group for the target group with same propensity of login user. The system can extract purchase data to join user's group with different features and RFM factors. The count of member is 12 customers as same user's propensity, the count of extracted purchase data is 53 counts, the count of level for score is 12 counts when user's score is greater than 90 point. In the step 3 in the sec. 3.2, the count of item category is 8 counts. we can show the Top-4 of the table of item category(8 counts) which is suggested to login user as the following table 3, then select item category by the highest probability in target group; for example, we can select the item category according to the probability for item category descendingly, from "AAC" to "AAA". .And then, we can select the brand item with the highest score in item category by each item category.

Table 3. Table of the Top-4 of item category and table of brand item with the highest score

Item Category	probability	score	Item Category	probability	score
AAD	50.84	80	ABA	42.86	83
BAB	50.84	80	ACC	37.04	71
CAE	44.44	80	AAC	17.38	46
AAA	42.86	85	CAD	4.55	26

Brand Item	probability	score
AAD20	50.84	80
BAB02	50.84	80

Brand Item	probability	score
CAE18	50.00	80
AAA35	44.44	85

4 Experimental Result

We make the implementation for prototyping of internet shopping mall which handles the cosmetics professionally and do experiments. It is the environment of

implementation and experiments in Jena semantic web framework 2.6.3 as inference engine, j2sdk 1.7.0_11 as Java environment, JSP/PHP 5.2.12 as server-side script, JQuery* mobile, XML/XHTML4.0/HTML5.0/CSS3/JAVASCRIPT as client-side script, C# .net framework 2.0, jakarta-tomcat, apache 2.2.14 as web server under MS Windows XP sp3(64).

4.1 Experimental Data for Evaluation

We used 319 users who have had the experience to buy items in e-shopping mall, 580 cosmetic items used in current industry, 1600 results of purchase data recommended in order to evaluate the proposing system[4]. In order to do that, we make the implementation for prototyping of the internet shopping mall which handles the cosmetics professionally and do the experiment. We have finished the system implementation about prototyping recommendation system. We'd try to carry out the experiments in the same condition with dataset collected in a cosmetic internet shopping mall. It could be evaluated in MAE and Precision, Recall, F-measure for the recommendation system in clusters. It could be proved through the experiment with learning data set for 12 months, testing data set for 3 months in a cosmetic cyber shopping mall[4]. The 1st system of clustering method using SOM based on purchase data, is proposing system called by "proposal", the 2nd system is other previous system(k-means) using k-means algorithm, the third system is existing system based on the whole data.

4.2 SOM Results

The clustering techniques in data mining can be used for the customer segmentation process so that it clusters the customers in such a way that the customers in one group behave similar when compared to the customers in the other group based on their transaction details. In this paper, we propose an efficient purchase pattern clustering based on SOM using RFM factors and compare the performance against the traditional techniques like K-means. The k-means algorithm which is built on user's setting initial seed can generate artificial clustering, but there is a sparse possibility of generating artificial clustering in SOM which is built into learning network. So, for this reason, we carry out the experiments in the same condition by comparing proposal system(SOM) with previous system(k-menas). SOM network is applied to segment purchase data to join the user's information, finally forms clusters of purchase pattern groups of user's information with different features, demographic variables and RFM factors as input vectors. In order to segment purchase data join to user's information into appropriate number of clusters, SOM is applied to determine the number of clusters, the output is a number between 0 and 1, so the value is needed to turn it back to the original measure. Below figure 1 and figure 2, nine clusters are recommended among 1,600 purchase data, 319 customers when recency(r), frequency(f) and monetary(m) are the three input variables and then divide the data

into five equal quintiles beside off demographic variables. From the SOM result, we find 9 segments for STR(Segment, Target, Recommendation) strategy modeling so as to recommend the items by the loyalty of RFM factors in real-time environment. The following figures show the result of SOM clustering using user's information based on RFM factors for the segmentation.

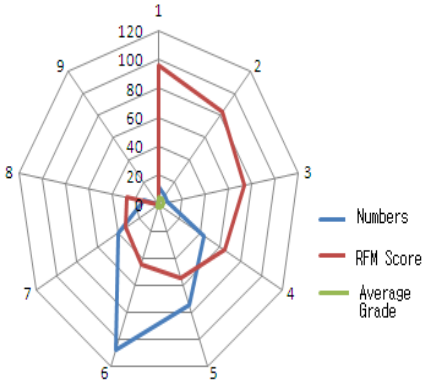


Fig. 1. Descriptive statistics of nine clusters based on SOM

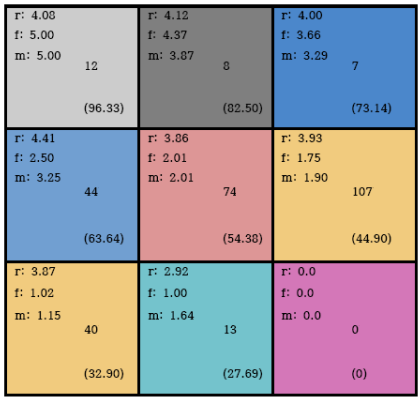


Fig. 2. Use a 3 x 3 Map neuron for data clustering

4.3 Experiment and Evaluation

We can make the task of clustering of SOM based on purchase data for preprocessing under ubiquitous computing environment. The proposing system's overall performance evaluation was performed by dividing the two directions. The first evaluation is mean absolute error(MAE). The mean absolute error between the predicted ratings and the actual ratings of users within the test set. The mean absolute error is computed the following expression-4 over all data sets generated on purchased data.

$$MAE = \frac{\sum_{i=1}^{N} |\epsilon_i|}{N} \tag{4}$$

N represents the total number of predictions, ε represents the error of the forecast and actual phase i represents each prediction.

The next evaluation is precision, recall and F-measure for proposing system in clusters. The performance was performed to prove the validity of recommendation and the system's overall performance evaluation. The metrics of evaluation for recommendation system in our system was well-known and used in the field of information retrieval commonly[10].

Table 4. The result for table of MAE by comparing proposal system with previous system and existing system

	P_count	Proposal (SOM)	Previous (k-means)	Existing
MAE	50	0.23	0.47	0.65
	100	0.13	0.23	0.32
	300	0.04	0.07	0.08
	500	0.03	0.05	0.06

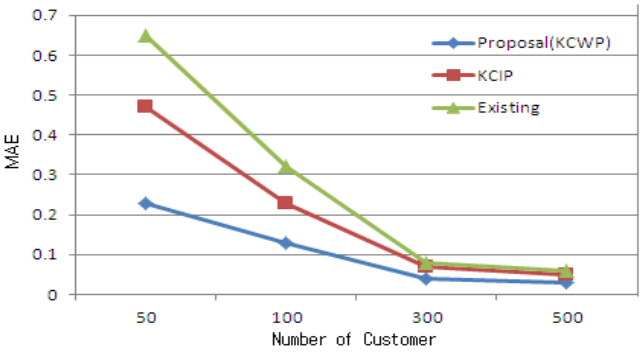


Fig. 3. The result for the graph of MAE by comparing proposal system with previous system and existing system

Table 5. The result for table of precision, recall, F-measure for recommendation ratio by each cluster

Cluster	proposal(SOM)			Previous (k-means)			Existing		
	Precision1	Recall1	F-measure1	Precision2	Recall2	F-measure2	Precision3	Recall3	F-measure3
C1	83.33	23.53	36.06	36.11	3.23	5.89	55.19	5.95	10.47
C2	61.11	17.65	26.39	34.52	16.13	21.14	38.97	15.18	20.88
C3	75.00	31.25	43.04	45.68	33.60	36.42	48.79	31.32	35.64
C4	75.00	12.50	21.11	51.67	9.60	15.40	50.22	11.74	18.28
C5	66.67	37.50	46.02	44.27	19.20	25.84	50.60	13.88	21.03
C6	79.17	57.14	64.05	38.83	32.49	33.64	44.26	21.81	27.65
C7	81.82	57.14	65.67	42.93	54.79	44.89	50.93	36.60	39.64
C8	71.43	33.33	44.29	35.33	39.73	36.60	43.60	36.60	37.82
C9	80.00	55.56	63.91	38.17	29.79	31.99	46.53	18.32	25.10

Fig. 4. The result of recommending ratio by precision

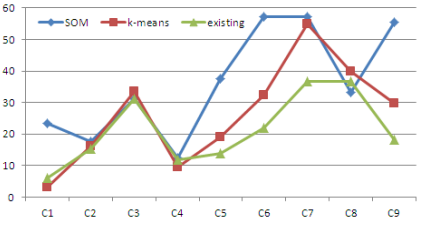


Fig. 5. The result of recommending ratio by recall

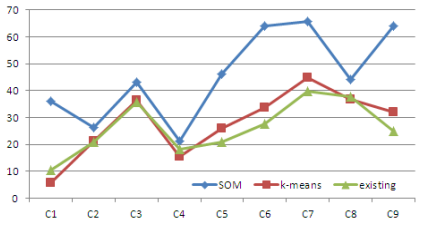


Fig. 6. The result of recommending ratio by F-measure

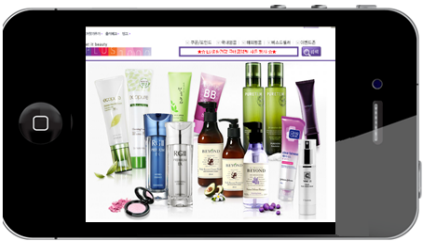


Fig. 7. The site of recommendation of cosmetics

Above Table 5 presents the result of evaluation metrics (precision, recall and F-measure) for recommendation system. An efficient purchase pattern clustering based on SOM(SOM) is improved better performance of proposing system than the previous systems(k-means). Our proposing system with an efficient purchase pattern clustering based on SOM is higher 34.01 % in precision, higher 9.67% in recall, higher 17.64% in F-measure than the previous system(k-means). As a result, we could have the recommendation system to be able to recommend the items with high purchasability. The figure 7 is shown in the site of recommendation of cosmetics on a smart phone. Our proposing system with an efficient purchase pattern clustering based on SOM is better performance than the previous method.

5 Conclusion

Recently u-commerce as an application field under ubiquitous computing environment required by real time accessibility and agility, is in the limelight[4]. SOM network is applied to recommend the items with an efficient purchase pattern by the loyalty of RFM factors in real-time environment, then we proposed an efficient purchase pattern clustering based on SOM for recommender system in u-commerce in order to improve the accuracy of recommendation by the loyalty of RFM factors. We have described that the performance of the proposing system with an efficient purchase pattern clustering based on SOM is improved better than the previous system(k-means) and existing system. We could simulate personal ontology recommender system, generate recommending items with high purchasability. Thus,

we could build inferential recommender engine of the system with focus on accuracy and efficiency, and our results validate the system. To verify improved better performance of proposing system, we carried out the experiments in the same dataset collected in a cosmetic internet shopping mall. It is meaningful to present a new framework of efficient purchase pattern clustering method for personal ontology recommender system in u-commerce to be recommended by the loyalty of RFM factors under ubiquitous computing environment. The following research will be looking for ways of a personalized recommendation using fuzzy clustering method to increase the efficiency and scalability.

Acknowledgements. This work[1)] was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST) (No. 2012-0000478) and this paper[2)] was supported by funding of Namseoul University.

References

1. Deboeck, G., Kohonen, T.: Visual Explorations in Finance with Self-Organizing Maps. Springer, London (1998)
2. Cho, Y.S., Jeong, S.P., Ryu, K.H.: Implementation of Personalized u-commerce Recommendation System using Preference of Item Category based on RFM. In: The 6th International Conference on Ubiquitous Information Technologies & Applications, pp. 109–114 (December 2011)
3. Cho, Y.S., Moon, S.C., Ryu, K.H.: Mining Association Rules using RFM Scoring Method for Personalized u-Commerce Recommendation System in emerging data. In: Kim, T.-h., Ramos, C., Abawajy, J., Kang, B.-H., Ślęzak, D., Adeli, H. (eds.) MAS/ASNT 2012. CCIS, vol. 341, pp. 190–198. Springer, Heidelberg (2012)
4. Cho, Y.S., Noh, S.C., Moon, S.C.: Weighted Mining Association Rules Based Quantity Item with RFM Score for Personalized u-Commerce Recommendation System. In: Park, J.J(J.H.), Arabnia, H.R., Kim, C., Shi, W., Gil, J.-M. (eds.) GPC 2013. LNCS, vol. 7861, pp. 367–375. Springer, Heidelberg (2013)
5. Cho, Y.S., Moon, S.C., Jeong, S.P., Oh, I.B., Ryu, K.H.: Clustering Method using Item Preference based on RFM for Recommendation System in u-Commerce. In: Han, Y.-H., Park, D.-S., Jia, W., Yeo, S.-S. (eds.) Ubiquitous Information Technologies and Applications. LNEE, vol. 214, pp. 353–362. Springer, Heidelberg (2012)
6. Wei, J.-T., Lin, S.-Y., Wu, H.-H.: The review of the application of RFM model. African Journal of Business Management 4(19), 4199–42060 (2010)
7. Smith, K.A., Gupta, J.N.D.: Neural Networks in Business: Techniques and Applications. The Idea Group Publishing (2001)
8. Kohonen, T.: Self-Organizing Maps. Springer (2000)
9. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning – Data Mining, Inference, and Prediction. Springer (2001)
10. Herlocker, J.L., Kosran, J.A., Borchers, A., Riedl, J.: An Algorithm Framework for Performing Collaborative Filtering. In: Proceedings of the 1999 Conference on Research and Development in Information Research and Development in Information Retrival (1999)

Collusion-Resistant Watermarking Using Modified Barni Method

Hyunho Kang and Keiichi Iwamura

Dept. of Electrical Engineering, Tokyo University of Science,
6-3-1 Niijuku, Katsushika-ku, Tokyo 125-8585, Japan
{kang,iwamura}@ee.kagu.tus.ac.jp

Abstract. Collusion attacks are effective against fingerprinting schemes. When a watermark is embedded as part of a fingerprinting schemes, it should be resistant to collusion by multiple users who each have a different version of the watermark. However, there is a limitation to detecting the watermark when classical spread spectrum watermarking is used. In this paper, we propose a negative correlation watermarking scheme to resolve this limitation using a modified Barni's watermarking scheme, which is a method that can be used in many application areas. In particular, the difference between the frequency coefficients and uniformly distributed real numbers is used as the embedded watermark. This method can also be applied to our previously proposed wavelet-based video fingerprinting system.

1 Introduction

Fingerprinting involves embedding a mark (a fingerprint) inside content to encode a user's identity. Fingerprints should not be easy to detect or remove, and they must be designed such that forgery is difficult or expensive [1].

There have been many techniques proposed in the literature to address the fingerprinting issue. Initially, digital fingerprinting was considered as a coding technique, such as the Boneh and Shaw (BS) algorithm [2]. Boneh and Shaw presented the concept and scheme of digital fingerprinting. However, when such fingerprint code is regarded as a watermark, embedded into a copy of the digital content, the copy constructed with the fingerprint cannot be denoted by the BS model, because the watermarks overlap each other in the digital content [3]. Moreover, the fingerprint code is too long to be implemented in practice.

One possible solution to overcome this problem is to use a spread spectrum (SS) sequence. Cox et al. first indicated that SS sequences have collusion resistance properties [4]. However, to recover the watermark, their algorithm requires the original image (which is called the non-blind technique) [5]. Existing detectors can be classified into two types: blind and non-blind detectors. Blind detectors are more focused because of the large amount of digital multimedia. The blind technique proposed in [6][7] is applicable to many application areas from a practical viewpoint.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_80,

However, there is a limitation to detecting the watermark when using Cox et al.'s algorithm in fingerprinting applications. In this paper, we propose a negative correlation watermarking scheme to resolve this limitation using a modified Barni et al.'s watermarking scheme. This method can also be applied to our previously proposed wavelet-based video fingerprinting system [8][9] and another application [10].

The rest of the paper is organized into the following sections: Section 2 briefly describes Barni et al.'s method and the proposed algorithm, including watermark construction, embedding, and detection. Experimental results and conclusions are given in Sections 3 and 4, respectively.

2 Proposed Method

2.1 The Preparatory Stage-Barni's Method

In [6], the watermark consists of M randomly generated real numbers, so $W = \{w_1, w_2, ..., w_M\}$, where each value w_i is a Gaussian random variable having zero mean and unity variance. The DCT is computed on the entire image, and the watermark is superimposed on some middle-range coefficients. The $L + M$ DCT coefficients, reordered into a zigzag scan as in the JPEG compression algorithm, are selected to generate a vector $E = \{e_1, e_2, ..., e_L, e_{L+1}, ..., e_{L+M}, ...\}$. To ensure perceptual invisibility without loss of robustness, the lowest L coefficients are not used for embedding; then, a watermark W is embedded in the last M numbers. A new vector $E' = \{e_1, e_2, ..., e_L, e'_{L+1}, ..., e'_{L+M}, ...\}$ is obtained according to the following rule:

$$e'_{L+i} = e_{L+i} + \alpha \cdot |e_{L+i}| \cdot w_i, \quad i = 1, 2, ..., M, \tag{1}$$

where α is a parameter for watermark strength. The vector E' is then reinserted into the zigzag scan, and inverse DCT is performed to yield the watermarked image I'.

2.2 Watermark Construction

In the proposed algorithm, first, for constructing a watermark W, we compute two elements (the largest and smallest values) from the embedded area. In this paper, we will use the term "E_a" to refer to the embedded area (see Fig. 1).

Second, we generate a uniform distribution of random numbers on a specified interval $[max(E_a), min(E_a)]$, denoted by P_{mark}. Third, we get the embedded watermark using Eq.2:

$$W = E_a - P_{mark}, \tag{2}$$

where P_{mark} is a sequence of uniformly distributed pseudorandom numbers.

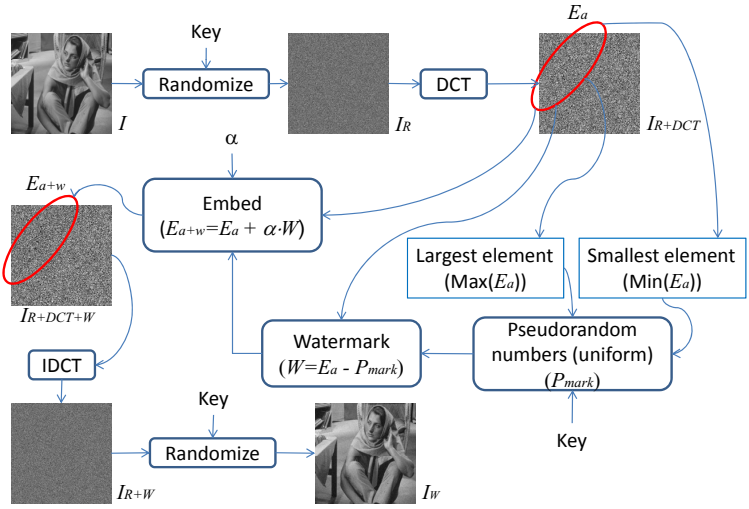


Fig. 1. Embedding process including watermark construction

2.3 Watermark Embedding

The embedding of the watermark w_i into the host signal x_i is usually either multiplicative or additive. Generally, the multiplicative rule $y = x_i(1 + \alpha_i \cdot w_i)$ is used for embedding the watermark. In the frequency domain, to improve watermark detectability, Barni *et al.* proposed the multiplicative rule $y_i = x_i + \alpha_i \cdot |x_i| \cdot w_i$ [6].

Because the watermark values have a negative property, we consider an additive embedding rule, as shown in Eq.3. In the proposed method, a unique watermark is inserted in the DCT domain with strength α. We define E_a as the embedded area after processing the image using randomization and DCT. We then define E_{a+w} as the watermarked area constructed using Eq.3:

$$E_{a+w} = E_a + \alpha \cdot W. \tag{3}$$

Figure 1 shows the embedding process that includes watermark construction.

2.4 Watermark Detection

The goal of this research is to efficiently detect the watermark (a fingerprint in a fingerprinting application) under collusion attacks. The proposed detection process uses linear correlation, and is illustrated in Fig. 2. We obtain the pseudorandom numbers relating to the embedded watermark in the embedding step using Eq. 4. Note that the numbers for detecting the watermark are not the same as for constructing it:

$$corr = \frac{1}{M} \sum E_{a+w} \cdot P'_{mark} \tag{4}$$

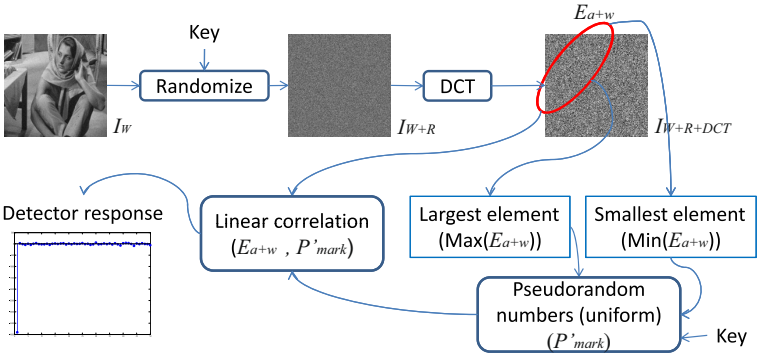


Fig. 2. Watermark detection process

where E_{a+w} is the watermarked area, M is the size of the embedded area, and P'_{mark} is a set of uniformly distributed pseudorandom numbers. However, P'_{mark} is a scaled and possibly shifted version of P_{mark}, although it is obtained using the same key.

3 Experimental Results

We tested the proposed algorithm on the 512×512 pixel "Barbara" gray-scale image. Figures 3∼8 show the results of the negative correlation detection experiment. In these experiments, we chose and $\alpha = 0.2$ and $M = 20\%$ (the length of a changed region). We examined 50 watermarks generated using different keys.

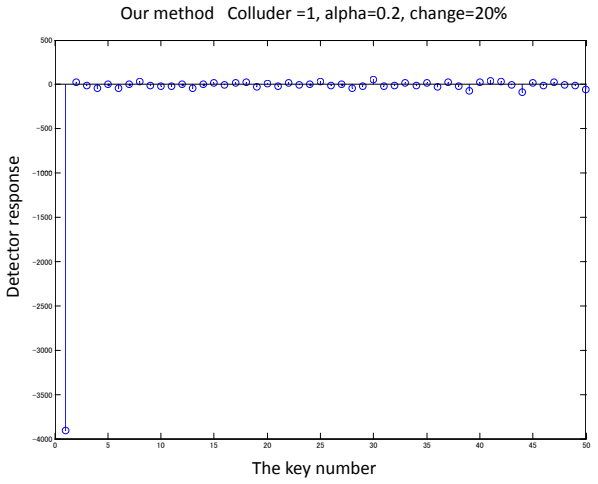


Fig. 3. Result of negative correlation (no colluders)

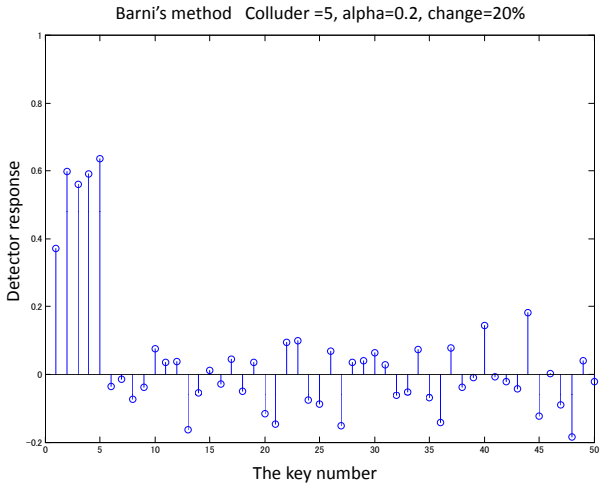


Fig. 4. Barni's detection result(with five colluders)

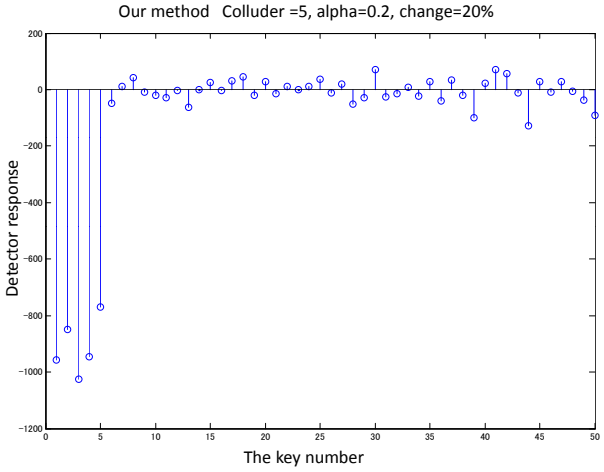


Fig. 5. Our detection result(with five colluders)

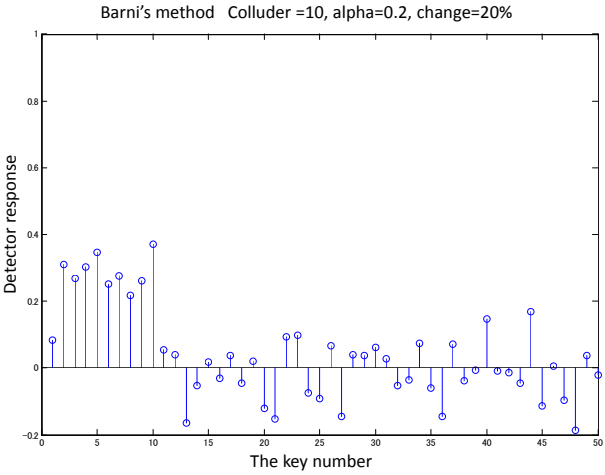


Fig. 6. Barni's detection result(with ten colluders)

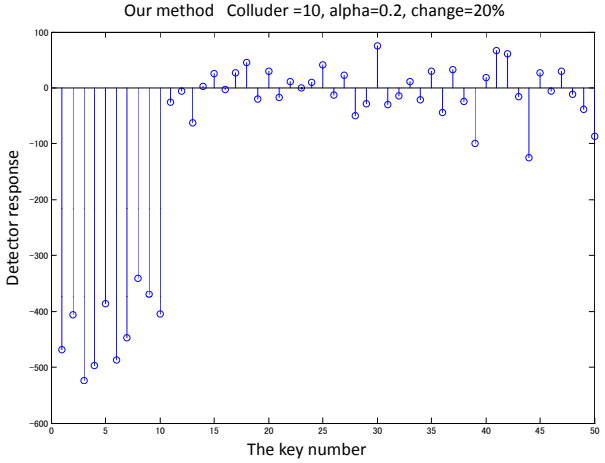


Fig. 7. Our detection result(with ten colluders)

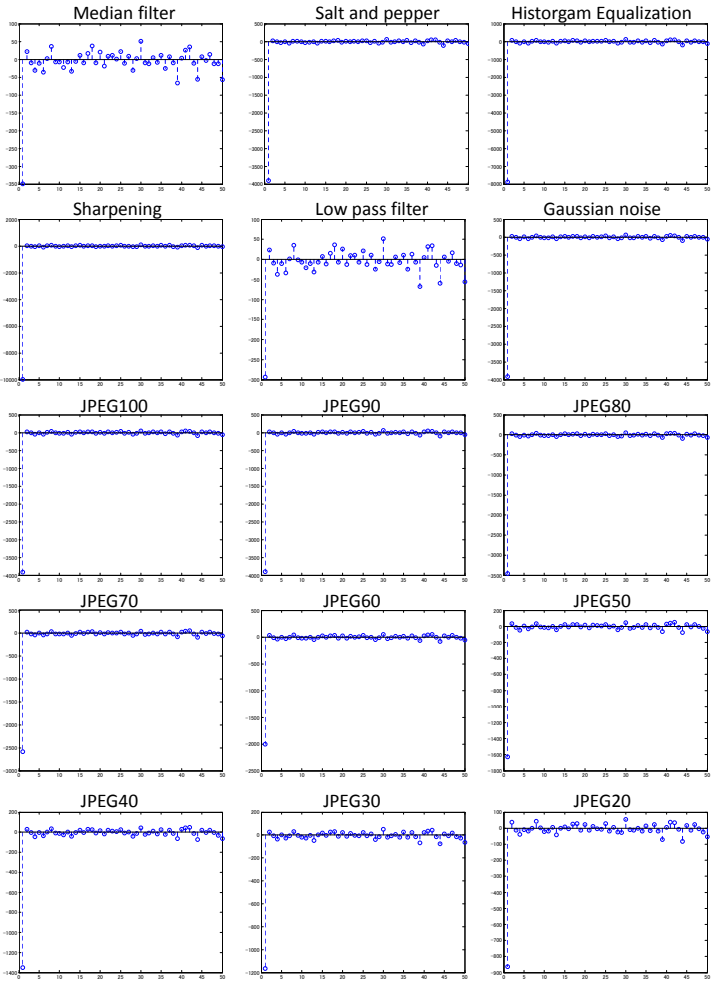


Fig. 8. Result of negative correlation detection after basic image processing and JPEG compression. The detector responded to 50 randomly generated watermarks, and the correct watermark with a negative direction is shown in the graph. The true key is found at the number "1."

As shown in Fig. 3, the correct watermark with a negative direction is detected at position key number "1." We also tested watermark performance under a collusion attack (using averaging). Figure 4∼5 show Barni *et al.*'s detection result and our result with five colluders. Figure 6∼7 show the results for when there are ten colluders, and Fig. 8 shows the results of negative correlation detection after basic image processing and JPEG compression.

4 Conclusions

In this paper, we proposed a framework that can improve correlation detection performance for collusion resistance. In particular, the difference between the frequency coefficients and uniformly distributed real numbers was used as the embedded watermark. This watermark had the important property of negative direction. Therefore, the experimental results demonstrated that the watermark had a negative direction and was resistant to collusion attacks, including basic image processing and JPEG compression.

References

1. van Tilborg, H.C.A.: Encyclopedia of Cryptography and Security. Springer Science+Business Media, Inc. (2005)
2. Boneh, D., Shaw, J.: Collusion-Secure Fingerprinting for Digital Data. In: Coppersmith, D. (ed.) CRYPTO 1995. LNCS, vol. 963, pp. 452–465. Springer, Heidelberg (1995)
3. Zhu, Y., Feng, D., Zou, W.: Collusion Secure Convolutional Spread Spectrum Fingerprinting. In: Barni, M., Cox, I., Kalker, T., Kim, H.-J. (eds.) IWDW 2005. LNCS, vol. 3710, pp. 67–83. Springer, Heidelberg (2005)
4. Cox, I., Kilian, J., Leighton, T., Shamoon, T.: Secure Spread Spectrum Watermarking for Multimedia. IEEE Transactions on Image Processing 6(12), 1673–1687 (1997)
5. Wang, J., Liu, G., Dai, Y., Sun, J., Wang, Z., Lian, S.: Locally optimum detection for Barni's multiplicative watermarking in DWT domain. Signal Processing 88, 117–130 (2008)
6. Barni, M., Bartolini, F., Cappellini, V., Piva, A.: Robust watermarking of still images for copyright protection. In: In Proc. 13th Inter. Conf. Digital Signal Processing, vol. 2, pp. 499–502 (1997)
7. Barni, M., Bartolini, F., Piva, A.: Improved wavelet-based watermarking through pixel-wise masking. IEEE Transactions on Image Processing 10(5), 783–791 (2001)
8. Kang, H.-H., Kurkoski, B., Park, Y.-R., Lee, H.-J., Shin, S.-U., Yamaguchi, K., Kobayashi, K.: Video Fingerprinting System using Wavelet and Error Correcting Code. In: Song, J.-S., Kwon, T., Yung, M. (eds.) WISA 2005. LNCS, vol. 3786, pp. 150–164. Springer, Heidelberg (2006)
9. Kang, H.-H., Kurkoski, B., Yamaguchi, K., Kobayashi, K.: Tracing Illegal Users of Video: Reconsideration of Tree-Specific and Endbuyer-Specific Methods. In: Gervasi, O., Gavrilova, M.L. (eds.) ICCSA 2007, Part III. LNCS, vol. 4707, pp. 1046–1055. Springer, Heidelberg (2007)
10. Kang, H., Park, Y., Kurkoski, B., Yamaguchi, K., Kobayashi, K.: Watermarking with permissible alterations. In: The 2007 Symposium on Cryptography and Information Security (SCIS 2007), Sasebo, Japan, January 23-26 (2007)

Design and Implementation of a High Integrated Noncontact ECG Monitoring Node for Wireless Body Sensor Networks

Fangmin Sun[1,2], Zhan Zhao[1], Zhen Fang[1], Lidong Du[1], Yangming Qian[3], Huaiyong Li[3], and Lili Tian[3]

[1] State Key Laboratory of Transducer Technology, Institute of Electronics, Chinese Academy of Sciences, Beijing, China
[2] Graduate University of Chinese Academy of Sciences, Beijing, China
[3] Navy General Hospital of PLA, Beijing, China
sfm0719@163.com, {zhaozhan,zfang,lddu}@mail.ie.ac.cn, qymbright@gmail.com

Abstract. An ease of use, convenient and high compact noncontact electrocardiograph (NCECG) monitoring node for wireless body sensor network is presented in this paper. Compared with many other two electrodes NCECG monitoring node, it has the advantages of noncontact, no infection and less time-consuming. It uses doubled shielded active ECG electrodes and, adds a right-leg-drive circuit to reduce the common mode noise. The experiment results show that the designed NCECG node could accurately monitor the ECG signal under the circumstances that the electrodes are insulated by one layer of clothes. Besides ECG monitoring function, the NCECG node also integrates with temperature, respiration and motion state monitoring function, while the size of the circuit board is just 93 mm *40mm.

Keywords: Noncontact electrodes, double shielded, ECG monitoring, wireless, body sensor network.

1 Introduction

Among BSN technology, the ECG plays a central role in the rapid diagnosis of heart diseases such as coronary heart disease, ischemic heart disease, myocardial infarction, arrhythmias, etc. Unfortunately, the adhesion of conventional electrodes to the skin sometimes is difficult if not impossible due to the wet skin caused by higher perspiration. Besides, for the application of long term ECG monitoring, the wet adhesive ECG electrodes are easy to cause infection of the wearers' skin. And moreover, in certain situations it is difficult or simply too time-consuming to undress the patient before acquiring an ECG. Therefore, novel ECG monitoring techniques are much in demand.

Nowadays, lots of new types of ECG monitoring electrodes are designed and proposed. Ha-Chul Jung described the fabrication of the PDMS/CNT flexible dry

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*, Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_81,

electrodes for long-term ECG monitoring [1]. It's electrodes have the advantages of flexible and comfortable for the wearers, nevertheless the electrical property can be affected by many factors such as the density of the CNT, the fabrication techniques, etc. Patrick Griss designed micro needles for bio-signal extracting [2]. Although it's electrodes have comparatively low contact impedance, it may cause pain to the wearer whose stratum corneum is thin as the electrode need to insert into the skin. The improvement method of micro needle: carbon nanotube (CNT) array is designed by Giulio Ruffini [3]. Though the CNT array has wonderful electrical performance and is comfortable for the patient to wear, the fabrication process of this kind of electrodes is complex. And another kind of dry electrode which is used in many studies is the active electrode [4], the active electrode is usually consist of a preamplifier circuit and an active shield, and this kind of electrode is easy to fabricate and has relatively better performance.

The main contribution of our work is the design of a noncontact ECG monitoring wearable chest belt which makes it possible to continuous and long-term monitoring of people's health conditions at anytime and anywhere and without causing irritation. Besides, compared with the studies of other noncontact ECG monitoring nodes, the NCECG node we designed has the advantages of small size, and ease of use, etc. And in order to reduce the common mode interference, we also added a right leg drive (RLD) circuit with the RLD electrode integrated in the back of the circuit board which could reduce wires to some extent. The structure of the chest belt NCECG node is shown in Fig.1.

(a) The outside of belt (b) The inside of belt

Fig. 1. The wearable chest belt

The remainders of this paper are divided as the following: in section 2, we introduce double shielded active electrode design. And in the section 3, we described the hardware design of our NCECG nodes. The performance of the NCECG node is tested and the experiment results are presented in section 4. And then we summarized the paper and described the future development of the NCECG belt.

2 Active Electrodes Design

According to the measurement principle of the active electrode, the capacitive coupling feature of the electrode is easy to be affected by the outside environment and the contact impedance between the electrode and the skin is usually higher than the contact

electrode. So shielding and improving the input impedance of the electrode are the two key techniques in the design of the active electrode.

In the designed NCECG node, the active electrodes is designed to sample the ECG signal indirectly trough clothes, and ECG measurement through clothes is characterized by very high electrode impedance (the impedance between the electrode and the skin). To improve the input impedance of the electrode and reduce the noise we take the following measures: 1). we designed the high impedance front end circuit to improve the input impedance of the electrode; 2). the double shield is designed and added to the electrode to reduce the electromagnetism interference from the outside environment. The architecture of the double shield active electrode can be shown in Fig.2.

In order to make up for the rather high coupling impedance, insulated electrodes need direct amplification close to the coupling surface. For the front end circuit design, each sensor consists of two small round electrically connected standard printed circuit boards and the radius of them just 15mm. The upper board contains a high input impedance amplifier LMP7701 which serves as an emitter follower to improve the input impedance, the bottom layer of the board is covered with copper which serves as one of the electrodes of the capacitor made from the skin and the electrode and to sensing the ECG signal.

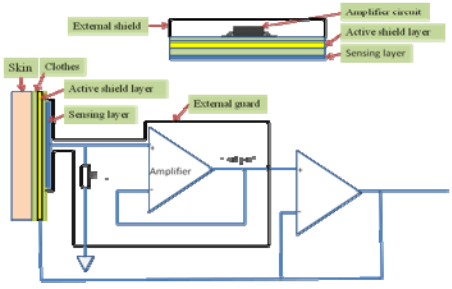


Fig. 2. The model of the double shield active electrode

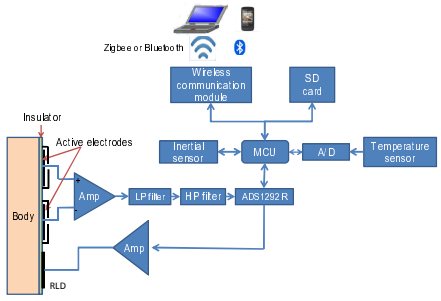


Fig. 3. The architecture of the NCECG monitoring node

Also, since capacitive measurements are rather sensitive to noise, these electrodes typically need electromagnetism shielding. We take the double shield structure to reduce the external electromagnetism interference: middle layer active shield and external ground shield. The middle layer active shield has the same voltage potential with the sensing layer and could effectively reduce the parasitic capacitance of the sensing layer. And the external shield is connected to the ground, and could largely reject the external electromagnetic interference such as the industrial frequency interference.

3 Hardware Design of the Node

The hardware framework of the NCECG monitoring node is depicted in Fig.3. It mainly includes four function modules: microcontroller module, sensing and signal processing module, wireless communication module and power management module. In this section, we will discuss these modules respectively.

3.1 Sensing Module

As introduced previously, the NCECG monitoring node not only could accurately sample the ECG signal but also could monitoring the body temperature and the motion state of the wearer. The ECG signal is sensed by the double shield active electrodes closed to the skin and a right leg driven electrode is added to reduce the common interference.

Body temperature is also an important factor for the diagnosis of many diseases, and the NCECG monitoring node also integrated the body temperature sensor for the continuous monitoring of the body temperature. For the temperature measurement, the YSI4000 family temperature sensor is selected to capture the body temperature signal, and The AD7783 is taken for the temperature measurement, which adopts the ratio metric method. The transducer excitation (voltage or current) on the analog input drives the reference voltage for this part, thus the effect of the low frequency noise in the excitation source can be removed.

In order to evaluate, select and self-correct the monitoring parameters according to the motion state of the wearer, which can avoid and reduce daily activity on the physiological signal measuring, and can improve the monitoring quality and health expert diagnostic efficiency, we integrated an inertial sensor CMA3000D01 in the ECG monitoring node for real-time monitoring of the wearer's behavioral state.

3.2 Signal Processing Module

The ECG signal sampled from the active electrode is first be filtered by a 100Hz LP filter and a 0.05Hz HP filter and a notch filter whose notch frequency is 50Hz is designed to reduce the power line interference. And then the filtered signal is converted into digital signal by the Analog/digital device.

We chose the analog to digital converter (ADC) for the node mainly based on the following principles: 1) ultra low power consumption; 2) high resolution; 3) highly integrated; 4) small size. According to these principles, we finally adopt TI's 24 bit ADS1292R as the ADC. The ADS1292R incorporate all of the features that are required in our design of a compact and low-power ECG and respiration monitoring node. With its exceptional performance and high levels of integration, the ADS1292R make the front-end ECG signal processing circuits very simple and largely reduced the physical size and power consumption of the nodes. As previously mentioned, the ADC digital data lines are connected as a daisy chain shifting serial data from the end of the sensor chain back to the host data acquisition module. A common clock and chip select line synchronizes the conversion and transfer of data at a sample rate of 500sps.

3.3 Wireless Communication Module

The sensor is integrated with two types of wireless communication modules: the Zigbee and the Bluetooth. The Zigbee transceiver circuit is made up of the radio of the CC2530 chip and a printed circuit antenna. The radio of the CC2530 chip can easily be built on top of the IEEE 802.15.4 based standard protocols such as RemoTI, TIMAC, and Z-Stack. However, take the power consumption of these protocols into consideration, we finally choose the Z-Stack protocol as our application platform. For the design of the antenna, we made a small 2.4GHz antenna in the printed circuit board which largely reduced the size of the node compared with the whip antenna. To realize the communication between the node and other Bluetooth devices such as mobile phones, the node is also integrated a Bluetooth module. With these two communication methods, the physiological information can be easily send to the remote data base for storage or clinical center for pathology analysis through the Zigbee, or just send to a nearby Bluetooth device to be shown.

3.4 Power Management Module

To improve the energy efficiency, we added a dynamic power management (DPM) module and a dynamic voltage scaling (DVS) module to the ECG monitoring node operation system, which would make the node using its resources more efficiently. When there is nothing interesting happen around the node, some modules of the nodes are in the idle mode, and the DPM module will turn off these modules or turn them into sleep mode to save more energy. The DVS module calculates the loads of the power source, and when the loads is lower than the threshold it will reduce the work frequency and voltage of the microprocessor to reduce the process ability of the microprocessor module and thus reduce the power consumption of the processor.

And in the low power mode which is set by the button in the device, the wireless transmission module is closed and the physiological signals could be stored in the mini SD card automatically, relative people could read the SD card when needed. This, to some extent, largely reduced the power consumption of the sensor node. And when the NCECG node is not in use, it could be set into the sleep mode through push the button longer, and in the sleep mode the MCU is in the low-power mode and the sensing module and the wireless communication module are all suspended. And in the same way, the device can be waked up by long time pushing the button.

4 Experiment Results and Conclusions

In this section, we tested the performance of the NCECG node included the frequency response of the electrode, the power consumption of the node and the monitoring function of the node. The test results are given and analyzed.

4.1 The Frequency Response of the Electrode

To know the frequency response of the active electrode, a differential test frequency sweep was applied to the metal plates from a function generator. The sensor output, along with the test input was recorded through a digital oscillograph, allowing the gain to be measured. And in this experiment, the insulator medium between the electrode metal plate and the function generator electrode is a piece of cotton cloth with the thick of 1.2μm, and the test result is shown in Fig.4. As the frequency of the ECG signal is mainly ranging from 2Hz to 20Hz, so we can see through the result that active electrode has a very good performance to sample the human ECG signal.

4.2 The Health Monitoring Function of the NCECG Node

Measurement of ECG, temperature and acceleration value of one of the wearers is shown in Fig.5. The ECG signal was sampled from the chest over a cotton T shirt which is about 1um thick.

A clear ECG signal could be observed when the sensor is placed directly above the skin's surface. However, when the recording is made over insulated medium, some 50Hz line noise is easy to be introduced as a result of capacitive mismatch due to the larger separation distance and corresponding smaller coupling capacitance. To reduce the 50Hz power line interference, we designed a 50Hz notch filter with the labview software.

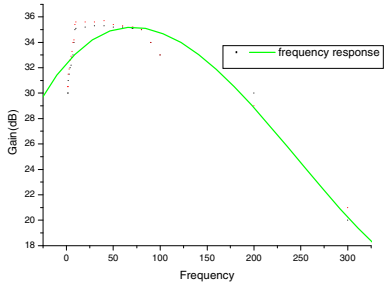


Fig. 4. The frequency response of the active electrode (insulator medium: cotton)

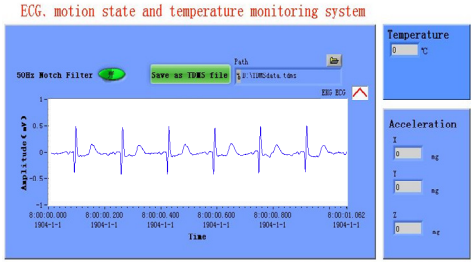


Fig. 5. The functional test result of the NCECG node

5 Conclusion

This paper described the design of a compact, low-power, small-size and high-integrated health monitoring sensor node which could be used to monitor multi physiological signals of human comfortably for a long time. And we mainly discussed the double shielded active electrode design and the hardware design of the electric circuit. At the end, we studied the frequency response of the active electrode and tested

the functions of the node and then we did experiments to analyze the power consumption of the node when it is in different mode. And the test result shows that this high compact noncontact ECG monitoring node could accurately measure the wearer's ECG signal, body temperature and motion state while consume relatively lower power.

Acknowledgment. The paper is based on research funded through 863 Program under Grant no. 2013AA041201 and 2012AA040504.

References

[1] Lee, J.H., Nam, Y.W., Jung, H.-C., Baek, D.-H., Lee, S.-H., Hong, J.S.: Shear induced CNT/PDMS conducting thin film for electrode cardiogram (ECG) electrode. BioChip J., 91–98 (2012)

[2] Griss, P., Enoksson, P., Tolvanen-Laakso, H.K., Meriläinen, P., Ollmar, S., Stemme, G.: Micromachined Electrodes for Biopotential Measurements. Journal of Microelectromechanical Systems 10(1), 10–16 (2001)

[3] Ruffini, G., Dunne, S., Farrés, E., Marco-Pallarés, J., Ray, C., Mendoza, E., Silva, R., Grau, C.: A dry electrophysiology electrode using CNT arrays. Sensors and Actuators A 132, 34–41 (2006)

[4] Chi, Y.M., Deiss, S.R.: Gert Cauwenberghs. Non-contact Low Power EEG/ECG Electrode for High Density Wearable Biopotential Sensor Networks. 2009 Body Sensor Networks, 246–250 (2009)

A Wearable Multi-parameter Physiological System

Zhihong Xu[1,2], Zhen Fang[1], Lidong Du[1], Zhan Zhao[1], Xianxiang Chen[1], Diliang Chen[1], Fangmin Sun[1], Yangming Qian[3], Huaiyong Li[3], and Lili Tian[3]

[1] Institute of Electronics, Chinese Academy of Sciences, Beijing, China
[2] Graduate University of Chinese Academy of Sciences, Beijing, China
[3] Navy General Hospital of PLA, Beijing, China
xuzhihong111@163.com, {zhaozhan,zfang,lddu}@mail.ie.ac.cn, qymbright@gmail.com

Abstract. We developed and tested a health monitoring device capable of measuring a subject's ECG, blood pressure, blood oxygenation, respiration, temperature and motion – almost equivalent to the feature set of a hospital bedside patient monitor. The main contribution of this paper include: the device has been a highly integrated design incorporating the radio and all associated circuitry on a single PCB; a new non-invasive and cuff-less measurement of blood pressure using pulse wave transit time has been designed and validated. The device stores data locally on microSD flash and /or transmits via Bluetooth and/or Zigbee. We have developed a bandage vest which embeds reusable electrodes for data acquisition as well as a desktop and mobile application for real-time data telemetry. We have evaluated the performance of the device in capturing and recording ambulatory data and found the device easy to use and with high precision.

Keywords: pulse transit time (PTT), blood pressure, R peak detection, wearable system.

1 Introduction

Nowadays, there are more and more people have the problem of chronic disease. And in order to effectively diagnose these diseases usually need to continuously monitor the patients' physiological information such as electrocardiograph, body temperature and respiration rate for a long term. The development of the Body Sensor Network (BSN) technology makes it possible to continuously monitor physiological parameters such as electrocardiograph (ECG), respiratory rate, electroencephalograph (EEG), blood oxygenation (SpO2), body temperature, etc for a long time. Convenient monitoring of all the aforementioned parameters raises the strong requirement of light-weight, small-size, low-power and multi physiological parameter monitoring systems [1]. However, the advancement of precision micro power amplifiers, microcontrollers, and MEMs technologies have enabled the development of very small, low power, wireless health monitoring technologies [2]. There have quite a few of groups worked on this problem. However, most devices developed only have one or two sensing modalities or require multiple networked devices attached to the

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*, Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_82,

subject. Our approach has been to develop a highly integrated ambulatory device which reduces cost and improves reliability by reducing the number of components, and the wearable technology is our focus in our research.

Compared to the above mentioned system, our wearable system has two following unique features:

1) High integration with multi-parameter: Full-featured physiological parameters monitoring equipment solve the adaptability of the sensor and anti-jamming technology, the wearable system technology, and highly integrated design technology.

2) Real-time Blood Pressure monitoring continuously: Based on the pulse wave propagation time sleeveless-belt continuous blood pressure monitoring sensor system, use hydrostatic method for online calibration and correction to overcome the impact of individual differences in the sensor detection accuracy and facilitate the achievement of the 24-hour continuous blood pressure monitoring.

2 The Principles of Physiological Parameters

Wearable system monitoring multiply parameters is shown in Figure 1. The system is divided into core processing module, electrocardiogram acquisition module, respiration acquisition module, blood oxygen acquisition module, temperature acquisition module, motion state acquisition module, and wireless communication module.

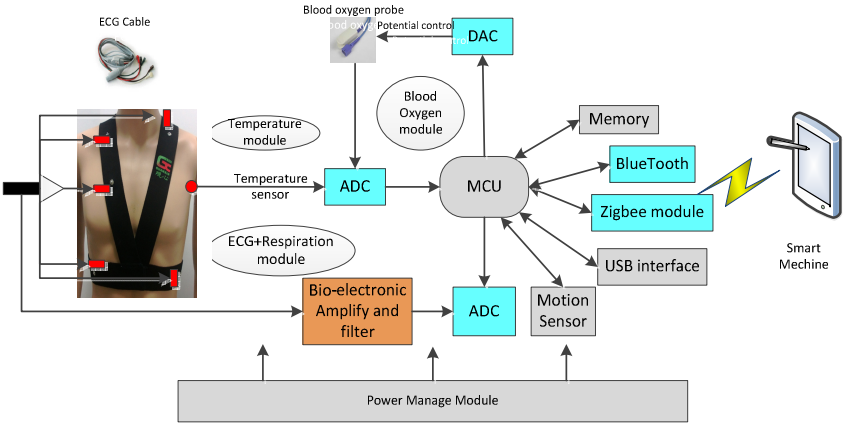


Fig. 1. The measure principle of system

The ECG electrodes are embedded into the inside of bandages, and the electrode is not easy to shift. Measurement in this way can accurately position signal's position of collection point, as well as maintain the consistency of signal measurement point. Temperature and oxygen probe are separately connected to the armpit and fingers as an attachment. Motion sensor real-time monitors the state. The wearable system wears next to the skin; wireless signal is absorbed by the body easily. To avoid the situation,

high frequency circuit matching near field impedance is designed. The modular design in the hardware structure makes each unit powered off completely. And the software system adopt polling mode of each sub-unit low duty cycle. The wearable system uses high input impedance amplifier circuit, correlation detection method and other methods to achieve weak signal detection; a variety of methods is adopted to improve the signal-to-noise ratio. Multiply sensors are integrated in the same unit, the small size of the sensor units and modules reduce the volume of the monitoring system.

3 Hardware Design

The main innovative hardware feature is high integration with electrocardiogram acquisition module, respiration acquisition module, blood oxygen acquisition module, temperature acquisition module, motion state acquisition module, and wireless communication module. The overview of system architecture is discussed in this section.

The architecture of the wearable node is depicted in Figure 2, which includes five function modules, i.e. the core processing unit, power, memory, transmission module and parameters acquisition module. Each hardware sub-circuit is isolated; power to the circuit can be turned on or off independently of the reset of the platform. This isolation provides a degree of robustness-in the event of a failure.

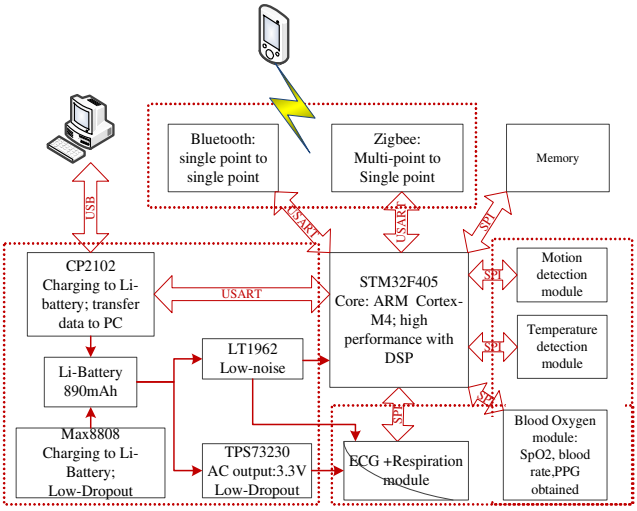


Fig. 2. The architecture of system

For the practical convenience, the package with ECG cable and PPG probe illustrated in Fig. 3 is designed for our wearable system by using plastic material. The system board was embedded in the package. There are an ECG interface, blood oxygen interface, SD interface and power interface in the sidewall.

Each system is equipped with a micro-controller (MCU) which acts as a local control center for collecting and processing data, arbitrating sensor behavior, maintaining communication with the wireless module, and timing events. Since our system needs multi-parameter continuous detection, the entire software algorithm is embedded in the MCU.

Real-time ECG signal is vulnerable to electromagnetic interference, then an EMI filter is needed, an amplifier to amplify the weak signal. Additionally, the right leg drive (RLD) block is used as a means to counter the common-mode interference in an ECG system as a result of power lines and other sources. Respiratory impedance is proportional to lung volume. Based on the collected ECG data, the system tests, and converses, calculates the changing state of lung volume, and achieves the DC signal varying with respiratory impedance, finally result the respiratory rate.

The measurement of blood oxygen is based on the different absorptivity to red and infrared. A single photodiode in response to the red and infrared light receives light. Transimpedance amplifier generates the voltage proportional to the received light intensity. In order to reduce interference with each other, a time-multiplexed manner is usually applied in Red and infrared LED.

Wireless communication module consists of two parts: Bluetooth and Zigbee. The Bluetooth module realizes reliable transmission from one point to another point and Zigbee makes sending and receiving from multi-point to single-point.

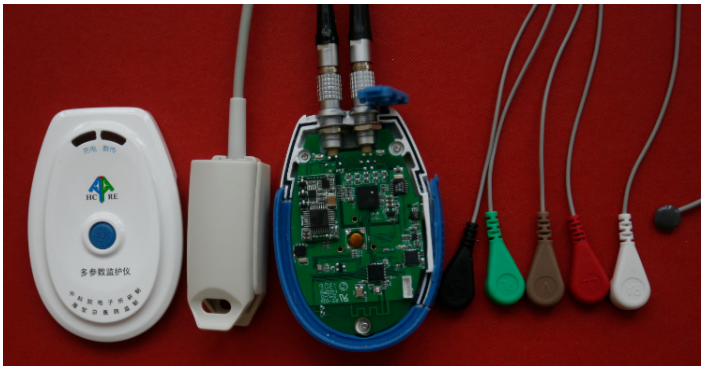


Fig. 3. The package with electrocardiogram cable and photoplethysmography probe

4 Software Design and Data Processing Algorithm

The software design is divided into sensor layer, sampling layer, core processing layer, transport layer and application layer. First, the original signal collected by sensor layer is routed to sampling layer to obtain digital stream. Based on the digital signal, core processing layer handles the data with the RR interval algorithm, respiration algorithm and blood pressure algorithm to obtain indirectly physiological parameters, i.e. Heart rate, respiration rate and blood pressure. Transport layer packets the original data and indirectly physiological information, then transfers the packets to the host computer via wireless module. Application layer in the host

computer displays. This sub-section is organized by five parts to show the whole software algorithm.

4.1 Improved Pan-Tompkins Algorithm

The position of R peak and heart rate are key physiological indicator, so as to the real-time detection, Pan-Tompkins algorithm is adopted [3]. The traditional Pan-Tompkins algorithm consists of low-pass filter, high-pass filter, derivative, square and average value in range of sliding window. There are three differences between the traditional and improved algorithm in Fig.4. One is that we use absolute instead of square, another is we use two single-order low-pass filter instead of a two-order low-pass filter to eliminate the effect of finite word length. The last one is the length of sliding window, which has close relationship with QRS width. And in principle, the length of sliding window must vary in the range between 80ms and 100ms, according to the standard width of QRS complex.

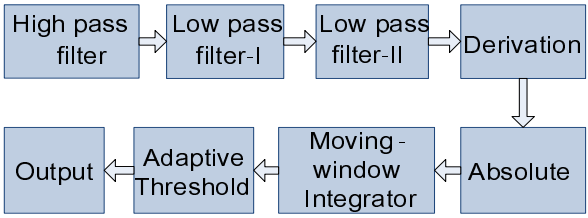


Fig. 4. Pan-Tompkins algorithm

4.2 Respiration Algorithm

The respiratory algorithm will be shown on the following process in Fig.5:

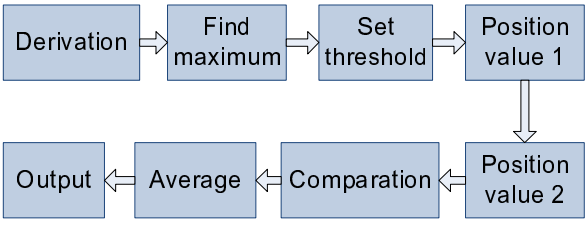


Fig. 5. Respiration Algorithm

First, store respiration data in 8 seconds, derivation to all the data, obtain array Respiration; Second, find the maximum MAX in Respiration [ ]; Third, set *threshold = 0.7*MAX*; Fourth, increase pulseperiod by 1, find the first position whose value is greater than threshold, note *P1*; Fifth, increase pulseperiod by 1 until finding the second position whose value is greater than threshold, note *P2*; Sixth, determine

whether P_1-P_2 >*300*, if the result is yes, we loop to step 4 another two times or go back to step 5; Seventh, set the *average_pulseperiod* = *pulseperiod/3;* The last procedure is, respiration_rate = 60*sample_rate/ *average_pulseperiod.*

4.3 Blood Pressure Algorithm

The measurement of blood pressure takes advantage of a linear relationship between blood pressure and pulse transmission time (PTT), which is the time interval between the R peak with the PPG feature point in the same heartbeat. There are several feature points in PPG. Experiments show the linear correlation with maximum slope on the onset of PPG is best. The whole algorithm is divided into 3 parts: R peak detection, the maximum slope on the onset of PPG and the synchronization between ECG and PPG. Fig. 6 shows the algorithm of maximum slope on the onset of PPG based on the following process:

Fig. 6. Photoplethysmography algorithm

5 Conclusions

In this paper, we have developed a wearable medical monitoring system aimed at the elderly monitor and chronic disease management. The system combines multi-parameter real-time measurement of vital signs, especially to the real-time blood pressure. The tests have provided a clear indication of the feasibility of the concepts and validity of the solutions adapted by the project.

Acknowledgements. The paper is based on research funded through 863 Program under Grant no. 2013AA041201, 2012AA040504 and 2012AA040506.

References

[1] Allen, J.: Photoplethysmography and its application in clinical physiological measurement. Physiol. Meas. 28, R1–R39 (2007)
[2] Deb, S., Nanda, C., Goswami, D., Mukhopadhyay, J., Chakrabarti: Cuff-less estimation of blood pressure using pulse transit time and pre-ejection period. In: Internal Conference on Convergence Information Technology (2007)
[3] Pan, J., Tompkins, W.J.: A real-time QRS Detection Algorithm. IEEE Transactions on Biomedical Engineering BME-32, 230–236 (1985)

On Mobility-Aware Dual Pointer Forwarding Handoff Scheme in Cost-Optimized Proxy Mobile IPv6 Networks

Seungsik Son, Sangik Jeong, Jaeyoung Choi, and Jongpil Jeong

College of Information and Communication Engineering, Sungkyunkwan University,
2066 Seobu-ro Jangan-gu, Suwon, Gyeonggi-do, Republic of Korea
{ssson69,sangikjeong}@naver.com, {jychoi1001,jpjeong}@skku.edu

Abstract. In this paper, a mobility-aware Dual Pointer Forwarding scheme (mPF) is applied in Proxy Mobile IPv6 (PMIPv6) networks. When the MN moves, this scheme can reduce the high signalling overhead for intra-handoff/inter-handoff, because the Local Mobility Anchor (LMA) and Mobile Access Gateway (MAG) are connected by pointer chains. In other words, a handoff is aware of low mobility between the previously attached MAG (pMAG) and newly attached MAG (nMAG), and another handoff between the previously attached LMA (pLMA) and newly attached LMA (nLMA) is aware of high mobility. Based on these mobility-aware binding updates, the overhead of the packet delivery can be reduced. Also, the binding update cost and packet delivery cost for route optimization are analysed, based on the mathematical analytic model. Analytical results show that our mPF outperforms the PMIPv6 and the other pointer forwarding schemes, in terms of reducing the total cost of signalling.

Keywords: PMIPv6, Dual Pointer Forwarding, Mobility-aware, mPF.

1 Introduction

In Mobile IPv6 (MIPv6) [1], signaling procedures are required to support the mobility of mobile terminals based on Mobile Node (MN) in MIPv6, and incur a higher signaling overhead on the network. The Internet Engineering Task Force (IETF) proposed Fast handover for MIPv6 (FMIPv6) [2] and Hierarchical Mobile IPv6 (HMIPv6) [3] to eliminate the weaknesses of MIPv6, but the waste of wireless link resources and handover delay problem was not solved.

PMIPv6 [4] proposed to solve the signaling overhead as a problem of mobility support, for MN-based has no different mobility management of the host-based. The only difference is that there is no request regarding mobility of the MN. But, PMIPv6 resolves the domain internal signaling overhead, while the cross-domain is not specified for the handover. The inter domain handover in PMIPv6 is handover between the LMAs. Because LMA, one of the PMIPv6 domains, has the function of Home Agent (HA), Binding Update (BU) sends the HNP from pLMA to nLMA to carry out, and nLMA access to correspondent node (CN) with the Home Network Prefix (HNP) information received from pLMA. However, the binding update occurs

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_83,

when moving between the LMAs, and this is the cause of the high signaling overhead and handover delay.

In this paper, a method is provided for improving the performance via a dual pointer transfer technique unlike [10], for recognizing the movement by the PMIPv6 network environment. In the case of inter-domain movement in PMIPv6, a binding update is performed among pLMA, nLMA, and CN (LMA). That is, binding updates occur in which HNP is passed to nLMA in pLMA, and nLMA connects to CN (LMA). However, these binding update techniques will have high signal overhead and delay of handover. To solve those problems, the pointer chain is connected, when handover between pLMA and nLMA occurs.

When MN accesses nLMA, nLMA requests the address of HNP and pLMA of pLMA, and a pointer chain is formed between pLMA and nLMA, by transferring the data packet through the LBU from pLMA to nLMA. This enhanced technique can reduce the signal overhead and delay of handover, when an MN moves between domains. Furthermore, the mobile-aware pointer forwarding technique recognizes that mobility adaptive pointer forwarding scheme (mPF) is applied. Mobility aware reduces the negative factor of pointer forwarding in the packet transfer process of MN. In mPF, the slow or fast moving speed of the MN is recognized. When the moving speed is faster than a pre-defined stay time, the length of chain extends to K, while the negative factor of pointer forwarding is reduced, by notifying of CN (LMA) from MN, when the speed of MN is slow, and the packet is transferred. In addition, a mathematical analytic model is used to calculate the updating of binding and transfer of the packet about the pointer forwarding scheme, to recognize the movement in the PMIPv6 network environment, and to evaluate the performance of the proposed method. The performance shows it is better, as compared with the conventional method, in terms of overall cost.

This paper is configured as follows; Chapter 2 explains about related works, and chapter 3 explains the dual pointer forwarding scheme considering mobility. Chapter 4 explains the analytic model and numerical result, and the conclusion is in chapter 5.

2 Related Work

The design of the pointer forwarding scheme has been proposed, to reduce the high costs of signaling in mobile network environments. A dynamic hierarchical mobility management scheme has been proposed [5]. This method calculated via analytical models the optimal length, derived from the chain generated in the home registration. However, based on MIPv6 (Mobile IPv4), the proposed structure does not consider the impact of path optimization. In addition, by assuming the inter-packet arrival process, according to the exponential distribution, we propose a cost analysis model. However, the inter-packet arrival process does not follow an exponential distribution.

A pointer forwarding scheme is proposed in the MIPv6 environment [6]. The pointer forwarding scheme provides service for a single mobility domain, and introduces a new PFMA (Pointer Forwarding Mobility Agent) service. When the MN enters a new mobility domain, it sends a BU message to the previous PFMA.

Therefore, it would be possible to reduce the number of binding updates. The proposed structure in MIPv6 is evaluated by using a simple analytical model. However, it shall not be calculated for the optimal forwarding pointer chain length. In addition, if mobility domain includes access router, PFMA and pointer forwarding occur frequently. In this case, the pointer-forwarding performance is certainly reduced. The pointer forwarding technique was applied in HMIPv6 networks [7]. In this way, the pointer chain is set up between the access routers, instead of the Mobility Anchor Point (MAP). Therefore, this approach does not reduce the binding update message, by handoff between the MAP. In this way, the inter-MAP handoff can't reduce the binding update message to the HA. As mentioned before, handoff between MAP does not happen, if the domain size is enough. Therefore, the pointer forwarding scheme brings limited improvement in the MAP. Moreover, it does not evaluate the impact of route optimization, and the optimal length of the pointer chain. The pointer forwarding mechanism can be used to enhance the performance of some network-based mobility management protocols, i.e., PMIPv6, proposed [10], but it didn't consider the inter-domain mobility.

3 Mobility-Aware Dual Pointer Forwarding Handoff

The handover between domains is not defined in PMIPv6, but the movement between domains is frequent, according to the development of transportation, and diversification into mobile devices. The mobility of Inter-domain causes high handover delay, packet loss and signaling overhead, according to the fast-moving of the MN. For this reason, two methods of a pointer forwarding scheme are proposed, as in Fig. 1. First, when MN moves inter domain, make a pointer forwarding chain between the domains (LMA).

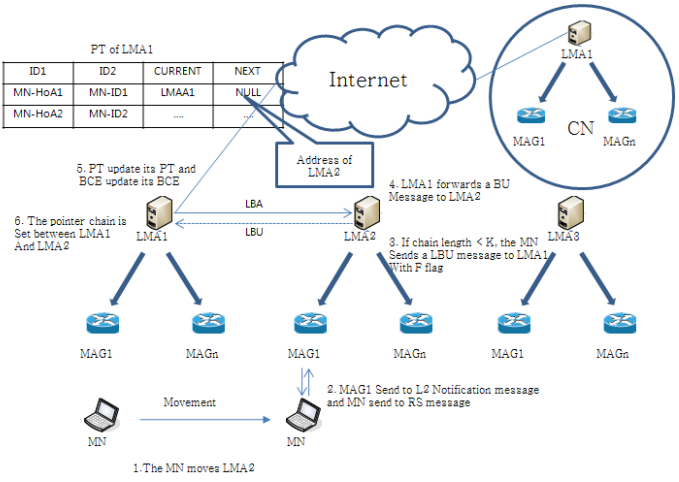


Fig. 1. Inter-domain Binding Update Procedure

If mPF is applied for the PMIPv6 network environment, between the CN (another LMA) and LMA, and LMA and MAG, BU signaling overhead and tunneling overhead can be reduced. Tunneling for PBU between the MAG and the LMA is required; but with the signaling for that, there is a possibility of causing overhead within the LMA. This is the motivation of the binding update considering mobility. MN can be classified into two types in MBU. The classification can be done by measuring the time interval between the router advertisement (RA) messages received from the other MAG. When T1 is the waiting time for the RA message from MAG and moving, and waiting, T2 is the RA from another MAG to move and the waiting time, the MN subnet residence time (t2-t1) can be explained. MN enters a new subnet of the MAG, and it is possible to estimate the residence time in the new MAG subnet, using the residence time of the subnets that are previously measured. Evaluation in the exponentially weighted moving average (EWMA) technique to mitigate the effects of a change in the measured residence time can be utilized. After that, the MN compares the expected residence time to a pre-defined threshold δ. If the expected value is smaller than δ, the MN is considered to be the fastest; if not, it is considered to be slow. PLCoA is the percentage of direct notifications to CN from the MAG. Therefore, while the MN is slow, and is then notified to the CN, the fast MN sends a BU to the LMA. Therefore, the mPF scheme includes the dual pointer forwarding mentioned earlier, and mobility awareness through direct notification between the CN and the MAG, and the MAG and LMA, and the tunnel between the LMA and the CN signaling overhead and packet transmission overhead are reduced. The numerical analysis for the calculation of PLCoA is in Chapter Ⅳ. Thus the fast MN in the mPF scheme using LMA sends BU in order to reduce the binding update traffic and at the same time, the slow MN can reduce the unnecessary LMA processing cost of packet forwarding procedure. Fig. 2 describes the procedure of packet transmission in the mPF scheme.

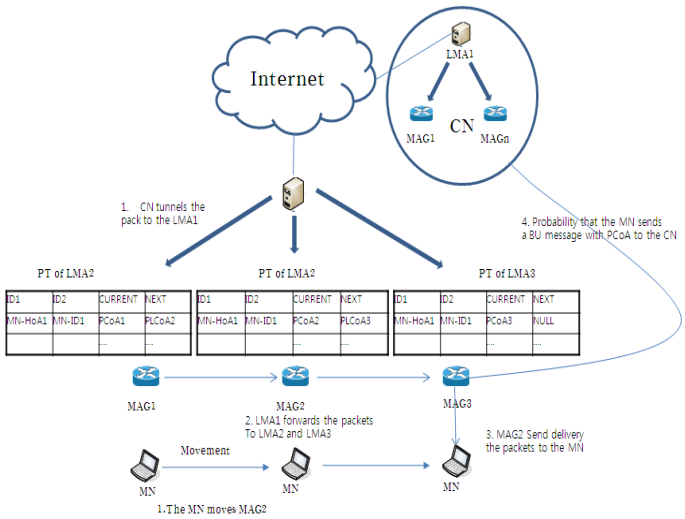


Fig. 2. Packet Transmission Procedure

4 Performance Analysis

In this chapter, we consider the binding update and packet transmission through improved analytical models, to quantify the total cost. The total cost is analyzed between the domains, and within the domain, by each scheme of PMIPv6, PF, mPF. And, we have modeled the following notation for analysis [11].

- $E(L_S)$: Average session length (number of packets)
- $E(N_D)$: The arrival time inter average cross-domain session
- $E(N_C)$: The arrival time inter average cell cross-session
- B_{F-LMA} : unit pointer installation costs(between LMAs)
- B_{F-MAG} : unit pointer installation costs (between MAGs)
- B_{LMA} : PBU cost per unit from MAG to LMA
- P_{F-LMA} : The cost of from previous LMA to next LMA
- P_{F-MAG} : The cost of from previous MAG to next MAG
- $C_{LMA-LMA}$: The number of hops between LMAs
- $C_{MAG-LMA}$: The number of hops between LMA and MAG
- $C_{MAG-MAG}$: The number of hops between MAGs
- w : The total number of data session packet before routing optimization
- PD^{PMIPv} : The packet delivery costs of from CN to LMA
- P_{LCoA} : The rate of sending message from MAG to CN (MAG)

In mPF, it is possible to receive PLCoA from CN in mPF. It is possible to calculate the lifetime T_{BU} that all the PcoA do binding updates across to MAG, considering the MN move slow. The cross ratio (μc) of the MN's MAG should be considered, to calculate the additional cost of BU. Then the average number of crosses to the subnet during T_{BU} is $\mu c T_{BU}$. Therefore the cost of BU in the mPF scheme is as follows.

$$C_{BU}^{mPF-Intra} = C_{BU}^{PF-Intra} + P_{LCoA} \bullet \mu c T_{BU} \bullet B_{F-MAG}$$

If the mPF scheme is applied to the pointer forwarding scheme, then the cost path of pointer forwarding is that MAG notifies direct to CN (another LMA), according to the packet time received from LMA. When MAG informs PCoA to CN, the CN receiving PCoA does packet delivery to MAG. It's the same as PMIPv6. Therefore, when MBU is applied to the packet delivery, the cost of PF is as follows.

$$C_{PD}^{mPF-Intra} = E(L_s) \bullet (P_{LCoA} \bullet PD^{PMIPv6} \bullet C_{MAG-MAG})$$

In the mPF environment, the difference of inter domains and within domain is the same as the difference of pointer forwarding scheme. mPF is the inter domain PF between LMA. In the PMIPv6 environment, the cost of mPF BU between domains is shown below.

$$C_{BU}^{mPF-Inter} = C_{BU}^{PF-Inter} + P_{LCoA} \bullet \mu cT_{BU} \bullet B_{F-LMA}$$

In the mPF environment, the cost of packet delivery is the same between domains and within domain. The packet delivery cost of mPF between domains is as follows.

$$C_{PD}^{mPF-Inter} = C_{PD}^{mPF-Intra}$$

Table 1. The parameter values for numerical analysis

Parameter	Value	Parameter	Value
B_{F-LMA}	2	$C_{LMA-LMA}$	8
B_{F-MAG}	1	$C_{MAG-LMA}$	2
B_{LMA}	4	$C_{MAG-MAG}$	1
P_{F-LMA}	3	PD^{PMIP}	15
P_{F-MAG}	1.5	$T_{BU}(S)$	180

In Table 1, these values applied to process costs and the number of hops for MAG and LMA [12]. SMR (session-to-mobility) = $\lambda s / \mu c$ is defined for mobility effect analysis. λs is the session arrival rate, and μc is the subnet cross ratio. Similarly, if the exponential distribution is set to the session arrival rate, and the subnet run residence time at [13] is $1/\lambda s$ and $1/\mu c$, the average number $E(Nc)$ to cross a number of MAG / LMA per session is $\mu c / \lambda s$. Specifically, the μ_D is an approximation of $\mu c / \sqrt{n}$, and n is the number of MAG within the domain [14]. Therefore, when described as an exponential distribution, the average residence time of the LMA cross ratio of the MAG domain per session, $1/\mu_D$ and $E(N_D)$ is the same as $E(N_C)/\sqrt{n}$. In the above analysis, N is 49 shall be determined by [11]. The subnet residence time t_C is a random variable. Then, P_{LCoA} can be obtained as follows.

$P_{LCoA} = \Pr(tc > \delta) = 1 - \Pr(tc \le \delta)\sigma^{-\lambda c \delta}$, δ is a pre-specified threshold.

The total cost due to SMR is variable. Fig. 3 shows the effect of the total cost of PMIPv6, PF, and mPF, due to change in the SMR.

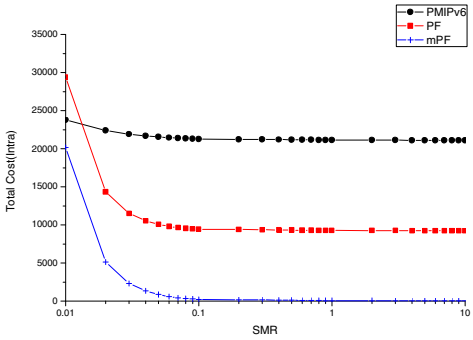


Fig. 3. Impact of the total cost of SMR within domain

It is time to analyze the effect on the length of the session between the domains and within domain, and in cases of the size of SMR being small and big (SMR = 0.1, SMR = 10). Fig. 4 shows the total cost of the session in the case of the value of SMR 0.1.

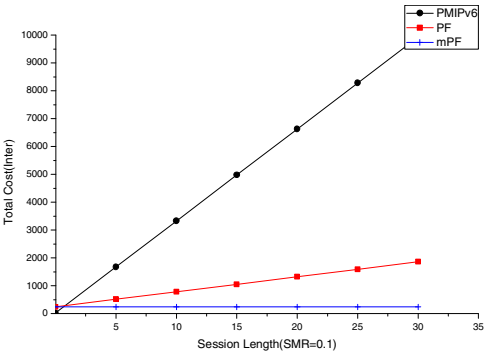


Fig. 4. The effect of session between session (SMR=0.1)

The size of LMA is n (the number of MAG in LMA). Two points of view affect the total cost. The first, LMA within the MAG is the $(E(N_C)\sqrt{n}$ cross. MN and MAG, and is proportional to the number of hops between the LMA and the MAG and the BU costs of the LMA, depending on the size of the LMA. The hops $\log_\beta(n)$ between the MN and LMA are the maximum number in the hierarchical structure. Therefore the BU cost in LMA is proportional to $\log_\beta(n)$.

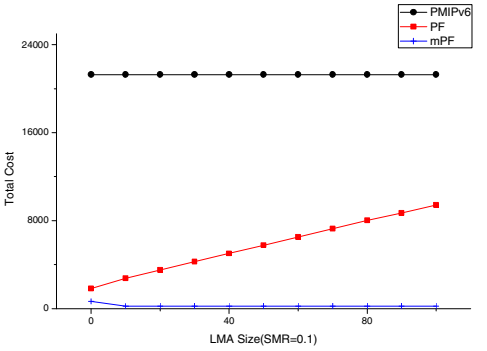


Fig. 5. The effect due to LMA size (SMR=0.1)

Fig. 5 shows the SMR value of 0.1, when the impact on the domain (LMA) in the MAG can. In the case of mPF, the total cost is reduced relatively due to the size of the domain increases. So, mPF is more effective when the domain size is large. In the PMIPv6, it does not affect the size of the domain (LMA). But the relatively total cost of PMIPv6 is significantly large, compared to PF and mPF. Therefore, the difference

is obvious between PMIPv6 PF and mPF. The BU cost is proportional to the size of the domain (LMA).

5 Conclusion

This paper proposes BU and PF methods that consider the mobility in PMIPv6 networks. The proposed dual pointer forwarding scheme can decrease the cost of BU, and solve the high signaling overhead problem generated in LMA. The dual pointer forwarding scheme is proposed with two methods in this paper. The first is the pointer chain connection method between domains (LMA), which restricts movement between domains in PMIPv6. The second is proposed to reduce the high signaling overhead and packet delivery delays within domain that are generated by the pointer chain connection between MAGs within domain, when MNs move within the domain (LMA). The PF scheme between LMA reduces high signaling overhead, due to when the MNs move between domains. A mobility adaptive mPF scheme is used, where the length of pointer chain becomes long, or packet delivery delays occur with simultaneous signaling overhead between LMA and MAG. When the delay of response from LMA reaches a scheduled time, the packet overhead that is delivered from MN is decreased, due to registering PLCoA to MAG, where CN belongs to. Numerical analytic result is presented in this paper, after analyzing the effect of session and domain (LMA) size with between domains, and within domain, respectively. In addition, the effect from the optimal pointer chain length and SMR is analyzed. Finally, the proposed mPF scheme shows that it is excellent in performance, and reduces the total cost from the side SMR environment, by comparing mPF to PMIPv6 and PF.

Acknowledgment. This research was supported by Next-Generation Information Computing Development Program (No.2010-0020737) and Basic Science Research Program (NRF-2010-0024695) through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning. Also, this research was supported by the Ministry of Trade, Industry and Energy (MOTIE), KOREA, and Korea National Industrial Convergence Center (KNICC) through the Special Education program for Industrial Convergence. Corresponding author: Jongpil Jeong.

References

1. Johnson, D., Perkins, C., Arkko, J.: Mobility Support in IPv6. RFC 3775 (June 2004)
2. Koodli, R.: Fast Handovers for Mobile IPv6. IETF RFC 4068 (July 2005)
3. Soliman, H., Castelluccia, C., ElMalki, K., Bellier, L.: Hierarchical Moble IPv6 Mobility Manangement (HMIPv6). IETF RFC 5380 (October 2008)
4. Gundavelli, S.: Proxy Mobile IPv6. IETF RFC 5213 (August 2008)
5. Ma, W., Fang, Y.: Dynamic hierarchical mobility management strategy for mobile IP networks. IEEE Journal on Selected Areas in Communications 22(4), 664–676 (2004)

6. Chu, C., Weng, C.: Pointer forwarding MIPv6 mobility management. In: IEEE GLOBECOM 2002 (December 2002)
7. Yi, M.-K., Hwang, C.-S.: A pointer forwarding for minimizing signing cost in hierarchical mobile IPv6 networks. In: Yang, L.T., Guo, M., Gao, G.R., Jha, N.K. (eds.) EUC 2004. LNCS, vol. 3207, pp. 333–345. Springer, Heidelberg (2004)
8. Chen, I., Gu, B.: Quantitative analysis of a hybrid replication with forwarding strategy for efficient and uniform location management in mobile wireless networks. IEEE Transactions on Mobile Computing 2(1), 3–15 (2003)
9. Ma, W., Fang, Y.: Two-level pointer forwarding strategy for location management in PCS networks. IEEE Transactions on Mobile Computing 1(1), 32–45 (2003)
10. Yan, Z., Lee, J.-H.: State-Aware Pointer Forwarding Scheme With Fast Handover Support in a PMIPv6 Domain. IEEE Systems Journal 7, 92–101 (2013)
11. Zhang, X., Castellanos, J., Campbell, A.: P-MIP: paging extensions for mobile IP. ACM Mobile Networks and Applications 7(2), 127–141 (2002)
12. Xie, J., Akyildiz, I.: A distributed dynamic regional location management scheme for mobile IP. IEEE Transactions on Mobile Computing 1(3), 1069–1078 (2002)
13. Xiao, Y., Pan, Y., Li, J.: Design and analysis of location management for 3G cellular networks. IEEE Transactions on Parallel and Distributed Systems 15(4), 339–349 (2004)
14. Wang, K., Huey, J.: A cost effective distributed location management strategy for wireless networks. ACM Wireless Networks 5(4), 287–297 (1999)

An Adaptive Teaching and Learning System for Efficient Ubiquitous Learning

Kil Hong Joo[1], Nam Hun Park[2,*], and Jin Tak Choi[3]

[1] Dept. of Computer Education, Gyeongin National University of Education, Korea
khjoo@ginue.ac.kr
[2] Dept. of Computer Science, Anyang University, Korea
nmhnpark@anyang.ac.kr
[3] Dept. of Computer Science, Incheon University, Korea
choi@incheon.ac.kr

Abstract. In this paper we present our pedagogical and technological approach for supporting the design of novel situated teaching and learning activities that can be conducted both, outside the school and in the classroom. Education has undergone major changes in recent years, with the development of digital information transfer, storage and communication methods having a significant effect. This development has allowed for access to global communications and the number of resources available to today's students at all levels of schooling. Therefore, ubiquitous learning is a new educational paradigm made possible in part by the affordances of digital information. Ubiquitous learning is characterized by providing intuitive ways for identifying right learning collaborators, right learning contents and right learning services in the right place at the right time. This paper first creates ubiquitous environment, providing function enabling learning to take place anytime and anywhere with any available learning device, for ubiquitous learning according to various properties. Also, in order to improve of proposed ubiquitous system, this paper proposes the scaffolding and mentoring system. If the scaffolding and the mentoring in the ubiquitous learning are provided, studying efficiency would be maximized. Furthermore, the adaptive teaching and learning system with ubiquitous computing may offer great innovation in the delivery of education, allowing for personalization and customization to student needs. The experiments in studying achievements and attitudes of students are performed and show the application possibility of the ubiquitous teaching and learning model.

Keywords: Ubiquitous learning, Ubiquitous environment, U-learning model.

1 Introduction

Ubiquitous computing can be considered as the new hype in the information and communication world. It is normally associated with a large number of small electronic devices (small computers) which have computation and communication

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*, Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_84,

capabilities such as smart mobile phones, smart pad, contactless smart cards, handheld terminals, sensor network nodes, Radio Frequency IDentification (RFIDs) etc. which are being used in our daily life [1]. These small computers are equipped with sensors and actuators, thus allowing them to interact with the living environment. In addition to that, the availability of communication functions enables data exchange within environment and devices. In the advent of this new technology, learning styles has progressed from electronic-learning (m-learning) to mobile-learning (m-learning) and from mobile-learning to ubiquitous-learning (u-learning) [2,3].

Ubiquitous learning, also known as *u-learning* is based on ubiquitous technology [4]. The most significant role of ubiquitous computing technology in u-learning is to construct a ubiquitous learning environment, which enables anyone to learn at anyplace at anytime. Ubiquitous learning or u-learning is a new learning paradigm. It is said to be an expansion of previous learning paradigms as we move from conventional learning to electronic-learning (e-learning) and from e-learning to mobile-learning (m-learning) and now we are shifting to u-learning[3,5,6].

Therefore, this paper proposes new teaching and learning model which enables efficient learning with the ubiquitous computing services in ubiquitous environment. By the proposing model, the school education system becomes familiar with the ubiquitous environment and to be advanced into a national education system. New teaching and learning model proposed in this paper uses the concepts of scaffolding and mentoring are adapted to the ubiquitous environment. It is provided to verify applicability and appropriateness of the new teaching and learning method for the u-learning environment.

2 Ubiquitous Learning (u-learning)

According to [7], "the evolution of ubiquitous computing has been accelerated by the improvement of wireless telecommunications capabilities, open networks, continued increases in computing power, improved battery technology, and the emergence of flexible software architectures". This leads to u-learning that allows individual learning activities embedded in daily life. However, as mentioned by [8], there is no clear definition of u-learning due to rapid changes of the learning environments. Until now, researchers have different views in defining the term "u-learning".

Figure 1 illustrates the classification of four learning environments according to [9] with reference to four dimensions of ubiquitous computing by [7]. From this figure, it was observed that the *Desktop-Computer Assisted Learning* systems provide low mobility and low level of embeddedness. Therefore, the learning environment is fixed. Compared to desktop-computer assisted learning, *Mobile Learning* is basically about increasing learners' capability in order to hold together their learning environment, thus enabling them to learn at anytime and anywhere. In *Pervasive Computing*, learner may obtain information from their learning environment via the communication between the embedded devices and environment. However, this makes the availability of pervasive learning are highly localized and limited. These limitations of

pervasive learning have been overcome by *Ubiquitous Learning* through the integration of high mobility into the learning environment. The communication between devices and the embedded computers in the environment allows learner to learn while they are moving, hence, attaching them to their learning environment. It is obviously shows that the level of embeddedness and mobility of devices do have a significant impact on the learning environment.

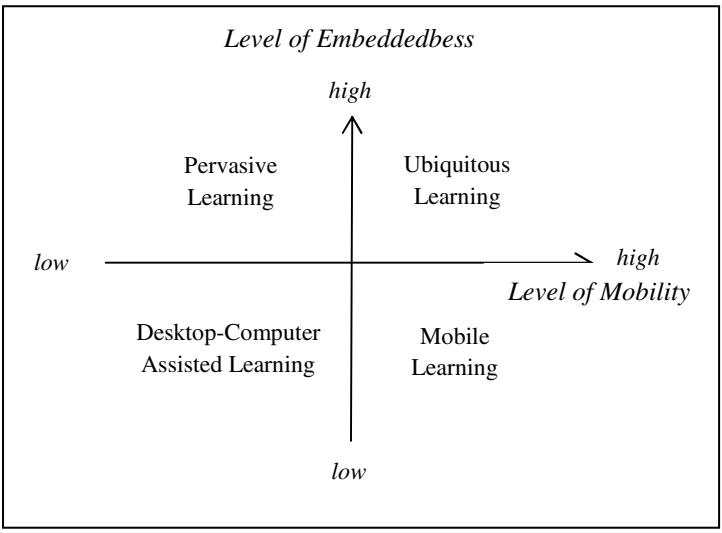


Fig. 1. Classification of learning environments

3 Adaptive Teaching and Learning Model for u-learning

In order to build an u-learning environment, we develop the LMS(Learning Management System)[10, 11, 12]. All students can set their learning goals according to their levels. When students log on to LMS, their attendances are automatically checked and the teacher knows who has attended. In LMS, students can investigate the contents related to a given task and set the goal according to their task after logging on the system. Figure 2 shows the LMS environment and the content.

In the u-learning environment, all students and teacher are connected via wireless network. Once tasks have been assigned access their tasks from anywhere. When students received a task, they determine whether it would be done as an individual assignment or group work. Then, they study without any limitation of place or time while at the use of the available scaffolding and mentoring. Students can supply and obtain materials via wireless networks at various places such as library, science room, laboratory, classroom and the likes. Figure 3 shows studying activities of students.

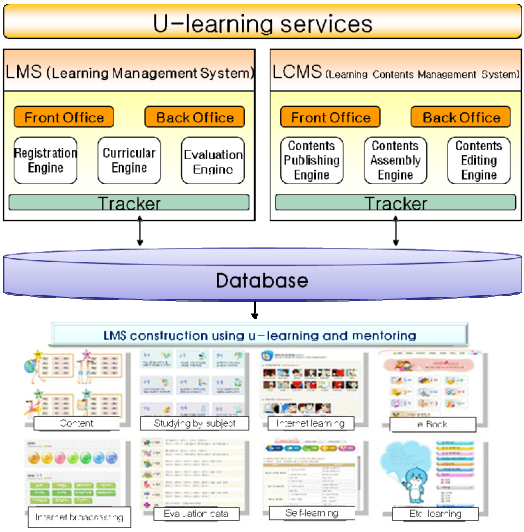


Fig. 2. Classification of learning environments

When students finish the prerequisite learning, they plan learning activities for solving the task. Also, they begin an examination and observation of the task. They receive mentoring to the prerequisite learning by teacher continuously. When students apply for mentoring, the mentee receives requests via smart device or wireless laptop computer. Students can reach the object with the helps from the mentor and they solve the studying problems.

Fig. 3. Studying activities

Figure 4 introduces adaptive teaching and learning model proposed in this paper. The teaching and learning model of u-SM learning proposed in this paper will be described in detail below.

- Learning task selection
- Object selection according to the student's individual learning level
- Assessment and scaffolding
- Mentoring through observation
- Providing scaffolding according to evaluation result

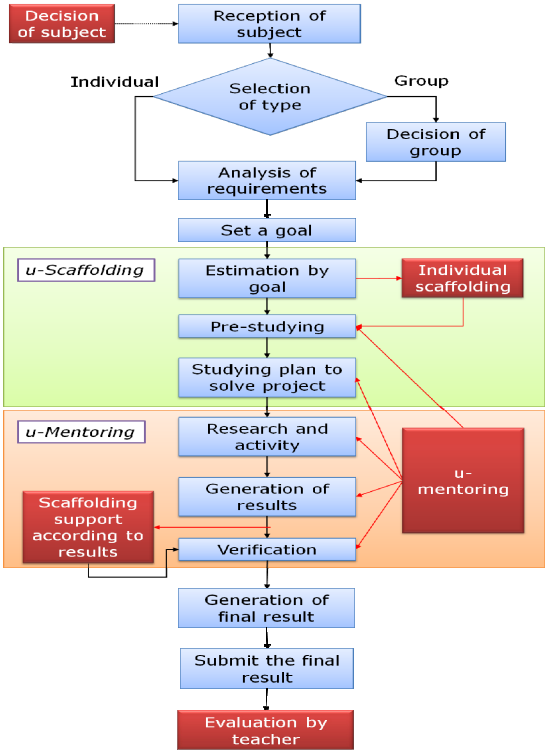


Fig. 4. Adaptive teaching and learning model for u-learning

The u-learning model proposed in this paper allows students to obtain materials anytime, anywhere to solve their tasks. However, if students are not assigned the appropriate learning task, their learning interest will decrease. Therefore, students should first undertake a diagnosis evaluation and then receive tasks according to the results. For this objective, the scaffolding method should be employed. The biggest problem in u-learning is that teacher may not be able to teach while simultaneously determining the extents of the student's understanding as he could in face-to-face education, or student may not be able to understand the educational purpose of

u-learning to utilize the benefits of off-line learning. In order to solve such problems, a mentoring system is adapted in the teaching and learning model proposed in this paper.

4 Experimental Results

The adaptive teaching and learning model proposed in this paper is applied to the fifth graders of 'H' elementary school and examined in connection to each subject across the whole field. The same learning time is provided to both groups of students taking u-learning classes through the adaptive teaching and learning model and taking classes in the conventional learning method. Since there are time differences for each student to be used to u-learning, it is difficult to include the time needed for the on-line scaffolding and mentoring. For this research, after students perform the studying, the results were analyzed by performing a t-test with a significance level of 0.5. Table 1 shows the results of the achievement test.

Table 1. Result of achievement test

subject	Group	Number	Average	Standard deviation	t	P
Korean	Experiment	35	87.5429	10.33652	2.993	.002
	Comparison	35	77.6571	17.05000		
Math	Experiment	35	80.6286	11.43546	1.404	.0825
	Comparison	35	75.8857	16.39184		
Society	Experiment	35	85.2286	9.01008	1.847	.0345
	Comparison	35	80.1714	13.46343		
Science	Experiment	35	89.3714	8.39848	3.248	.001
	Comparison	35	78.5714	17.78679		

Figures 5 and 6 are results of a comparative analysis according to gender of the degree of satisfaction and effectiveness. In both genders, the u-learning level of satisfaction and effectiveness were high. We compared the degree of satisfaction between genders in 10 categories. As a result, there was no difference between genders as described in figure 5. In addition, the degree of satisfaction in all categories was higher than 90 percent. However, the degree of satisfaction in the off-line communication category and private education category were lower than other categories. This is because the degree of satisfaction in communication with parents related to learning was remarkably low in the off-line communication category. It can be posited that students cannot get much help from their parents because of their parents' insufficient understanding of the personal sensor that students use. In the private education aspect, analysis suggests that the parents' lack of understanding does not allow them to readily give up private education, even though u-learning is a more desirable method which can be used as a substitute for private education.

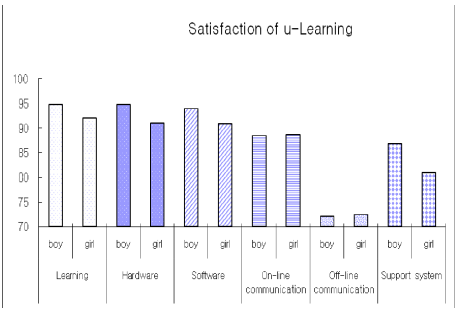


Fig. 5. Satisfaction of u-learning

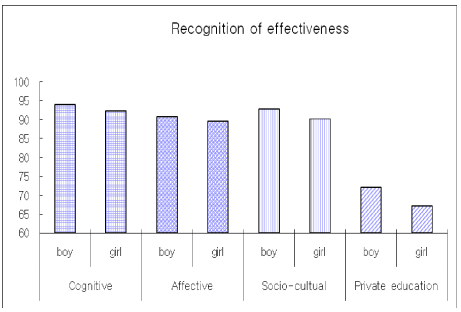


Fig. 6. Recognition of u-learning

5 Conclusion

In this paper, the adaptive teaching and learning model is proposed for effective u-learning. Students are provided with the scaffolding and mentoring and thus can improve their self-directed learning ability, achievement and satisfaction. By providing the adaptive teaching and learning model with the scaffolding and mentoring to lead students into a successful experience in the task-solving process, learning achievement improves and the students' learning attitudes become more positive. Therefore, the adaptive teaching and learning model can be used to solve academic differences in the individual or group task. When the proposed system is applied to the elementary school, students are very positive in their degree of satisfaction/grade/self-directed learning ability. Students are remarkably superior in the categories of self-directed learning ability, including open-minded character, ego, initiative leadership, independence, responsibility, enthusiasm for learning, future-oriented, and problem solving ability. A more effective study will be conducted in the future by adapting this model to ubiquitous learning using a digital text book.

Acknowledgements. This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science Technology(NRF-2011-0025300). In addition, this work was supported by the Incheon National University (International Cooperative) Research Grant in 2013.

References

1. Sakamura, K., Koshizuka, N.: Ubiquitous Computing Technologies for Ubiquitous Learning. In: Proceedings of the 2005 IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE 2005), pp. 11–20 (2005)
2. Yahya, S., Ahmad, E.A., Jalil, K.A.: The definition and characteristics of ubiquitous learning: A discussion. International Journal of Education and Development using Information and Communication Technology (IJEDICT) 6(1), 117–127 (2010)

3. Yang, S.J.H.: Context Aware Ubiquitous Learning Environments for Peer-to-Peer Collaborative Learning. Educational Technology & Society 9(1), 188–201 (2006)
4. Joo, K.-H.: A practical method of a distributed information resources based on a mediator for the u-Learning Environment. Korea Association of Information Education 9(1), 79–86 (2005)
5. Wen, J.R., Chen, C.P.: The strategy of implementing the Internet and cloud computing in teaching. International Journal of Research and Reviews in Computer Science 2(1), 83–87 (2011)
6. Wen, J.-R., Cheng, K.-M., Chen, C.-P., Hsieh, Y.-H.: A study on the application of ubiquitous learning environment to english learning in elementary schools. Universal Journal of Education and General Studies 2(2), 53–65 (2013)
7. Liyytinen, K., Yoo, Y.: Issues and Challenges in Ubiquitous Computing. Communications of the ACM 45(12), 62–62 (2002)
8. Hwang, G.-J., Tasi, C.C., Yang, S.J.H.: Criteria, Strategies and Research Issues of Context-Aware Ubiquitous Learning. Educational Technology & Society 11(2), 81–91 (2008)
9. Ogata, H., Yano, Y.: Context-Aware Support for Computer-supported Ubiquitous Learning. In: Proceedings of the 2nd IEEE International Workshop on Wireless and Mobile Technologies in Education, pp. 27–34 (2004)
10. Deur, P., Murray-Harvey, R.: The inquiry nature of primary schools and student's self-directed learning knowledge. International Education Journal, ERC 2004 Special Issue 5(5), 166–177 (2005)
11. Yoon, D.-H., Joo, J.-H.: Study on Basic Technology for Implementing Ubiquitous Learning system. Industrial Science Research 23(2), 243–252 (2006)
12. Klopfer, E., Yoon, S., Perry, J.: Using Palm Technology in participatory Simulations of Complex Systems: A New Take on Ubiquitous and Accessible Mobile Computing. Journal of Science Education and Technology 14(3), 285–297 (2005)

Enhanced Indirect-Broadcasting Synchronization Protocol for Wireless Sensor Networks

Shi-Kyu Bae

Computer Eng. Dept., DongYang Univ., Korea
skbae@dyu.ac.kr

Abstract. Time synchronization in Wireless Sensor Networks (WSN) is critical to many WSN applications. A new time synchronization protocol for WSN called Indirect-Broadcast Synchronization (IBS) [1] has been already proposed. As the protocol operates in cluster tree topology, network lifetime may be shortened mainly by cluster head node[s] (called as a Leader Node in IBS), which consumes more power than cluster member (i.e. non-leader) nodes usually. In this paper, we propose enhanced version of IBS (called EIBS) which saves overall energy and prolongs network lifetime by re-constructing partial cluster tree locally. Compared with other tree construction approaches, this tree reconstruction algorithm is not only simpler, but also more efficient in the light of overall power consumption and network lifetime.

Keywords: Wireless Sensor Networks, Time Synchronization, Energy-efficient, Network Lifetime, Cluster, Tree.

1 Introduction

Time synchronization in WSN is also necessary because time plays a crucial role for many WSN applications such as data fusion, assembly of distributed observations, duty cycling, transmission scheduling, localization, security, tracking etc. Time synchronization scheme for WSN should meet following requirements; 1) limited resources and cost such as computing power and communication capability, 2) scalability for working well with any number of nodes. Energy consumption is most important among requirements.

A time synchronization scheme for WSN called Indirect-Broadcast Synchronization (IBS) has been already proposed [1], which is energy efficient by reducing communication overhead and has good performance in synchronization accuracy and scalability. The scheme synchronizes a multi-hop network cluster by cluster with the global time from the referenced node.

A network with hierarchical structure is better than for scalability than flat structure [2]. Cluster-based structure has been used to improve the routing efficiency such as scalability, load, robustness, and energy consumption, in a dynamic network [3]. Many clustering schemes with different characteristics and design goals have been developed till now.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_85,

In this paper, we propose enhanced version of IBS (called EIBS) which saves whole energy and prolongs network lifetime by re-constructing partial cluster tree locally. This scheme provides the capability of tree reconstruction algorithm which is not only simpler, but also more efficient in the light of overall power consumption and network lifetime, compared with other cluster construction approaches.

The remainder of this paper is organized as follows. After surveying existing clustering schemes in section 2, we propose cluster-based time synchronization scheme with the capability of partial cluster reconstructing in section 3. And then analysis of proposed algorithm has been performed in section 4. This paper ends with some concluding remarks in Section 5.

2 Related Works

There have been proposed some synchronization schemes. Synchronization approach is classified into 3 categories in terms of message direction; Sender-receiver, Receiver-receiver [4] and Receiver-only approach. Among three approaches, the sender-receiver type indicates that one node send a message while the other receive it.

A network with hierarchical structure is better than for scalability than flat structure. Cluster-based structure has an advantage of scalability than flat structure also. So, it has been used to improve the routing efficiency in a dynamic network [2][5]. Authors in [2] proposed a distributed algorithm that reconstructs the aggregation tree from the initial aggregation tree excluding the faulty sensor node. In the algorithm of [5], a cluster tree is constructed by the sink node which selects cluster heads using power and location information of all nodes. These existing cluster tree construction methods perform full reconstruction for the entire network whenever topology changes by at most one node.

Similar to routing techniques, several clustering-based synchronization schemes have been developed recently [1],[6-8]. Clapping and Broadcasting Synchronization (CBS) [6] synchronizes Ordinary nodes in each cluster by using Broadcasting and Clapping node. Authors in [8] presented a cluster-based synchronization scheme in heterogeneous WSN, where H-sensors which have a larger transmission range (power) than L-sensors play a role of cluster header.

3 Proposed Tree Reconstruction Algorithm

3.1 Network Model in IBS Protocol

IBS[1] synchronizes whole network on a cluster tree, similar to TPSN [9]. So, IBS has 2 phases; *Setup* and *Synchronization* phase [1]. Setup phase constructs cluster tree using various clustering algorithms. Many clustering algorithms with different performance have been proposed. As developing of an efficient clustering algorithm was not part of of IBS, but a new method of synchronization, a simple clustering algorithm is used, which is similar to tree construction algorithm used in TPSN [9], as shown Fig. 1.

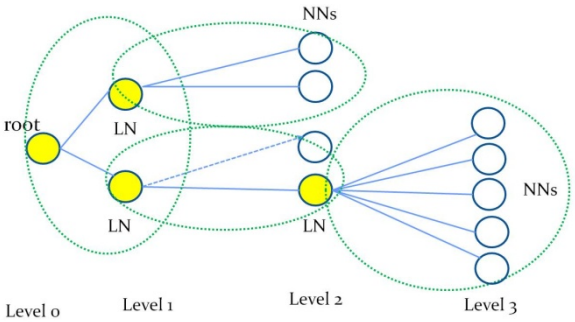


Fig. 1. Cluster tree used in IBS

In synchronization phase, synchronization from the root node together with the next neighbor nodes is performed consequently through the whole network. Nodes in IBS are classified into Root, Assistant, Leader, and Normal node according to their role. And they are simply grouped into *Leader node* (LN) and *Normal node* (NN).

LN plays a main role for synchronization of normal nodes within a cluster. And NN, a member of the cluster, is led by LN and does very simple task relatively. Thus, LNs consume most of power while NNs hardly do (See Fig. 2).

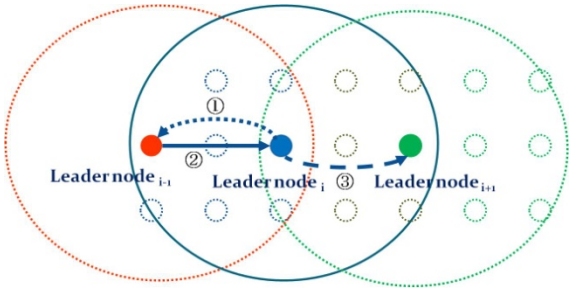


Fig. 2. LN's role in basic operation of IBS

3.2 Cluster Head Change Procedure

When any LN in the tree detects that its residual energy is less than a threshold value, it tries to find another LN which is able to replace itself among NNs in the cluster. The candidate LNs should have the following conditions;

- They have more residual energy than the current LN in order to act as LN for longer time.
- They should reside in the limited and specified area, where is cross-section between the communication area of both the parent and child nodes of the current LN (shown in Fig. 3). It is because that whole tree structure of clusters should not be affected after LN replaced with another node.

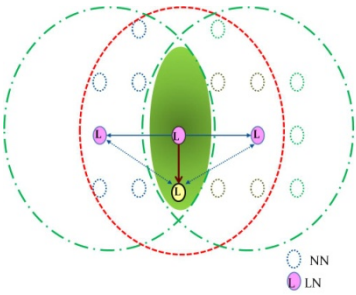


Fig. 3. Area for new LN candidates

The procedure for changing LN is as follows;

(a) The current LN start to find a new LN by broadcasting *"Help"* message with IDs of itself, its parent and child nodes (see Fig 4-(a)).
(b) Receiving *"Help"* message from the current LN, its parent and child nodes (other LN, as well) send broadcast *"Agree"* message with it family's ID (with the child node's ID for a parent node and the parent node's ID for a child node, each). *"Agree"* messages are used for NN to determine if they are new LN candidate[s] or not. That is to say, NNs which can receive all the messages from the current LN, its parent and its child nodes will be a candidate of a new LN (see Fig 4-(b)).
(c) Candidate[s] of a new LN check if its residual energy is more than the current LN's one. Being enough to be a new LN, they ask for a new LN by sending *"Request"* message to the current LN. See Fig 4-(c).
(d) On receiving requests from the candidates for a certain interval, the current LN determines a new LN by choosing the requesting node which has the most energy value. And the current LN broadcasts the notification that the new LN is selected with *"Nominate"* message (see Fig 4-(d)).
(e) New LN notifies that it will begin its term after a specified time with *"Inaugurate"* message (see Fig 5-(a)).
(f) All NNs, except the new selected LN, update information about relationship. There are four cases;
    - Change to new selected LN. If NN could listen to the messages from both the current and next LN, it just changes its LN ID to new LN's ID (see the left figure of Fig 5-(b)).
    - Move to other cluster. If NN cannot hear the new LN for the specified time, after listening to the messages from the current LN that new LN was elected, NN will move to new cluster by joining to other LN than the previous LN.
        - If NN has heard the parent node of the previous LN (say, "Agree" message before) it joins to the cluster of the previous LN's parent (see the right figure of Fig 5-(b)). For example, they are node #1 and node #2 in blue-colored area of the right figure of Fig 5-(b).

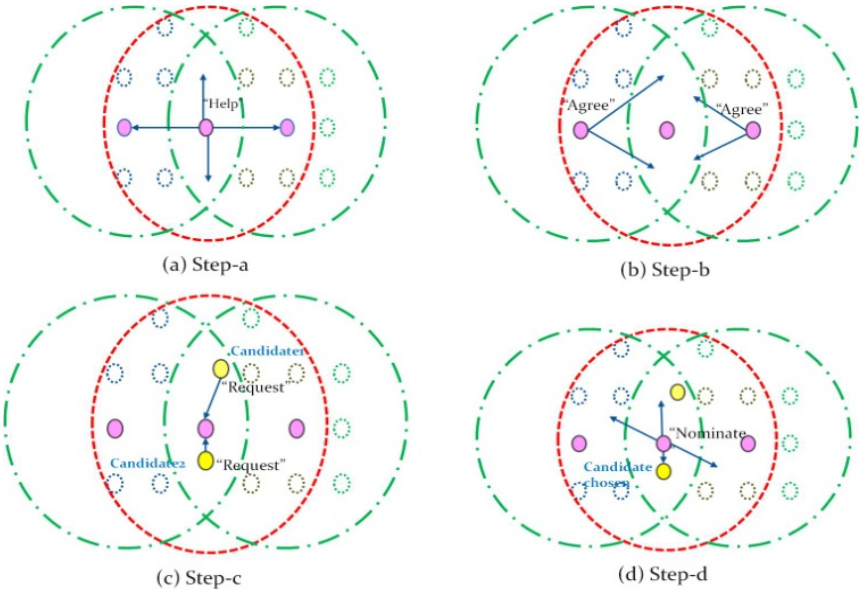


Fig. 4. Procedure for new LN selection

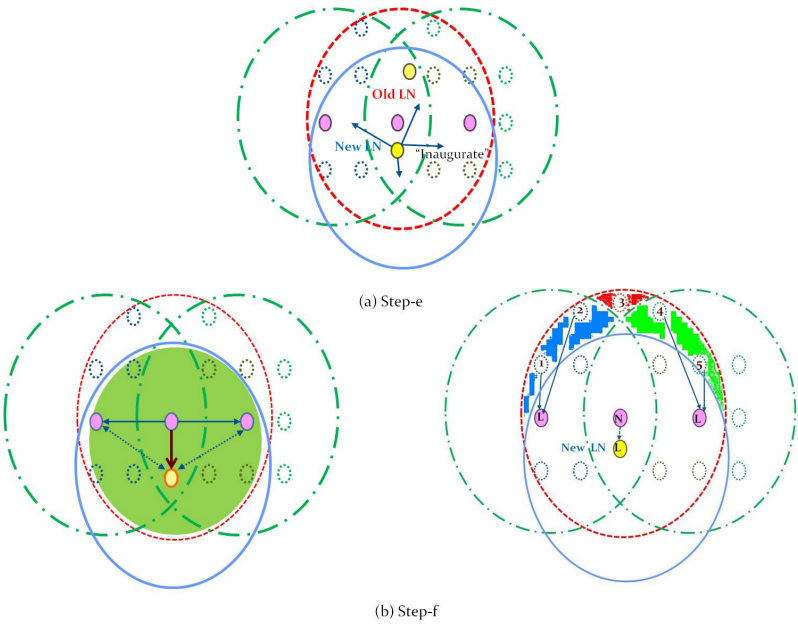


Fig. 5. Rejoining of NNs according to the new LN

- If NN has heard only the child node of the previous LN, not the parent one, it joins to the cluster of the previous LN's child. These examples are node #4 and #5 in green-colored area of the right figure of Fig 5-(b).
- If NN has heard neither the parent nor the child node of the previous LN, it tries to join to other cluster to which the parent or child node of the previous LN belongs. See node #3 in red-colored area of the right figure of Fig 5-(b).

4 Analysis of the Proposed Algorithm

The simple clustering algorithm used in IBS transmits packets as many times as the number of the nodes in the whole network, which is same as the tree constructed in TPSN [9]. In contrary, for partial tree reconstruction by replacing LN in the cluster, at least six messages are required; "Hello", 2 * "Agree", several "Request", "Nominate", and "Inaugurate".

If partial reconstructions at k CH[s] are needed, the total number of required message transmission will be $(6 * k)$.

Partial tree reconstruction in the proposed algorithm would be more efficient than the full tree construction, as long as the following condition is satisfied.

$$(6 * k) < N \tag{1}$$

where k is the number of CH to be changed.

Let define Cluster Head Ratio (CHR) be the ratio of the number of cluster heads to the number of entire nodes.

$$\mathrm{CHR} = \frac{H}{N},\ 0 < CHR < 1 \tag{2}$$

Where H is the number of cluster heads and N is the number of all nodes in a network.

To replace all the CH, Eq. 1 is rewritten by using Eq. 2, as below,

$$(6 * \mathrm{CHR}) < 1 \tag{3}$$

This means that in the case of CHR=0.15 (15 %) , for example, the partial reconstruction even for all CHs would be better than the full reconstruction in the light of energy consumption.

5 Conclusion

A time synchronization protocol for WSN called IBS has been already developed. As the protocol operates in cluster tree topology, network lifetime may be shortened mainly by LNs in IBS, which consumes more power than NNs, usually. In this paper, we propose enhanced version of IBS (called EIBS) which saves overall power consumption and prolongs network lifetime by re-constructing partial cluster tree

locally. Compared with full tree reconstruction approaches, this partial tree reconstruction method is not only simpler, but also more efficient in the light of overall power consumption and network lifetime. The proposed partial tree reconstruction algorithm, we believe, will be useful for other cluster tree-based protocols for WSN.

References

1. Bae, S.: Time Synchronization by Indirect-Broadcasting for Wireless Sensor Networks. In: Park, J.J(J.H.), Jeong, Y.-S., Park, S.O., Chen, H.-C. (eds.) EMC Technology and Service. LNEE, vol. 181, pp. 659–666. Springer, Heidelberg (2012)
2. Sharma, P., Mandal, P.: Reconstruction of Aggregation Tree in spite of Faulty Nodes in Wireless Sensor Networks. In: 6th IEEE International Conference on Wireless Communication and Sensor Networks (WCSN 2010), Allahabad, India (December 2010)
3. Liu, X.: A Survey on Clustering Routing Protocols in Wireless Sensor Networks. Sensors (2012)
4. Sundararaman, B., et al.: Clock Synchronization for Wireless Sensor Networks: A Survey. Ad Hoc Networks (2005)
5. Yen, Y., et al.: Tree-Clustered Data Gathering Protocol (TCDGP) for Wireless Sensor Networks. Journal of the Chinese Institute of Engineers 32(7), 1025–1036 (2009)
6. Qian, X., et al.: Clapping and Broadcasting Synchronization in Wireless Sensor Network. In: 2010 Sixth International Conference on Mobile Ad-hoc and Sensor Networks (2010)
7. Gautam, G., et al.: Time Synchronization Protocols for Wireless Sensor Networks using Clustering. In: IEEE Int. Conf. on Recent Trends in Information Technology (ICRTIT 2011) (2011)
8. Du, X., Guizani, M., Xiao, Y., Chen, H.: Secure and Efficient Time Synchronization in Heterogeneous Sensor Networks. IEEE Transactions on Vehicular Technology 57(4) (July 2008)
9. Ganeriwal, S., et al.: Timing-Synch Protocol for Sensor Networks. In: ACM Sensys, USA (2003)

Multimodal Combination of GPS, WiFi, RFID and Step Count for User Localization

Hung-Long Nguyen, Eric Castelli, Trung-Kien Dao,
Viet-Tung Nguyen, and Thanh-Thuy Pham

Dept. of Pervasive Spaces and Interaction, MICA Institute (HUST-CNRS/UMI 2954-INP Grenoble), Hanoi University of Science and Technology, Hanoi, Vietnam
{hung-long.nguyen,eric.castelli,trung-kien.dao,
viet-tung.nguyen,thanh-thuy.pham}@mica.edu.vn

Abstract. In this paper, a conceptual design of an indoor/outdoor localization system using combination of multiple localization technologies based on the idea of dividing spaces into grid points is presented. In implementation part, we have built the system with GPS, RFID, WiFi positioning and step count techniques, in which the system has been proven to improve accuracy and availability compared to each technology individually.

Keywords: multimodal localization, WiFi-based localization, sensor fusion, indoor localization.

1 Introduction

Many user localization technologies and methods have been proposed for either indoor or outdoor environments. Among the most used technologies are GPS, RFID, others methods based on WiFi, camera, accelerometer, microphones, etc [1]. However, each technology has its own backwards, for example GPS is not good with indoor localization when number of visible satellites is reduced cause of walls blocking; WiFi positioning is only suitable for user localization with low precision because of its accuracy varying from few meters to tens of meters; RFID, a proximity scheme, is limited in a small range since RFID readers could not be installed at every location.

Recently, many researches and designs have been proposed to build a combination of multiple localization technologies system which can provide higher precision results and solve the limitation in each localization technology alone. Pfeifer [2] proposed a design to extract results from localization technologies as useful information in real time, however it lacks ideas and algorithms about how those results should be fused and analysed to produce better results. Few systems already combined some localization technologies, which showed to improve in precision, but these systems depend on specific technologies and lack of availability characteristics: GPS, WiFi, Zigbee [3]; RFID, WiFi, camera [4]; WiFi, step count [5].

In this paper, a general approach to multimodal localization system combining multiple technologies based on the idea of dividing spaces into grid points where

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_86, © Springer-Verlag Berlin Heidelberg 2014

location result will be chosen as the point with highest probability depends on a given precision is proposed. The system does not depend on any specific localization technologies but is built to be an open platform so that multiple heterogeneous localization technologies can be integrated, since more technologies applied means we have more information to improve precision as well as availableness.

2 General Approach

To determine the user location, space is divided into grid points. At each point, the user appearance probability and precision is calculated with information provided from all available localization technologies. Since each localization technology provides results from at different moments, the time past from the last result received from a technology to the current system time also affects the final result. Besides, each application may require a different acceptable precision, so on the basis of the given precision from each application, the location with highest probability is chosen as user localization result:

$$\mathbf{x}_{res} = \mathbf{x} \text{ if } \mathrm{Prob}(\mathbf{x}) = \max\{\mathrm{Prob}(1), \mathrm{Prob}(2), ..., \mathrm{Prob}(n)\} \wedge \mathrm{Pre}(\mathbf{x}) \le Pre^{*}, \quad (1)$$

where $\mathrm{Prob}(\mathbf{x})$ and $\mathrm{Pre}(\mathbf{x})$ are probability and precision at point $\mathbf{x}$, Pre^{*} is the acceptable precision given by the application, n is the number of grid points, and $\mathbf{x}_{res}$ is the localization result.

The probability at point $\mathbf{x}$ is defined as

$$\mathrm{Prob}(\mathbf{x}) = \sum_{i=1}^{k} P_i(\mathbf{x}) R_i \exp(-\lambda_i \Delta t_i), \quad (2)$$

where $P_i(\mathbf{x})$ is probability at point $\mathbf{x}$, R_i is reliability constant (i.e., RFID results is more reliable compared to those of GPS), λ_i is time decay constant of the i^{th} technology, Δt_i is the time difference from the last result current processing time, and k is the number of localization technologies in use. The precision at point $\mathbf{x}$ is defined as

$$\mathrm{Pre}(\mathbf{x}) = \min_{i=1..k}\{\mathrm{Pre}_i(\mathbf{x})\}, \quad (3)$$

where $\mathrm{Pre}_i(\mathbf{x})$ is the precision of the i^{th} technology at point $\mathbf{x}$.

For localization technologies (such as GPS, RFID, camera) providing results consisting of location and precision, $P_i(\mathbf{x})$ can be extracted from normal distribution using the empirical rule:

$$P_i(\mathbf{x}) = \left(1/\sigma_i\sqrt{2\pi}\right)\exp\left(-d_i^2(\mathbf{x})/2\sigma_i^2\right), \quad (4)$$

where $\sigma_i = 3Pre_i$, $d_i(\mathbf{x})$ is the distance from point $\mathbf{x}$ to user location provided by the i^{th} technology. For technologies based on user movement (such as accelerometer, step count), probability information $P_i(\mathbf{x})$ can be determined by combining the last localization result and normal distribution using Eq. (4) however the value of σ_i is replaced by $3LastPre_i$.

If the environment is large where the number of points is big, it is possible to reduce the searching time by first gridifying the space with fewer points to find a rough position, then repeating the same process once or twice with the subspace around this rough position for fine tuning.

The system uses results from multiple technologies as inputs and gives out user location to applications as output. Event-driving approach is used, i.e., each input will provide new localization results to ensure real-time availability. API module contributes as an open platform to extract useful information from inputs and store in the database. Information extraction module plays the part to extract necessary information from database as well as receives precision limitation from applications and provides them to next module. Calculating module does the main processing function and produces user localization result. Depending on applications and their required precisions, results would be different. Afterwards, the outputs including user location, precision and probability are sent to applications and feed back to store in database for next calculating usage.

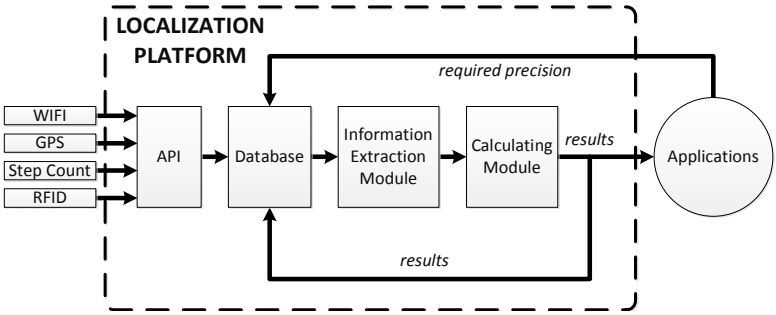


Fig. 1. General system architecture

3 Implementation

Recent studies to monitor location using smartphones have increased due to their advantages and popularity: embedded GPS radio, WiFi functioning and others sensors (accelerometer, orientation, magnetic field, etc.) RFID tags are widely deployed in many applications with the advantages of its small size and variables in contents. We decided to first implement four technologies as inputs in the system: GPS, WiFi, step count from Android smartphones, and RFID positioning by RFID tags.

3.1 WiFi

With conventional receivers, distance to WiFi APs can be estimated from the measured RSSI with help of a RF propagation model which is constructed based on the fact that a radio wave traveling through a certain environment will undergo specific types of signal attenuation. To start off, the empirical model widely used in previous works [6] is considered, with extension to include attenuation due to walls and floors:

$$P = P_0 - 10n\log\left(r / r_0\right) - k_d \sum_{i=1}^{n_w} d_i / \cos\beta_i \,, \tag{5}$$

where P_0 is the known signal power at a reference distance r_0 in dBm; P, the signal power at an unknown distance r; n, the path-loss exponent; n_w, the number of walls and floors in the middle; d_i and β_i, the thickness and angle of arrival to the i^{th} wall/floor, respectively; and k_d, the attenuation factor per wall/floor thickness unit.

In reality, given the RSSI P, the distance r might not be exactly the value calculated from Eq. (5), but is within a range around this value, which is denoted by $\bar{r}$. To be more precise, $\bar{r}$ will be the nominate value of the distance r with highest probability. The distribution of the distance is assumed to follow the normal distribution with median $\bar{r}$:

$$\rho\left(r, P\right) = \Pr\left(r \middle| P\right) = \left(1/\sigma\sqrt{2\pi}\right)\exp\left(-\left(r-\bar{r}\right)^2 / 2\sigma^2\right), \tag{6}$$

where $\sigma = k_\sigma \bar{r}$ is the standard deviation, with k_σ being a constant.

From a tuple of RSSI information received from a WiFi receiving device, the system determines the location with maximal summation of probabilities corresponding to the visible APs, i.e., maximizing $\rho_\Sigma\left(x, y, z\right) = \sum_{i=1..n_{AP}} \rho\left(r_i(x, y, z), P_i\right)$, where n_{AP} is the number of visible Aps; $\rho_\Sigma\left(x, y, z\right)$, the probability that the user is located at position (x, y, z); and $\rho\left(r_i(x, y, z), P_i\right)$, the probability component based on the i^{th} visible AP. The search process is achieved by gridifying the space surrounding the environment into a number of points and calculate ρ_Σ for each of them to find the point that maximizes ρ_Σ.

3.2 GPS, RFID and Step Count

For GPS technology, we use the built-in GPS API from Android platform with ex-traction information including location (latitude, longitude, altitude) and its precision.

The step count is implemented using built-in sensors from Android smartphones. Using derivative applied to the smoothened signal from accelerometer, when the derivative is greater than threshold value, we detect a step [9]. Direction of users is determined by geometry sensor with azimuth, roll, and pitch angle information. After a high precision localization from RFID, step count can make a good localization in a short time.

With characteristics of RFID radio signal, it can be assured that at the moment user's RFID tag is recognized by a reader, the user is within 2 meters around it. The RFID reader position is stored in database and every time a RFID tag is read from a reader, information including user ID and RFID reader ID will be sent to system API. Afterwards user location will be calculated on the basis of the RFID reader location.

4 Experiment Results

Experimental scenario is taken in the 8th floor of an 11-floor building, as shown in Fig. 2. A user holds a smartphone, walks starting from Point 1 straight to Point 2, Point 3, respectively, and then returns to Point 2 and Point 1 at the same path, with the total distance of 160m. In this experiment, 3 RFID readers, 16 WiFi access points are set up for RFID and WiFi technologies. Results are compared and analysed between multimodal system and WiFi alone system. Figure 3 shows user walking path and localization results in two cases using multimodal localization system, compared with the system using WiFi alone. It can be observed that multimodal-based results almost coincide with user walking path while WiFi-based results have big jumps and it is impossible to track position of user with only WiFi-based system.

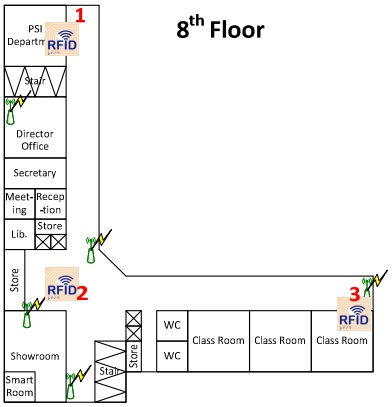


Fig. 2. Testing environment

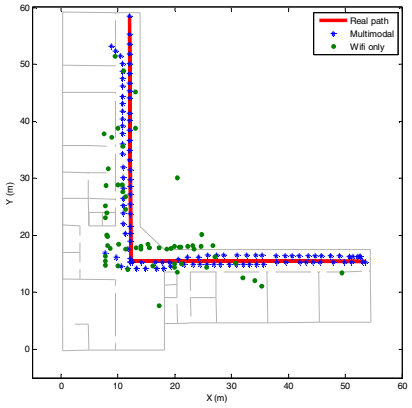


Fig. 3. Localization results

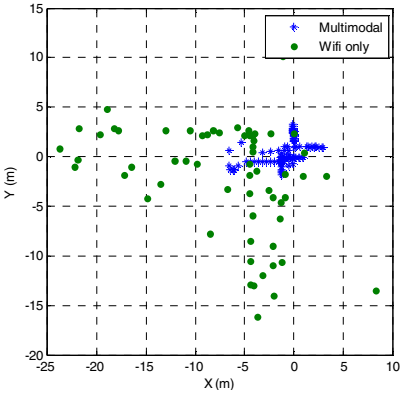


Fig. 4. Localization error distribution

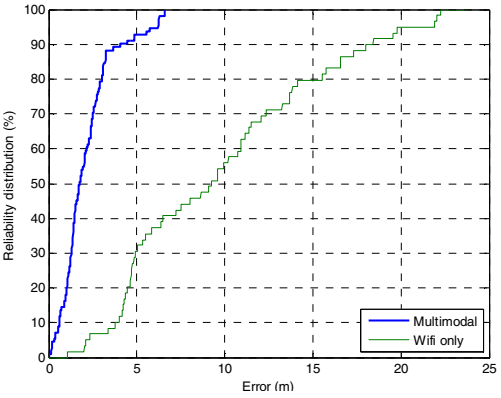


Fig. 5. Localization reliability distribution

Table 1. Localization results

	Average error (m)	Maximal error (m)	Error at reliability of 90% (m)
Multimodal	2.18	6.62	4.07
WiFi only	9.99	23.84	18.44

Figures 4 and 5 present the error distribution and reliability distribution. While multimodal results gather in around 5 meters of error, WiFi results error distribution exceeds 20 meters. It is shown that current implementation for multimodal would be useful for applications with required precision of 5 meters. The error results are summarized in Table 1. Average error of multimodal system is about 4.5 times better, the maximal error is 3.6 times smaller and the error at reliability of 90% is 4.5 times smaller compared to WiFi-based system. It can be seen that the multimodal system do a much better work, and localization results are improved significantly.

5 Conclusion and Future Works

In this paper, an approach to build an indoor/outdoor multimodal system combining multiple localization technologies has been introduced. An implementation with WiFi, GPS, RFID, step count technologies has been successfully built and demonstrated. Experiment results show a promising approach for this system to continue in the future. For future works, the WiFi implementation algorithm will be improved. Step counter with smartphones also needs to be improved to work in pocket situation (currently smartphones need to be held in hand). Second, it is necessary to think of some algorithms to turn on/off smartphone sensors on the basis of activities to save battery. Environment information constraints combining user's history records and Kalman filters could also be used to smoothen the output results. At last, next localization technology based on video processing will be implemented.

Acknowledgment. The authors would like to thank the project "*Visually Impaired People Assistance using Multimodal Technologies*" funded by the VLIR's Own Initiative's Program under grant reference VLIR-UOS ZEIN2012RIP19, the National Project "*Develop an Optimal Path Finding System using Localization Information based on WiFi, Camera, RFID and Application in Aiding Blind People in Unconstrained Environments*" under grant reference B2011-01-052, and the project members for their supports.

References

1. Ferdous, S., Vyas, K., Makedon, F.: A survey on multi-person identification and localization. In: 5th Int. Conf. on Pervasive Technologies Related to Assistive Environments, Crete, Greece, p. 36 (2012)

2. Pfeifer, T.: Redundant positioning architecture. J. of Computer Communications 28(13), 1575–1585 (2005)
3. Yeha, S.C., Hsu, W.H.: Adaptive-weighting schemes for location-based services over heterogeneous wireless networks. In: IEEE Vehicular Technology Conf., Ottawa, Canada (2010)
4. Anne, M., Crowley, J.L., Devin, V., Privat, G.: Localisation intra-bâtiment multi-technologies: RFID, WiFi et vision. In: 2nd French-Speaking Conf. on Mobility and Ubiquity Computing, pp. 29–35 (2008)
5. Martin, E., Vinyals, O., Friedland, G., Bajcsy, R.: Precise indoor localization using smart phones. In: Int. Conf. on Multimedia, Firenze, Italy, pp. 787–790 (2010)
6. Figueiras, J., Frattasi, S.: Mobile Positioning and Tracking: From Conventional to Cooperative Techniques. John Wiley & Sons (2010) ISBN: 978-0470694510
7. Wu, S.S., Wu, H.Y.: The design of an intelligent pedometer using Android. In: Int. Conf. on Innovations in Bio-inspired Computing and Applications, Shenzhen, China, pp. 313–315 (2011)

User Authentication Mechanism Based on Secure Positioning System in RFID Communication

Xinyi Chen, Inshil Doh, and Kijoon Chae*

Dept. of Computer Science and Engineering,
Ewha Womans University, Korea
chloexiny@ewhain.ac.kr, {isdoh1,kjchae}@ewha.ac.kr

Abstract. Radio Frequency IDentification (RFID) has gathered a lot of interest as a good wireless technology in pervasive computing. However, the security challenge should be addressed for secure and reliable communication in RFID environment. The main goal of this paper is exploring a way to authenticate the RFID users with the cost efficiency and secure estimation advantages. The proposed mechanism inherits the benefits of existing positioning and authenticating technologies and applies a dynamic authentication method with a random generated node to help the reader estimates the distance and authenticates the user simultaneously. Our design focuses on determining the secure location of user extension by relying on RFID reader and computes distance deviation between time-based and coordinate-base estimation method.

Keywords: RFID, Security, Authentication, Positioning System.

1 Introduction

Radio Frequency Identification (RFID) is a rapidly developing wireless technology with key features which anticipate its outstanding position in the upcoming era of pervasive computing [1]. Today, RFID technology is constantly being improved and refined, and more and more industries are adopting it in new innovative ways throughout manufacturing and commerce. An RFID system always consists of two components: a transponders, or tags, which are attached to the objects intended for identification; a reader, or interrogator, which can read data from and possibly write data to transponders. The reader is typically connected to a computer or other control unit which runs higher level software for controlling the reader. The transponder usually consists of a coupling element and an electronic microchip.

Key features of RFID systems include a lack of physical contact between readers and tags, and tag scanning out of the line of sight [2], [3]. Moreover, a smart tag possesses storage and processing capabilities, and can also perform lightweight cryptographic functions. It is generally accepted that RFID systems can revolutionize various commercial applications, such as product management, transport payments, livestock tracking, library book administration, patient medical care and e-passports.

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_87,

However, the technology also poses threats including the possibilities of user information leakage and location tracking. A considerable amount of research has been published to provide possible solutions to these RFID security challenges. One approach to protect against such security threats is to use a tag authentication scheme in which a tag is both identified and verified in a manner that does not reveal the tag identity to an eavesdropper. Another possible requirement for RFID systems is secure tag ownership transfer. However, at the moment of tag ownership transfer, both the old and new owners have the information necessary to authenticate a tag, and this fact may cause an infringement of tag owner privacy.

The goal of this paper is to propose a secure RFID user authentication mechanism by using time-based distance estimation and coordinate-based localization. We introduce a new verification processing method on a basic RFID communication range and develop an authentication range and a verification range to provide user authentication and location verification. It can prevent users in the communication range from relay attacks and user impersonation attacks.

We list the current security issues for RFID system and related works in chapter 2. The proposed user authentication mechanism will be described in chapter 3 with a new secure distance estimation algorithm. In chapter 4, we analyze the proposed mechanism against two of typical attack models. Finally, we summarize the contributions of the paper and some future works will be introduced.

2 Related Works

2.1 Relay Attack on RFID Systems

RFID systems have their own vulnerabilities and security threats that are separate from the network model [1]. Especially, relay attack, known as wormhole attacks, which was already proposed in the 1970s have also become a recognized threat in wireless network security [3]. To execute a relay attack the attacker needs to build two devices locate between the real reader and tag. These devices act as a proxy-token and a proxy-reader respectively and they are connected via a suitable communication channel in order to relay information over a greater distance. Any information transmitted by the real reader is received by the proxy-token and relayed to the proxy reader, which will transmit the information to the real tag. The real tag assumes that it is communicating with the real reader and responds accordingly. The real tag's response is then relayed back to the proxy-token, which will transmit the information to the real reader.

2.2 Secure Issues in RFID Communication

Relay attack will cause the transmission delay and the only way to defend it is to use the distance estimation method or to locate the position of transmitter.

- *Distance Bounding Protocol*

Distance bounding techniques are used to upper bound the distance of one device to another (compromised) device. It is used by a verifier to verify that a claimant node being at a distance from a verifier node, cannot claim to be at a distance. This protocol was first introduced by Brands and Chaum [4] to prevent mafia fraud attacks.

- *Secure Positioning System*

Secure Positioning in Wireless Networks proposed by Srdjan Capkun [5] described security problems related to various positioning and distance estimation techniques. They showed how the devices can upper bound their mutual distances. They proposed a technique for position verification called VM algorithm. This technique enables a secure computation and verification of the positions of mobile devices in the presence of attackers.

- *RFID-assisted indoor localization*

The benefits of RFID motivated many researchers in exploring its potential for indoor localization proved in [6]. This method can be broadly divided into two classes: tag and reader localization, depending on the RFID component type of the target. In tag localization schemes, readers and possible tags are deployed as reference points within the area of inter positioning technique applied for estimating the location of a tag. The reader embedded at each user device queries for reference tags within its coverage in order to retrieve their IDs. Then, the list of the retrieved tag IDs with the corresponding RSS levels is forwarded to the location server within a TAGLIST message. The location server estimates the location for all users by employing a RFID-based positioning algorithm and finally returns the estimated locations back to the corresponding users.

3 Proposed User Authentication Mechanism

We now propose an authentication technique to verify user's position based on the round-trip time estimation and coordinate computation. The proposed user authentication mechanism is a technique for identifying the secure position of a normal RFID user with his/her RFID devices by using a reference verification point defined by the reader and unknown to the users.

3.1 Setting of Communication Environment

The necessary settings phases are described as below and table 1 lists the useful notations:

Table 1. Notations

R	the RFID reader
V	the reference verifier
U	the RFID communication user
K_E	pre-shared symmetric key
(u_x, u_y)	the coordinate of U
(v_x, v_y)	the coordinate of V
(T_U^r, T_U^s)	U's message sending and receiving time
(T_R^r, T_R^s)	V's message sending and receiving time
D_u	the commit value compute by (u_x, u_y)
D_v	the commit value compute by (v_x, v_y)
D_{UR}	the distance between user and verifier

- R sets the RFID communication range as the center of the communication range, with coordinate of (0, 0).
- R sets the authentication range when it detects U in its communication range and the distance between U and R is the radius of the defined verification range: $D_u = \sqrt{x^2 + y^2}$. The authentication range is a circle passed position of U around the origin R. The position of R is set to be the origin of the authentication range.
- R chooses two random numbers: x and y to construct a coordinate (x, y). Defines a point as a reference verifier V locate at the randomly generated (x, y), its position is represented as.
- R defines a verification range and the distance between V and R is the radius of the defined verification range: $D_v = \sqrt{x^2 + y^2}$. As the reference point, V is right located at circle of the verification range. The position of R is also set to be the origin of the verification range.

3.2 Proposed Mechanism

Related authentication and verification execute procedure for the communication system will be described and we initially model some kinds of attacks for simulating the most assumable attack situations. Here, we make some assumptions:

- *For RFID reader R*:
  - Its position is set to be the center of communication range as (0, 0).
  - It can estimate the distance with the users by using the proposed algorithm.
  - It cannot estimates user's location.
- *For user U with the RFID devices*:
  - It can estimate its own location
  - Location coordinate information is unknown to any others unless it is told.
  - R and U share a secret key K_E.

The operation of the authentication and verification procedures will be processed when a user comes into the RFID communication. Two sessions, communication session and authentication session, are described.

Session 1. Detection and Communication

- When a User with its RFID device comes into the communication range it will be detected by reader R and as a prover, communicate with R.
- The User estimates its own location (u_x, u_y).
- Compute the commit value: commitment (u_x, u_y) = commit (D_u) and then send u_x to R.
- R received the u_x and generates a circle with radius of u_x.
- Random choose a point on the circle called verifier V.
- Let V has the position of (v_x, v_y): commitment (v_x, v_y) = commit (D_v).

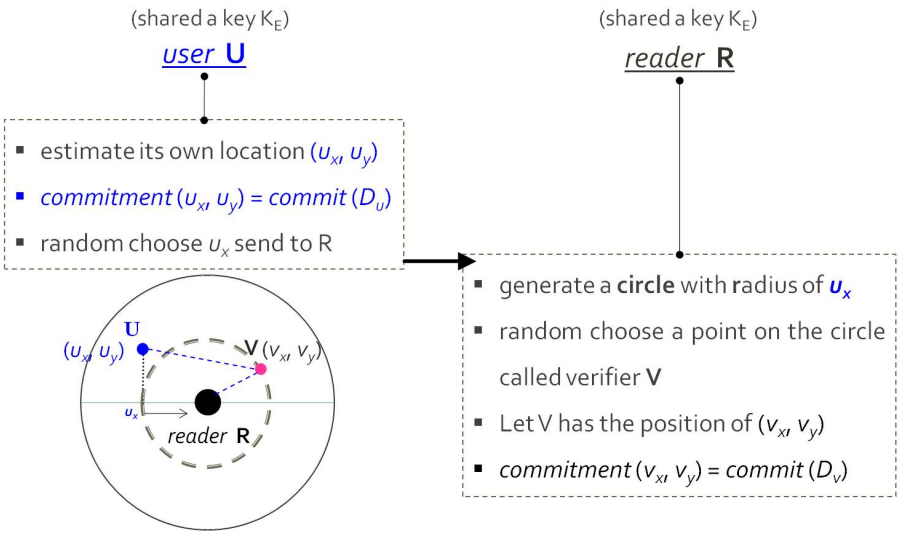


Fig. 1. Detection and Communication Process

Session 2. Authentication

- R starts the clock to record the message sending time when send D_v to user. The record is T_R^s.
- U records the receiving time as T_U^r.
- U computes the value $G_{uv} = y_u \oplus (Du - Dv)$ by using received D_v.
- The next message sending time recorded as T_U^s. (T_U^r, T_U^s) is encrypted with Guv by Message Authentication Code as $MAC_{KE} = (T_U^r, T_U^s, G_{uv})$.
- Then U will send the confirmed message together with MAC to R.
- After received the message send by U, R will records the receiving time as T_R^r.
- R will verify the received message by using the shared key K_E:
- MAC verification:
  $MAC_{KE} = (T_U^r, T_U^s, G_{uv})$
- Distance verification:
  $D_{UR} = (T_R^r - T_R^s - (T_U^r - T_U^s))/2)*c$
- User Authentication:
  Compute: $u_y = G_{uv} \oplus (D_{UR} - D_v)$
  Compare: $D_{UR} = d_{commit}(commitment(v_x, v_y), u_y)) = d_{commit}(u_x, u_y))$

If two values are equal, U confirms its authentication and R accepts the access of U; if they are not equal, U is considered to be an attacker who tries to disturb the communication between normal user and reader, the access will be rejected and communication session terminates.

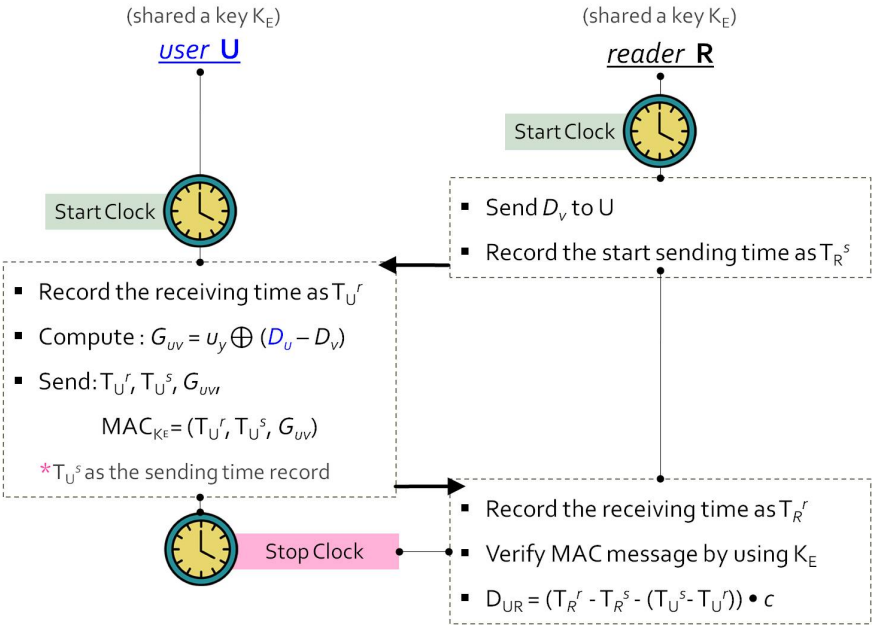


Fig. 2. Authentication Process

4 Security Analysis

We analyze the security statements for the defense of possible attacks in our proposed communication environment. We already introduced how our mechanism can provide user authentication and location verification. We now prove it can prevent users in the communication range of the transportation from relay attacks and user impersonation attacks. We define the attack models, and then analyze the secure requirements.

4.1 Attack Models

We briefly present our attacker model construct with the external and internal attacks. We call an attacker external if it cannot authenticate itself as an honest one to a central reader. We call an attacker internal if user controlling RFID device is compromised or if the RFID device is malicious. An internal attacker can report a false position or convinces the positioning infrastructure that it is at a false position. External attacks can convince an honest user and the positioning infrastructure that the user is at a different position from its true position (i.e., the attacker spoofs user's position).

Our proposal is based on a user-centric positioning system. By a user-centric positioning system a user computes its own position by observing signals received from public base stations with known locations mentioned in our assumption. The internal attacks can cheat on its position by simply lies about the position.

The external attacker should cheat on whole proposed protocol with timing performance in some way.

The internal attacker (User Impersonation attack) should obtain the shared secret key KE in a short time and therefore it has to cheat to successfully complete the protocol. The external attacker (Relay or Wormhole attack) is not in possession of the secret key KE, but should close enough to V to run the protocol successfully.

4.2 Protection Statements

In our security analysis, we describe relevant standard attacks on round trip time based distance measurement protocols and how to either minimize their impact or to defeat them fully.

- ***Relay or Wormhole Attacks***

If the attacker cannot speed up the transmission of the forwarded message, high additional transmission delay would be the result, and the attack will be detected by reader. Because the signal propagation is limited to light-of-speed, the attacker cannot speed up the transmission of the forwarded message.

Between the user and the reader, the verifier V can be placed randomly in a verification range circle. The attacker still cannot get any location information of the verifier but only knows the distance between reader and verifier D_v, and it is not enough to know the coordinate. The attacker cannot place any device near the verifier with the unknown position. Therefore, relay attacks are defeated by the proposed protocol.

- ***User Impersonation Attack***

An internal attacker close to user could try to obtain s in an online or offline way by eavesdropping the transmission message $\{T_U{}^r, T_U{}^s, G_{uv}, MAC_{KE} = (T_U{}^r, T_U{}^s, G_{uv})\}$. Because we use a 32-bit key size, the attacker should speed down the processing for key guessing online and re-transmit the key encrypted message in a certain short time. To reduce the delay of processing and re-transmission, the attacker should combine with the "place the device near the target" method again, which is difficult to succeed with the unknown position of verifier. It is especially difficult in a large communication range.

Although the attack can get close to the verifier in a short communication range, the delay of processing and re-transmission will be reduced more than in the large communication range case. In conclusion, user impersonation attack is not possible since reader can detect this distance fraud because of the processing and re-transmission delay.

5 Conclusion and Future Work

Recent advances in the use of RFID technology have generated significant interest in society, not only because they have brought change to the industry and business sectors, but also because they begin to influence our daily life more and more.

Although each RFID application has its own special requirements, security vulnerability is always a major concern when deploying RFID system. Case studies have been presented and discussed. We also discussed that our proposal is secure against the internal and external attacks. In our future work, we will simulate our proposed mechanism and show its efficiency.

Acknowledgement. The work was partially supported by Ewha Global Top 5 Grant 2011 of Ewha Womans University and World Class University Program (R33-10085) through National Research Foundation of Korea funded by the Ministry of Education, Science and Technology, and also supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2013R1A1A2011788). Kijoon Chae is the corresponding author.

References

1. Song, B., Mitchell, C.J.: Scalable RFID pseudonym protocol. In: Third International Conference on Network and System Security, pp. 216–224 (October 2009)
2. Avoine, G.: Cryptography in radio frequency identification and fair exchange protocols. Ph.D. Paper, Swiss Federal Institute of Technology (December 2005)
3. Juels, A.: RFID security and privacy: a research survey. IEEE Journal on Selected Areas in Communications 24, 381–394 (2006)
4. Brands, S., Chaum, D.: Distance-bounding protocols. In: Helleseth, T. (ed.) EUROCRYPT 1993. LNCS, vol. 765, pp. 344–359. Springer, Heidelberg (1994)
5. Capkun, S.: Secure Positioning in Wireless Networks. IEEE Journal on Selected Areas in Communications 24(2) (February 2006)
6. Papapostolou, A., Chaouchi, H.: RFID-assisted indoor localization and the impact of interference on its performance. Journal of Network and Computer Applications 34(3), 902–913 (2011)

A Study on a Bio-signal Biometric Algorithm on the Ubiquitous Environments

Sangjoon Lee, Sung Yun Park, Sung Jae Kim, Jae Hoon Joeng, and Sung Min Kim*

Department of Medical Biotechnology, Dongguk University-Seoul, 100715 Seoul, Republic
{sangjoon,sypark,jjh,joy4110,smkim}@dongguk.edu

Abstract. This paper is about the personal identification algorithm for adapting ubiquitous environment using electrocardiogram (ECG) that has been studied by a few researchers recently. The main characteristic of proposed algorithm uses together features analysis and morphological analysis method. The Principle Component Analysis (PCA) algorithm was applied for morphological analysis method and the features analysis method adapting to Support Vector Machine (SVM) classifier algorithm. We choose 18 ECG files from MIT-BIH Normal Sinus Rhythm Database for estimating algorithm performance. The algorithm extracts 100 heartbeats from each ECG file, and use 40 heartbeats for training and 60 heartbeats for testing. The proposed algorithm shows clearly superior performance in all ECG data, amounting to 93.89% heartbeat recognition rate and 100% ECG recognition rate.

Keywords: Biometric, Ubiquitous, Bio-signal recognition, Pattern recognition.

1 Introduction

AS internet and ubiquitous technology have rapidly developed recently, there is an increasing demand for biometric technologies for information security and for the prevention of invasion of privacy. Conventional biometric systems employ the unique appearance and characteristics of the human body including fingerprints, iris, face, hand vein, and gait. Additionally, biometric systems that have been commercialized up until now have had the trend that identification is performed based on only the features appearing on a human body, and most of them employ a camera as a sensor to measure the features of a human body. The trend causes problems including an increased erroneous identification ratio due to a change in appearance of the identification target, an increased identification time due to complex biometric algorithms, an increased erroneous identification ratio due to variations in the surrounding environment, an increased number of identification targets, and the identification of duplicated images (non-living target identification). To minimize these problems, conventional biometric systems focus on increasing the identification ratio by employing two or more identification elements rather than one[1]. Up until

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_88,

now in studies on ECGs and biosignals, algorithms have been developed to detect abnormal cardiac activity that accounts for 0.1% of the 24-hour cardiac activity. However, the biometric system is characterized by a method that extracts and identifies the features in the ECGs that take place in a steady state. A biometric system that uses biosignals is in its early stages of development, and there are no commercialized products yet. The initial approach to an ECG biometric system was introduced by Lena Biel [2] in 2001, and later, it was studied by a few scientists. Lena Biel measured the ECG of 20 subjects using the Megacart ECG signal measurement equipment from SIMENS and provided 30 features automatically from the ECG signal measurement equipment. Individual identification was done based on the extracted features using the SIMCA model [3], and 100% ECG signals identification ratio was obtained as a result. John M. Irvine [4] and Steven A. Israel [5] proved the validity that ECG can be used for individual identification even under a stressed heart state (from excitement or physical exercise), and designed the ECG identification method based on not the fiducial point but the ECG shape analysis. Although the recognition of the entire ECG signal had been studied even before the algorithm was proposed by John M. Irvine, the method was not appropriate for real-time identification of individuals. The morphological analysis method distinguishes the heart beat signals (the signals between RRIs) from the ECG data stream and performs identification by creating the Eigen pulse of the heart beat signals by means of PCA. The individual heart beat signals are identified from the entirety of ECG signals. Besides, Yongjin Wang [6] distinguished a few seconds of ECG signals from the entirety of the ECG signals and identified the ECG signals using the AC (Auto Correlation) and DCT (Discrete Cosine Transform) methods. Moreover, This paper proposes a hybrid type algorithm which uses morphological analysis and features analyzing method. It extracts candidate subset into the training data using morphological analysis and then compares between extracted candidate subset and testing data using features analyzing method.

2 Material and Method

2.1 Introducing the Proposed Algorithm

This paper introduces a hybrid type algorithm for ECG biometric. Previously published ECG biometric algorithms can be classified as two methods. One is the method that analyzes ECG features and the other is the morphological analysis of ECG waveform. The main characteristic of the hybrid type algorithm uses together two methods. Firstly, the morphological analysis can obtain the candidate subset among the training subset using PCA(principal component analysis). Secondly, the R-R intervalsegmentation of ECG was classified by Down Slope Tracing Waveform (DSTW) and extracted 7- features among the object heartbeat (testing subset) and candidate heartbeat (training subset), respectively. Lastly, the Support Vector Machine (SVM) classifier was performed to recognition among candidate subsets. The total block diagram of proposed algorithm is shown in Figure 1.

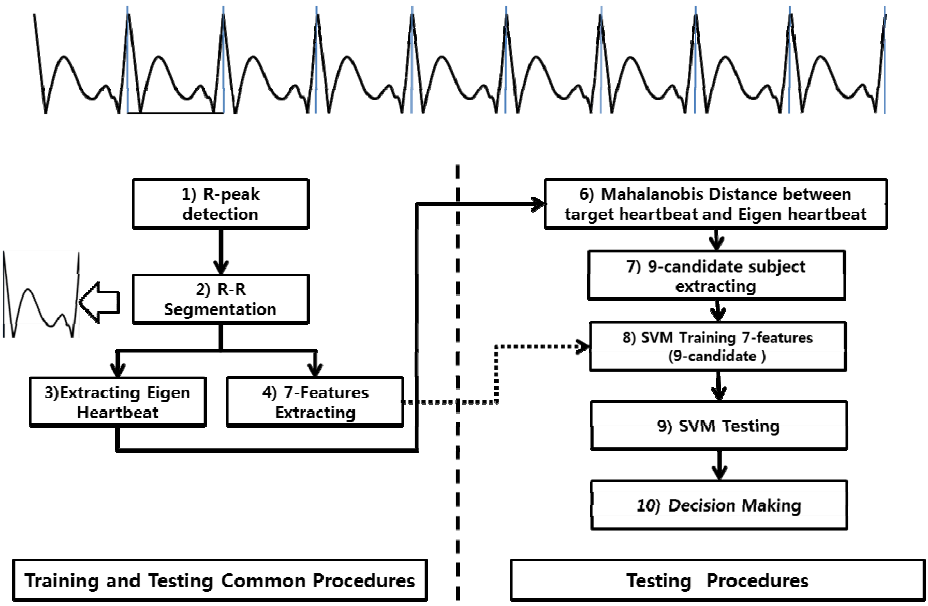


Fig. 1. The main block diagram of propose algorithm

2.2 Peak Detection and Feature Extraction

One of the most unique characteristic of ECG is its periodicity. A typical ECG signal has 3 important features, the P-wave, the QRS complex, and the T-wave. In addition, the R-peak can use a reference point to extract a periodic ECG. A DSTW (down slope trace waveform) is applied for detecting the R peaks of the ECG signal, easily detecting the peaks of the signal in real time. The DSTW algorithm can be used in various fields which detect the (P,Q,R,S,T) segments of the ECG, such as with fibrillation, reducing a power source noise, and a baseline wandering filter, showing an excellent peak detection rate in a signal with mixed amounts of various types of noise [7]. We surveyed the performance of the peak detection of the DSTW algorithm for all 48 instances in the MIT-BIH arrhythmia database. It was shown that 97.42% of the sensitivity and the 95.13% specificity could be determined. The proposed algorithm segments between current R peak and previous R peak when it finished R detecting procedures which is called RR segmentation data and then detects S, T peaks among RR segmentation data. The S, T detecting rule can be defined equation (1) and (2), respectively.

$$\mathrm{S(k)} = \mathrm{Index}(Arg_{min}[x(n)]),\ R(k) \le n \le R(k) + 30mS \tag{1}$$

$$\mathrm{T(k)} = \mathrm{Index}(Arg_{max}[x(n)]),\ S(k) \le n \le S(k) + 30ms \tag{2}$$

The k indices of detected peaks from 1 to M ($1 \le \mathrm{k} \le \mathrm{M}$), n also indices of sampled signal of $x(n)$. The actual obtained data is sampled with 125Hz.

The preprocessing procedure is finished for extracting of features by R, S, T peak detecting throughout the DSTW generation. The 7-features were chosen by detected R, S, T peaks as shown in Equation (3), (4), (5), (6), (7), (8), (9).

$$F_1(n) = R_{(n+1)} - R_{(n)},\ 1 \le n \le M \tag{3}$$

$$F_2(n) = Arg_{max}[X(n)],\ R(n+1) \le n \le R(n) \tag{4}$$

$$F_3(n) = Arg_{min}[X(n)],\ R(n+1) \le n \le R(n) \tag{5}$$

$$F_4(n) = R_n - S_n, 1\ \le n \le M \tag{6}$$

$$F_5(n) = S_n - R_{n+1},\ 1 \le n \le M \tag{7}$$

$$F_6(n) = R_n - T_n,\ 1 \le n \le M \tag{8}$$

$$F_7(n) = {}^{R}/_{T}\ , \text{Ratio of peaks value} \tag{9}$$

2.3 Eigen Heartbeat Extracting by PCA(Principle Component Analysis)

The algorithm should be normalized using previously setting average and standard deviation for all candidate heartbeats data to have same time and amplitude bandwidth by Equation (10).

$$S_k = \left[\frac{1}{n}\sum_{i=1}^{n}\left(x_k^i - Xu_k\right)^2\right]^{\frac{1}{2}},\ 0 \le k \le N-1 \tag{10}$$

Where $Xu_k$is average value of kth heartbeats data in the Equation (11)

$$Xu_k = \frac{1}{n}\sum_{i=1}^{n} x_k^i,\ 0 \le k \le N-1 \tag{11}$$

The Equation (12) is the amplitude normalization procedure.

$$\overline{y_k} = \frac{\left(x_k^i - Xu_k\right)\times\mu}{S_k+\sigma},\ 1 \le i \le n,\ 0 \le k \le N-1 \tag{12}$$

$$\overline{Y_k} = \text{Spline}(\overline{y_k}),\qquad 0 \le k \le N-1 \tag{13}$$

Where σ, μ are previously setting standard distribution and average, respectively. Where $\overline{y_k}$ is normalized vector data for amplitude, which is inputted heartbeats. The heartbeat data should be normalized for time bandwidth to perform PCA because it has a different time bandwidth although it comes from a same person. The heartbeat data length is arranged 1,000 point and applied to cubic spline interpolation method[8]. Where $\overline{y_k}$is normalized vector data for amplitude and time bandwidth in Equation (13). Proposed algorithm is implementing the PCA algorithm for extracting Eigen Vector. The Eigen Vector looks similar to heartbeats waveform so it is called Eigen heartbeat in the proposed algorithm. Figure 2 shows the shape of eigen heartbeats waveforms for 18 MIT-BIH normal sinus rhythm database[9]. If new heart beat come into for recognition, proposed algorithm calculates Eigen heartbeat for

inputting heartbeat and compares with the distance of prepared 18-MIT-BIH database Eigen heartbeat using the Mahalanobis distance method, and then 9-candidate Eigen heartbeats were selected by descending order of shortest distance.

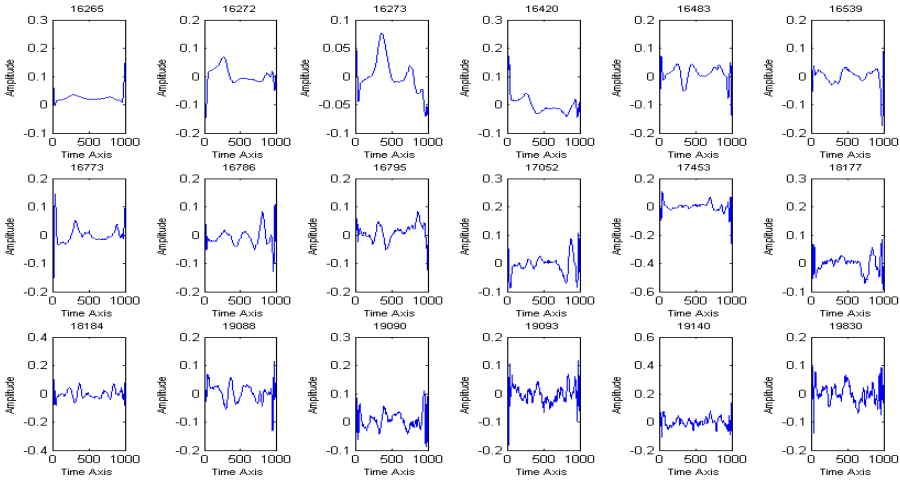


Fig. 2. Theeigen heartbeats of 18 MIT-BIH normal sinus rhythm ECG signals

2.4 Support Vector Machine (SVM)

The SVM (support vector machine) is new paradigm of pattern recognition for learning system which designed by Cortes and Vapnik[10]. The SVM was not noticed by the early, but recently has been widely used in various fields such as bio-informatics, letter, face and fingerprint recognition because of the superior performance in the supervised pattern recognition cases. The aim of SVM designs the hyperplane($w^T x + b = 0$) which having the maximum margin between decision boundary and training data and among the set of patterns that determine the boundary closer to the called support vector. The basic equation form of SVM can define the equation (14)

$$\min_{\mathbf{w},\varepsilon} \frac{1}{2}(\mathbf{w^T}\,\mathbf{w}) + C\sum_{i=1}^{l}\xi_i \tag{14}$$

Where the C is undecided constant variable, which is called normalized variable, can adjust the balance between classification error and the maximum margin.The two classes SVM case defines $y_i \in \{-1, +1\}$ but One against One SVM cases which has k classes, can define $y_i \in \{1, \ldots, k\}$which can expand 2-class SVM classifier to k-class SVM classifier using simple mathematical manipulating [11]. In order to recognize for inputting heartbeat, 7-features were acquired by section 2.2 method and then the k-class SVM algorithm performs among9-candidate subset which were selected by PCA performing. The 9-candidate subsets consist of 7-features subsets of each heartbeat.

3 Experiment

In order to evaluate the performance of proposed algorithm, 100 heartbeats are classified in each 18-MIT normal sinus rhythm and extracted 7-features at the each heartbeats. The experiment method is shown Figure 1. The input parameter of One Against One SVM has C that is the maximize error rate between the hyperplane and the maximize margin, is fixed 13 and evaluating the inter-symbol heartbeats recognition performance. The training data are adjusted from 10 to 50 and surveyed recognition rate about the remaining testing data. The recognition rate calculation is performed by the equation (15).The following table.1 is the recognition rate when the training data are adjusted from 10 to 50. The average recognition rate is shown 49.2% when PCA only performing but the hybrid type ECG biometric algorithm is shown the lowest recognition rate is 88.26% in average when the training data are 20 and the highest recognition rate is 93.89% when the training data are 40.

$$\text{Accuracy}(\%) = \frac{\text{Correct Match}}{\text{Total testing number of heartbeats}} \times 100 \tag{15}$$

Table 1. heartbeats recognition rate depending on training number of heartbeats

Recognition Rate depending on the training number(%)						
Training Number		**10**	**20**	**30**	**40**	**50**
Subject	16265	96.67	98.75	98.57	100.00	100.00
	16272	92.22	71.25	95.71	100.00	96.00
	16273	66.67	72.50	75.71	80.00	88.00
	16420	76.67	73.75	92.86	96.67	96.00
	16483	96.67	97.50	95.71	98.33	98.00
	16539	64.44	65.00	78.57	80.00	86.00
	16773	90.00	90.00	91.43	93.33	92.00
	16786	98.89	98.75	98.57	98.33	92.00
	16795	97.78	98.75	98.57	98.33	98.00
	17052	96.67	95.00	90.00	93.33	94.00
	17453	93.33	92.50	91.43	91.67	90.00
	18177	91.11	92.50	80.00	100.00	100.00
	18184	78.89	80.00	84.29	80.00	78.00
	19088	94.44	95.00	91.43	90.00	90.00
	19090	98.89	98.75	97.14	98.33	98.00
	19093	97.78	97.50	98.57	98.33	98.00
	19140	88.89	95.00	95.71	95.00	94.00
	19830	72.22	76.25	100.00	98.33	98.00
Average Rate(%)		88.46	88.26	91.90	93.89	93.67

4 Discussion and Conclusion

We has been tested a new hybrid type of ECG biometric algorithm, which uses together morphology and features analysis method. The recognition performance is shown 93.89% which meaning is possible to uses commercial type biometric system if it would be solve some problems. The reason why only choosing 7-features of heartbeat which shows a good representative of person identity. In the morphological analysis, we find out a high recognition rate when 9-candidates were chose by PCA The biosignal biometric system suggested to be equally evaluated in comparison with other commercialized biometric systems, identification performance should be verified in terms of the convenience of the measurement method using a large number of subjects with a wide range of age distribution. The private service application would be rapidly increasing issue and cyber security would be more important on the ubiquitous environment. To embed the ubiquitous device, the recognition algorithm needs to fast recognition time, less than computation time and ease measurement method. From the perspective of the convenience of the measurement method, a two-lead electrode ECG measurement system[12] needs to be fabricated to measure the ECG signals in a simpler manner.

References

1. Jain, A.K., Ross, A.: Multibiometric systems. Communications of the ACM 47(1), 34 (2004)
2. Biel, L., et al.: ECG analysis: A new approach in human identification. Ieee Transactions on Instrumentation and Measurement 50(3), 808–812 (2001)
3. Esbensen, K., Schönkopf, S., Midtgaard, T.: Multivariate Anal. in Practice, 1st edn., vol. 1. Camo, Trondheim (1994)
4. Irvine, J.M., et al.: eigenPulse: Robust human identification from cardiovascular function. Pattern Recognition 41(11), 3427–3435 (2008)
5. Israel, S.A., et al.: ECG to identify individuals. Pattern Recognition 38(1), 133–142 (2005)
6. Wang, Y., et al.: Analysis of Human Electrocardiogram for Biometric Recognition. EURASIP Journal on Advances in Signal Processing 2008, 1–11 (2008)
7. Kim, J., et al.: Development of an automatic external biphasic defibrillator system. Journal of Biomedical Engineering Research 25(2), 119–127 (2004)
8. Conte, S.D., De Boor, C.: Elementary numerical analysis: an algorithmic approach, 3rd edn., vol. xiii, p. 432. McGraw-Hill, London (1980)
9. The MIT-BIH normal sinus rhythm database, Physionet, `http://www.physionet.org/hysiobank/database/nsrdb/`
10. Cortes, C., Vapnik, V.: Support-vector network. Machine Learning 20, 273–297 (1995)
11. Kreßel, U., et al.: Pairwise classification and support vector machines. In: Advances in Kernel Methods-Support Vector Learning, pp. 255–268. MIT Press, Cambridge (1999)
12. Dobrev, D., Neycheva, T., Mudrov, N.: Simple two-electrode bio signal amplifier. Medical & Biological Engineering & Computing 43, 725–730 (2005)

Energy-Aware Profiler: An Energy Consumption Analysis Techniques for Offloading Communication-Intensive Mobile Apps*

K.O. Kwangman[1] and P.A.E.K. Yunheung[2]

[1] School of Computer and Information Engineering Sang Ji University, Seoul Korea
[2] Department of EECS, Seoul National University, Seoul Korea
kkman@sangji.ac.kr, ypaek@snu.ac.kr

Abstract. This study conducted analysis of traffic (sending/receiving) packet contents between smartphone and server through AndTweet and determined offloading candidates as 1) method irrelevant with UI generating traffic events in local smartphone 2) method not accessing computing resources of local smartphone after offloading 3) method irrelevant with maintaining state synchronization with local smartphone in the traffic process 4) method aiding traffic-related methods. This study was able to detect methods consuming energy mainly in communication-intensive applications and is expected to contribute to reduction of burden of battery consumption of mobile devices such as smartphone by means of offloading technique.

Keywords: Communication Offloading, Energy Optimization, Mobile Computing, Cloud Computing.

1 Introduction

In line with continuous increase of use of mobile devices such as smartphone and tablet PC with limited computing resources and capability, research on how to cope with mobile apps which need high performance computing resources for complicated computation has come to the forefront as an important issue. Particularly along with steady efforts for performance improvement of processing speed, storage capacity, is emphasized the importance of technical management in the area of software to prolong the use hours of battery with limited capacity. Under the circumstances, research on how to minimize and manage battery consumption of applications processed in mobile devices must be an important research subject in the related field[1].

There have been previous studies on a diverse range of offloading frameworks with a view to reducing of computing load on mobile system through offloading. That said, most of them focused on how to overcome computing burden by means of

* This work(Grant No. C0102874) was supported by Business for Cooperative R&D between Industry, Academy, and Research Institute funded Korea Small and Medium Business Administration in 2013.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*, Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_89,

computation offloading of units requiring complicated computation or long time of processing by determining in a static or dynamic way, and consequently to reduce burden of battery consumption. However, according to recent study reports, communication-intensive movement becomes the main cause of battery consumption not less than complicated and repeated computation, as one of characteristics of mobile apps[2].

This study conducted analysis of traffic(sending/receiving) packet contents between smartphone and server through AndTweet, android open source, and determined offloading candidates as 1) method irrelevant with UI generating traffic events in local smartphone 2) method not accessing computing resources of local smartphone after offloading 3) method irrelevant with maintaining state synchronization with local smartphone in the traffic process 4) method aiding traffic-related methods.

In Section 2, we first introduce Offloading meanings and benefits with related works. Then in Section 3, we describe our communication-intensive offloading model and energy profiling approaches. Our final experimental result demonstrates on conference sites for more reliability and complementary results. Section 4 we conclude this works and explain future research directions.

2 Related Works

2.1 Communication Offloading

Offloading is a solution to augment these mobile systems' capabilities by migrating computation to more resourceful computers (i.e., servers). This is different from the traditional client-server architecture, where a thin client always migrates computation to a server. Computation offloading is also different from the migration model used in multiprocessor systems and grid computing, where a process may be migrated for load balancing [1.3]. The key difference is that computation offloading migrates programs to servers outside of the users' immediate computing environment; process migration for grid computing typically occurs from one computer to another within the same computing environment, i.e., the grid[1].

In general offloading approaches, the common and core issues to make offloading decisions based on various factors, such as "offloading goals(improve performance or reduce energy dissipations), offloading decision time(static, dynamic, or hybrid), infrastructure for offloading, and so on.

2.2 Energy Profiling Tools

Analysis of factors and causes of battery consumption by applications processed in smartphone shows that frequent number of traffic between mobile devices or with server, volume of traffic, pattern of traffic are also crucial factors contributing to battery consumption, while many resources and battery are consumed by long processing time because of complicated computation and execution. According to recently reported papers, research on optimization of traffic patterns of smartphone and improvement of throughput through offloading can be effectively used as a means

to reduce battery consumption of smartphone. It is allegedly due to the fact that traffic signs [4], interval pattern of packets or throughput are core elements causing battery consumption [5]. Eprof's research [6] analyzed energy consumption volume of smartphone apps and suggested measuring instruments as well as the result of concentrated energy consumption in components, advertisement modules related with I/O events including 3G, WiFi, GPS [7].

SmartDiet[8] toolkit to identify the constraints that reduce offloading opportunities and to calculate the energy-saving potential of offloading communication-related tasks. SmartDiet traces the method-level application execution and estimates the allocation of communication energy cost from traffic traces. We discuss key features of SmartDiet and show some preliminary results using a prototype implementation.

3 Energy-Aware Profiler

3.1 System Configurations

In almost previous researches, it overcomes computation burdens and reduces battery consumptions on local devices through computation offloading that decides higher computation complexity and consumes long execution time units on static and dynamic decision. But, Communication-intensive methods are main causes that consume battery life –time as complicated and repeated computation on modern mobile applications.

In this paper, we reduce the burden of energy consumption by communication-intensive offloading. For this works, we implements a source-level application analysis tool, an energy profiler based on Eprof[6] and SmartDiet[8], migrates offloading informations and methods to nearest offloading server. This approaches has a some differential challenges that automatically changes an offloading restricting elements with consideration of communication and server status. Specially, Optimization techniques are used to information that migrated to offloading server.

As following Fig. 1, In Non-offloading case, Smartphone has communication burdens which consumes energy with send and receive messages. In Communication-Offloading case, energy profiler detects hot energy consume regions and migrates to remote potential server. Offloading server send/receive messages to communication server with heavy traffic patterns and a lot of send and receive message, but smartphone has light traffic and message passing

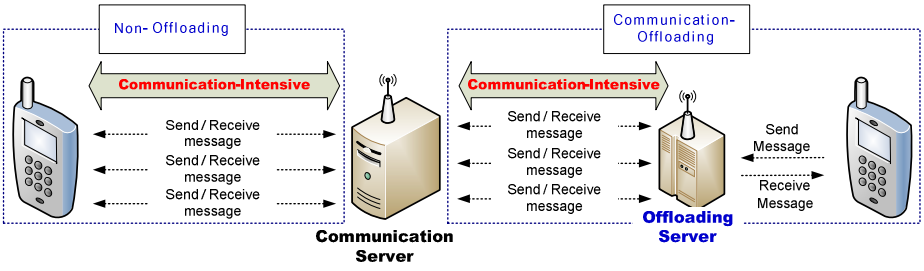


Fig. 1. Non-offloading vs. Communication-intensive Offloading

Through this approach, smartphone's communication processing burdens are reduced and eventually battery consumptions are minimized.

3.2 Energy-Aware Profiling

Energy-aware profiler detects offloading candidate methods causing concentrated energy consumption by means of compiler static source analysis for communication-intensive applications and creates energy-aware profiles & classes suitable to the model with ability to measure communication-intensive consumption. For this, profiler constructs data flow graph (DFG) and control flow graph (CFG) in applications for analysis of source codes and collects information such as computing amount, traffic amount and flow during applications processing. The core factor to determine offloading in this study is for smartphone, a local device to capture all traffic IP and monitor information of traffic number and amount and analyze traffic packets and then profile related methods.

This study conducted analysis of traffic (sending/receiving) packet contents between smartphone and server through AndTweet-Android open source - and determined offloading candidates as 1) method irrelevant with UI generating traffic events in local smartphone 2) method not accessing computing resources of local smartphone after offloading 3) method irrelevant with maintaining state synchronization with local smartphone in the traffic process 4) method aiding traffic-related methods.

4 Conclusions

There have been previous studies on a diverse range of offloading frameworks with a view to reducing of computing load on mobile system through offloading. That said, most of them focused on how to overcome computing burden by means of computation offloading of units requiring complicated computation or long time of processing by determining in a static or dynamic way, and consequently to reduce burden of battery consumption. However, according to recent study reports, communication-intensive movement becomes the main cause of battery consumption not less than complicated and repeated computation, as one of characteristics of mobile apps.

This study was able to detect methods consuming energy mainly in communication-intensive applications and is expected to contribute to reduction of burden of battery consumption of mobile devices such as smartphone by means of offloading technique.

References

1. Dinh, H.T., Lee, C., Niyato, D., Wang, P.: A survey of mobile cloud computing: architecture, applications, and approaches. In: Wireless Communications

2. Saarrinen, A., et al.: Can Offloading Save Energy for Popular Apps. In: Proceedings of the 7th ACM International Workshop on Mobility in the Evolving Internet Architecture (MobiArch 2012), pp. 3–10 (2012)
3. Powell, M.L., Miller, B.P.: Process Migration in demos/mp. ACM SIGOPS Operating System Riviews 17(5) (1983)
4. Signals Research Group. Smartphones and a 3g network: Reducing the impact of smartphone-generated signaling traffic while increasing the battery life of the phone through the use of network optimization techniques. Technical report, LLC (2010)
5. Xiao, Y., Savolainen, P., Karppanen, A., Siekkinen, M., Yla-Jaaski, A.: Practical power modeling of data transmission over 802.11g for wireless applications. In: e-Energy 2010: Proc. of the 1st Int'l Conf. on Energy-Ecient Computing and Networking. ACM (2010)
6. Pathak, A., Hu, Y.C., Zhang, M., Bahl, P., Wang, Y.-M.: Fine-Grained Power Modeling for Smartphones Using System Call Tracing. In: Proceedings of the 6th ACM European Conference on Computer Systems (EuroSys 2011), pp. 153–168 (2011)
7. Pathak, A., Hu, Y.C., Zhang, M., Bahl, P., Wang, Y.-M.: Where is the energy spent inside my app? Fine Grained Energy Accounting on Smartphones with Eprof. In: Proceedings of the 7th ACM European Conference on Computer Systems (EuroSys 2012), vol. 42, pp. 29–42 (2012)
8. Saarinen, A., Siekkinen, M., Xiao, Y., Nurminen, J.K., Kemppainen, M., Hui, P.: SmartDiet: offloading popular apps to save energy. In: Proc. of ACM SIGCOMM 2012, pp. 297–298 (2012)

A Study on CKP Signal Collection Algorithms for Knocking Identification and Development of Engine Diagnosis System in CRDI ECU

Hwa-seon Kim, Seong-jin Jang, and Jong-wook Jang

995 Eomgwangno, Busan jin-gu, Busan, 614-714, Computer Engineering,
Dong Eui University, Korea
rainwood@dreamwiz.com, ch99jin@hanmail.net, jwjang@deu.ac.kr

Abstract. As an abnormal combustion caused by the compression and spontaneous ignition of unburned mixture during the latter process of combustion, if knocking occurs, it will result in loss of engine power as well as severe damage to engine. Therefore, it is very significant to monitor knocking. Generally, a practical method of monitoring knocking is to use a knocking sensor. However, by this method it is not possible to distinguish between engine's own vibrations and knocking if the vibrations match up with the range of the knocking sensor signal for the use of detection of knocking. In this study, knockking was diagnosed by way of using the signals of CKP sensor which is basically installed in the engine, instead of using such a problematic knocking sensor. Moreover, a mobile diagnostic system based on OBD-II for industrial CRDI engine has been developed. Since the developed engine diagnosis system makes it possible for an administrator to monitor automotive information in real time wihtout any other equipment, the person may promptly respond to occurance of malfunction in engine.

Keywords: CKP(Crankshaft Position Sensor: CPS), CMP(Camshaft Position Sensor: TDC), Knocking, Longtooth, OBD-II, DTC(Diagnostic Trouble Code).

1 Introduction

1.1 Background and Purpose of This Study

An internal-combustion engine generates power with force pushing down on the top of the piston by the ignition and explosion of the fuel/air mixture at a precise timing. At this time, if the mixture is unusually ignited and exploded at the other times rather than the "precise timing", improper combustion occurs followed by exceptionally high pressure as a result of which the piston hits the cylinder. Such a phenomenon is called knocking[1-3].

Strong shockwaves caused by knocking destroy the thermal boundary layer between the combustion chamber and the surface of cylinder, resulting in sharp increase in the amount of heat transferred to the surface of cylinder. On this account, surface ignition occurs on the cylinder head and piston, and the function and

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_90, © Springer-Verlag Berlin Heidelberg 2014

efficiency of engine deteriorate. In addition, a sharp increase in pressure which partially occurs in the combustion chamber causes vibrations of cylinder block. If such a condition continues, shockwaves function as fatigue load over all the engine parts, which damages the components of engine [4-6]. Therefore, it is very significant to detect knocking.

It is a general and practical method of using a knocking sensor detecting vibration signals in order to diagnose knocking. When knocking is identified with the sensor adhered to the cylinder, however, there is a difficulty in making an accurate diagnosis because it is impossible to distinguish between vibrations caused by knocking and automobile's own vibrations. In order to solve this problem, it is necessary to install an extra sensor for detection of automobile vibrations besides the sensor adhered to the cylinder. Then, this would affect the automobile weight and bring on assembling matters such as the location to install the sensors during the course of assembly. The number of knocking sensors is a factor that affects the weight of automobile, and the weight influences the fuel efficiency. Of course, a question may arise in regard to how much the weight of the two sensors would affect the weight of automobile but in case of automobiles recently manufactured, various sensors and convenient equipments are installed, and the efforts to reduce the weight of automobile have been made. Moreover, since knocking detection may differ depending on the location of the knocking sensor, if knocking can be detected by another method besides the knocking sensor, such a study is needed. By this reason, this study intends to make a detection of knocking by using a built-in sensor of vehicle instead of the knocking sensor which is separately attached thereto. Such a built-in sensor is called crankshaft position ("CKP") sensor. As the most significant sensor to determine the amount and timing of basic fuel injection with calculation of engine rpm and crank angle, the CKP sensor detects the location of tone wheel [7-9]. Knocking can be detected by using such features of this sensor and knocking detection can be more accurate together with use of a camshaft position ("CMP") sensor to verify knocking in each cylinder.

2 Development and Implementation of Signal Collection Algorithms

As shown in Figure 1, ECU controls knocking by way of receiving knocking signals from the knocking sensor. In this study, among input values of CRDI system, the values of the CKP and CMP sensors for use of determining the injection timing at the

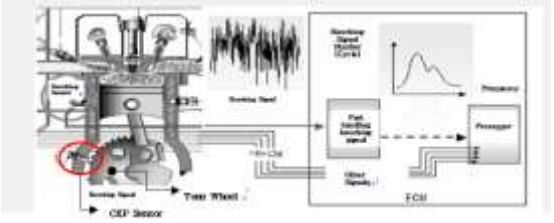


Fig. 1. ECU & Knocking Signal[10]

time of engine start, instead of the knocking sensor, were used to detect knocking. The CKP sensor described in Fig.1 is the sensor to be used in this study.

2.1 Waveform Analysis of CKP and CMP

In the 4-cycle engine, in order to calculate the ignition timing and fuel injection timing, each stroke's identification and specially knowing exactly when the compression TDC is coming on are important. If what degrees of BTDC(Before Top Dead Center) should be more efficient to ignite can be calculated in advance, the ignition at BTDC can be implemented. The ECU should know reference points (missing tooths) to calculate the exact TDCs(top-dead center) and BDCs(bottom-dead center). After occurring the CMP signal, number 1 TDC is the 19th tooths' position; based on the missing tooth, this point is before 114°, so that the ignition timing can be known by calculating degrees of the BTDC if the number of tooths from the missing tooth is calculated[7-9].

2.2 System Configuration

In this study, algorithms, which can provide baselines to identify the car's knocking by collecting control sensor values from the simulator, are implemented. Figure 2 is a configuration diagram to receive control sensor values from the simulator. On the car or simulator mounted the CRDI engine, through the Encoder or CPS(Crankshaft Position/angle Sensor), to measure the knocking sensor and important engine control sensor, the sensor values are collected by using the DAQ board; the values are transmitted to a laptop via USB communications connected with it; and the values are analysed on it. By using these values, to customize the mapping for the improved. CRDI engine control, an algorithm for knocking identification and correction, which can provide the optimal Knocking identification baseline by analysing and processing the useful sensor information, is implemented.

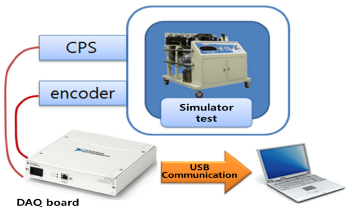


Fig. 2. System Configuration

2.3 Algorithm Development

The knocking of the vehicle can occur diesel knock when the ignition delay period is getting longer; to prevent this diesel knock, one of the methods is to control the injection timing. Therefore, if algorithms controlling the fuel injection timing and

injection amount are implemented, the fuel injection timing can be controlled through the knocking identification.

2.3.1 Knocking Identification Algorithm

If the measured acceleration is greater than the previous acceleration compared with the acceleration of each CPS as shown in Equation (1), the fuel injection timing is controlled by identifying the knocking[11][12]. Figure 3 is a flowchart of the algorithm determining the knocking.

$$\Delta t = \text{last timing} - \text{initialtiming} = t_1 - t_0 \quad (1)$$

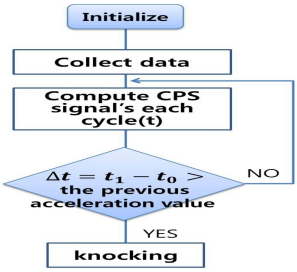


Fig. 3. Flowchart of knocking identification algorithm

3 Development and Implementation of Engine Diagnosis System

By way of using the OBD-II standard, automobile-centered diagnostic equipment was developed to provide driver-centered diagnostic services. By way of using cable and Bluetooth module which is a wireless system, it made it possible to provide real-time communication over signals from automotive malfunction diagnosis and sensor output.

3.1 Design of Bluetooth OBD-II Protocol

The developed OBD-II protocol has been manufactured based on the existing BOD-II standard. This differs from the existing BOD-II standard protocol structure. In case of the OBD-II protocol standard, automotive information of only one PID which was requested can be read and responded. However, the developed industrial automotive OBD-II protocol read all the automotive information and transmits such information at once.

3.1.1 Structure of Bluetooth OBD-II Protocol

The OBD-II message may be obtained from the automotive ECU by using automotive diagnostic tools. As seen in Table 1, the message consists of Header, Data and Checksum and saves 12 bytes of data in total and uses HEX codes.

In case of the proposed OBD-II protocol, it is designed to read the whole sensor information of vehicle by a response message at once when automotive information is requested to ECU. ECU provides 31 types of sensor information which actual service centers practically use. The below Table 2 shows the proposed OBD-II protocol response message structure.

Table 1. Proposed OBD-II Protocol Request Message Structure

Command STX	Command ID	Info	Opt1	Opt2	Checksum	Command ETX

Table 2. Proposed OBD-II Protocol Response Message Structure

Data STX	Data1	.....	Data31	Checksum	Data DTX

3.1.2 Structure of Engine Trouble Codes

There is a function to inform drivers that there is malfunction in the electronic control engine by lighting up the malfunction indicator lamp ("MIL") and to set diagnostic trouble code ("DTC") according to the details of malfunction and to automatically record such codes in the RAM of ECU if there is malfunction in electronic control engine or in exhaust gas related parts. This function was originally to set the BOD in order to easily verify the location to be inspected if automotive malfunction occurs but thanks to speedy development of computers, it came to play a role of conducting ready-test (monitoring of exhaust gas equipment) as well as making freeze frame (function to record DTC on ECU) when malfunction occurs in input and output of ECU (computer).

Therefore, a self-diagnosis function is the priority to be inspected when malfunction occurs in the car equipped with electronic control engine. The below Table 3 shows the structure of DTC response message of ECU.

Table 3. ECU DTC Response Message Structure

Command STX	Command ID	MODE	DTC Code	Checksum	Command ETX

3.2 Automotive Information Collection Algorithm

In order to collect the information of ECU, data are transmitted through the process as described in the below Figure 4. First of all, if Bluetooth communication is connected, a data request message is transmitted to ECU. If the input request message is identical to 023134303030303030454303, ECU transmits OBD-II response message of 130 bytes to temporary buffers. All the data between STX and ETX are converted into HEX codes and sent. The calculation of Checksum is of longitudinal redundancy check ("LRC") and the lower byte of the sum of data between STX and ETX and Checksum should be zero. If the response message is ACK(0x06) and the Checksum value is 0x00, 31 automotive data of 130 bytes are finally saved.

In order to collect DTC of ECU, such codes are transmitted through the process as described in the below Figure 5. Collection process of ECU diagnostic trouble codes is similar to the automotive information collection algorithm as explained above. First of all, if Bluetooth communication is connected, data request message is sent to ECU. If the input request message is identical to 023135303030303030454203, ECU transmits OBD-II response message of 14 bytes to temporary buffers. Then, if response message is ACK(0x06) and the Checksum is 0x00, automotive information of 14 bytes becomes finally saved.

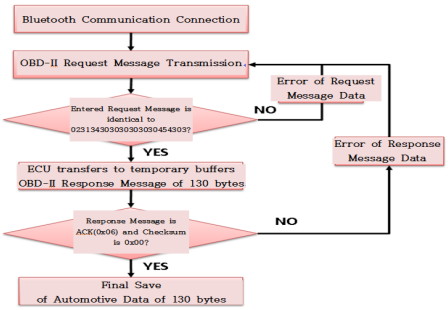


Fig. 4. Flowchart of Automotive Information Collection Algorithm

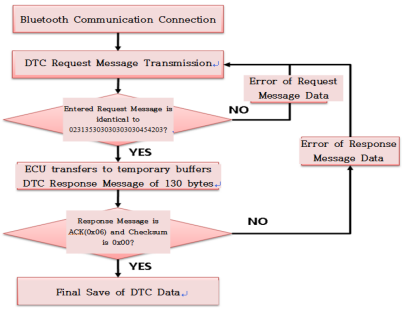


Fig. 5. Flowchart of Automotive Diagnostic Trouble Code Collection Algorithm

4 Experiments and Results

4.1 Knocking Identification

In this section, by using values of the sensors collected on the designed simulator, algorithms of knocking identification and engine balance correction are developed. Figure 6 is a screen of the program developed to collect data signals, and it shows waveforms of CPS and TDC signals collected at real-time on the simulator.

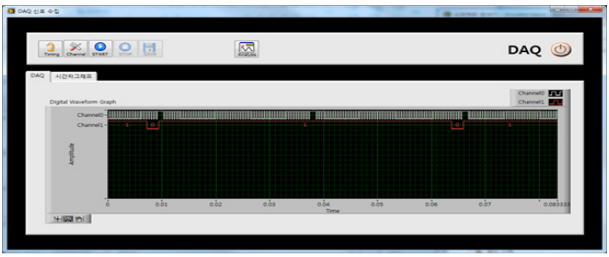


Fig. 6. Screen of data signal collection program

Figure 7 shows the result of using the knocking identification algorithm calculating the acceleration difference between the current time and the previous time. The long

waveform is a reference point(missing tooth); the short one, which is between one missing tooth and another missing tooth, is the position occurring knocking.

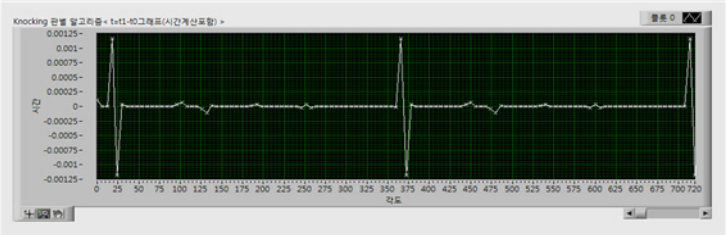


Fig. 7. Result Screen of Knocking Identification algorithm

4.2 Bluetooth Mobile Application Software for OBD-II Protocol Diagnosis

The following shows a screen used for communication between devices. If clicking the button named “Data Request” in order for application to request the connected device for automotive status information, the connected device transmits the status information to the application. Figure 8 shows status information communication between devices. If phone orders ECU to read status information through communication protocol, ECU sends out a data response. Figure 9 is a screen showing DTC information transmission between devices.

Fig. 8. Device status information Communication

Fig. 9. Device DTC Communication

The communication between devices is made with HEX codes, so it is difficult for a user to verify the information instinctively. In order to make a user promptly understand the status information, such information was made go through parsing process so that it may be shown on the screen as in Figure 10.

When malfunction occurs in ECU, the concerned trouble codes are searched and if the data which are identical to codes saved in DB exist, information related to such codes are notified.

If DTC is found, such found DTC is shown on the screen as in Figure 11.

Fig. 10. Status information

Fig. 11. DTC Information

5 Conclusion

In this study, instead of using a knocking sensor, algorithm to identify knocking by using the CKP sensor which is basically installed in almost all the engines has been developed.

While it is said that a use of knocking sensor is convenient and highly accurate, it cannot be said that it is the best solution for knocking identification because there exists a problem: if internal/external automotive environment and engine condition are not favorable, ignition timing may be delayed due to excessive knocking detection, resulting in power reduction. If there is a method free from such problem, such method should be researched. Therefore, in this study knocking was identified through the difference between angular speeds of the crank by using CKP sensor signals. Factors that cause excessive knocking detection were not applied to this method, so more stable identification of knocking could be made. In addition, through Longtooth identification algorithm, accurate knocking identification was made possible. Since the engine balance also has direct impacts on engine power and durability, algorithm to verify cylinders whose balance is to be adjusted has been developed so that all the cylinders could keep the same balance.

Moreover, with OBD-II protocol, a mobile engine diagnostic system using Bluetooth communication was developed. In this study, instead of handling information that can be controlled only by manufacturing companies, it was made possible to select necessary information only and take control at first hand.

It is unnecessary to passively receive and deal with all the data including even needless one, so the administrator may handle information satisfying his needs only. Therefore, with this system it was made possible that information of engine condition may be identified in real time and that if engine has malfunction, by notifying diagnostic trouble codes and information, the user and administrator may promptly respond to such malfunction.

Acknowledgments. This work was supported by the Brain Busan 21 Project in 2013.

References

1. Heywood, J.B.: Internal Combustion Engine Fundamentals. Mc-Graw Hill (1988)
2. Renault, F.: A New Technique to Detect and Control Knock Damage. SAE Paper820073 (1982)
3. Maly, R.R., Klein, R., Peters, N., Kang, G.: Theoretical and Experimental Investigation of Knock Induced Surface Destruction. SAE Paper 900025 (1990)
4. Kim, J.-J.: Basic Research on Knocking Control Using Block Vibration Signal by Spark Ignition Engine. Graduate School of Yonsei University. Thesis for Master's Degree (1993)
5. Lee, J.-H.: Development of Spark Ignition Engine Knocking Control Algorithm. Graduate School of Hongik University. Thesis for Master's Degree (1996)
6. Ham, Y.-Y.: Development of Spark Ignition Engine Knocking Control Algorithm and Establishment of Knocking Identification Base Value. Graduate School of Hongik University. Thesis for Doctor's Degree (1996)
7. Jung, Y.-G.: Electrical Electronic and Diagnosis of Common Rail Engine Automobile, pp. 99–220. Naeha Publisher (2008)
8. Park, J.-R., Baek, T.-S., Ahn, Y.-M., Choi, D.-S.: Automotive Engine. Golden Bell, 307–334 (2003)
9. Yoon, J.-G., Chun, D.-J., Cho, I.-Y., Ha, J.-S., Choi, D.-S.: Automotive Diesel Engine. Mijeon Science, 246–264 (2002)
10. http://blog.naver.com/hsj3117/140108324901
11. Shin, S.-Y., Jang, D.-H., Lee, H.-C.: Telematics Specific Horizontal Distance Traveled by a Falling Car. Journal of Information and Communication Convergence Engineering 10(2), 181–186 (2012)
12. Angkititrakul, P., Miyajima, C., Takeda, K.: Stochastic Mixture Modeling of Driving Behavior During Car Following. Journal of Information and Communication Convergence Engineering 11(2), 95–102 (2013)

How to Detect Obstacles within the Lane through Smartphone-Based Lane Recognition*

Hwan Heo, Taeg-Keun WhangBo, and Gi-Tea Han

Gachon University, 1342 Seongnamdaero, Sujeoung-gu, Seoung-si, Gyeonggi-do, Korea

Abstract. This paper proposes a method of detecting obstacles existing on the lane making use of lane recognition based on smart-phone. In the proposed method, after detecting lanes by means of inverse perspective transformation, the 1st step is to detect a obstacle candidate region in terms of dispersion map using dispersion values at the interested regions set up in the detected lane, and when multiple candidate regions are detected, the 2nd step is to detect characteristic points at the interested region with a FAST corner detector and select the region that has the overlapping obstacle candidate position with the 1st step result as the obstacle. The proposed method showed good obstacle-detection performance of 80~90ms for processing 1 frame image by reducing processing region and simplifying processing process.

Keywords: smart-phone base, lane recognition, obstacle detection, image processing, safe driving service.

1 Introduction

With the prevailing future prospect that the value of an automobile will depend on IT technology such as communication within a car rather than mechanical specification, auto makers are concentrating on high added value system such as intelligent cars. One of composition elements of such an intelligent car is Lane Departure Warning System which uses numerous input sensors, such as imaging, radar, and laser, to analyze and process imaging information received from sensors and support safe driving [1, 2, 3]. This paper proposes a smart-phone base obstacle detection method making use of this image processing technology.

Real time algorithm to judge lane departure in terms of smart-phone camera should adopt simple and small arithmetic operation to reduce processing time. Recently, instead of simply using color or edge information directly from image, it is under study to transform image in the pre-processing stage and make the information easy to use [2, 3].

As a method to use transformation of image, Alberto Broggi proposed a method which calculates information necessary for perspective transformation using camera

* This research was supported by MSIP (the Ministry of Science, ICT and Future Planning), Korea, under the IT-CRSP(IT Convergence Research Support Program) (NIPA-2013-H0401-13-1001) supervised by the NIPA(National IT Industry Promotion Agency)

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_91,

parameters and eliminates distance effect to create top-view for lane detection [4]. This method makes lane information easy to use by transforming image to Top View by means of IPM (Inverse Perspective Mapping). Also, it was possible to save detection time by designating only the region where lane exists, instead of using the entire image. Although there was a disadvantage of requiring camera parameters, recent research [5] developed an algorithm to obtain the same result as IPM without camera parameters. This is a method that can extract the homography matrix H necessary for inverse perspective transformation through lane prediction in image instead of using camera parameters [5].

In this paper, inside of detected lane is set up as interested region and obstacle is detected making use of dispersion and characteristic points in the interested region. At this moment, lane is detected using inverse perspective transformation without camera parameters.

Extraction of characteristic points is to find characteristic points existing inside a image. Well known methods for this include Harris corner detector, hessian detector, fast hessian detector, FAST corner detector.

Hessian detector uses hessian matrix to extract characteristic points and is suitable for detecting characteristic points such as spots. Harris corner detector detects characteristic points using the property that the corner has high curvature in two directions, eigenvalues, and corner response function.

Instead of considering all nearby pixels to detect corner, FAST corner detector apprehends nearby pixels through machine learning and sets up the order of nearby pixels using this. It is possible to detect characteristic points relatively accurately and very fast by removing what is not corner in advance using this order [8]. Therefore, this research uses FAST corner detector to detect obstacles.

2 Proposed Content

To detect obstacles in the car forward direction under the smart-phone base environment, lane detection is conducted first. Lane detection uses inverse perspective transformation [5] which does not need camera parameters, but blurring effect may appear in this method and may drop the obstacle detection rate. To eliminate this blurring effect, ideal lane is normalized to a certain width, and after inverse perspective transformation is applied, the region inside lane is set up as interested-region. To detect obstacles, the 1st step is to create dispersion map in the interested-region and set up the obstacle candidate region. The 2nd step is to detect characteristic points in the interested-region with a FAST corner detector and detect the final obstacle among the obstacle candidate regions.

2.1 Reset of Ideal Lane

The result of inverse perspective transformation used in lane recognition appears mostly as a top-view type image, such as the one seen from sky.

Because the information existing at the image upper part is spread to wider space by inverse perspective transformation, it looks as if blurring was done, while the image lower part looks sharp because it is projected from wide space to narrow space. The result of applied inverse perspective transformation [5] which needs no camera parameter looks as in below (Fig. 1).

a. Input image b. Transformed image

Fig. 1. An examle of IPM algorithm

The purpose of this paper is to detect obstacles in the drive way forward direction, and to make easy detection of obstacles, the width of ideal lane used in inverse perspective transformation is normalized to a certain constant number as in below (Fig. 2) to show equivalent results for all lanes.

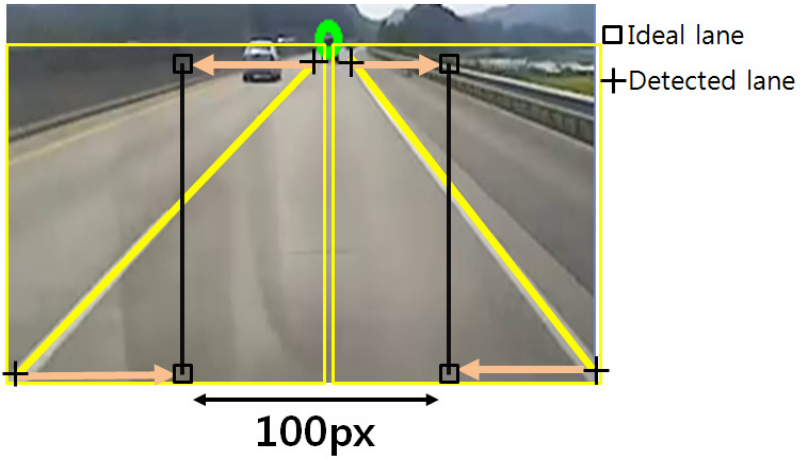
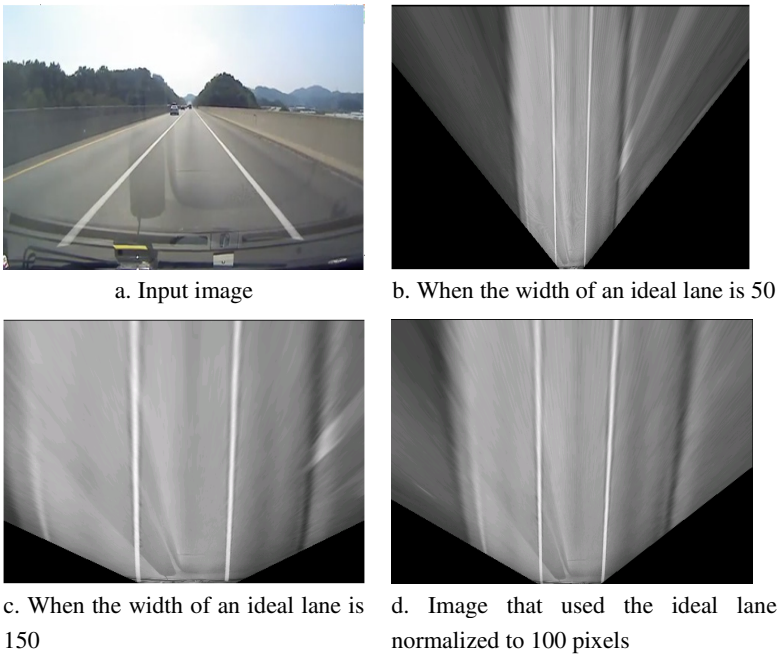


Fig. 2. Normalization of an ideal lane

Results of the case that the width of an ideal lane is changed through an experiment is shown (Fig. 3).

According to experimental results, when the ideal lane width is normalized to 50 pixels, the lane width became too small to be detected, and when normalized to 150



a. Input image

b. When the width of an ideal lane is 50

c. When the width of an ideal lane is 150

d. Image that used the ideal lane normalized to 100 pixels

Fig. 3. Results in accordance with the ideal lane width

pixels, blurring effect remained. When the ideal lane width is fixed to about 100 pixels, the spreading of lane information existing at the image upper part could most be effectively reduced and the best result of obstacle detection was obtained. Therefore, in this paper, the ideal lane width was normalized to 100 pixels.

2.2 Set-Up of an Obstacle Candidate Region Using Dispersion Map

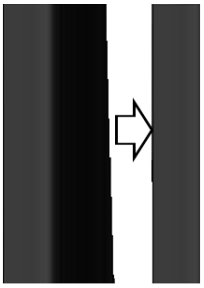
In this paper, location of lane is detected using inverse perspective transformation [5] which needs no camera parameter, inside the detected lane is set up as interested-region.

In general, when there is no obstacle, the road in front of a car is comprised of asphalt and lanes. The road image in the event of no obstacle does not show large change overall in pixels in the region excluding lanes. On the contrary, when obstacles exist, the pixel value on the road is distributed constantly on the asphalt part and is different from the asphalt pixel at the place where an obstacle exists. Hence, the dispersion value in the event of obstacles can be considered higher than the dispersion in the event of no obstacle. Out of road images with no obstacle on high ways and local roads, the average of dispersion values of 15,000 frames is about 15 and this value was set up as the maximum permitted value in the dispersion.

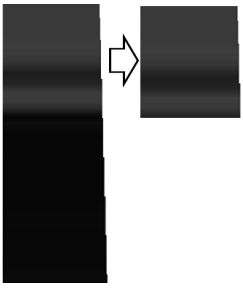
In the event that obstacles exist in the interested region, the road dispersion will be higher than 15, the maximum permitted dispersion value. Using this result, dispersion map is created with respect to horizontal and vertical direction in the interested

region. The next (Fig. 4) shows the process of making dispersion map using horizontal and vertical dispersion values in the interested region.

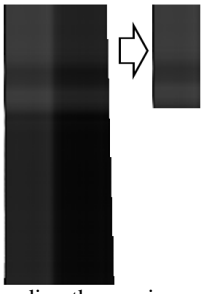
a. Interested region



b. The part exceeding the maximum permitted dispersion in the vertical dispersion map



c. The part exceeding the maximum permitted dispersion in the horizontal dispersion map

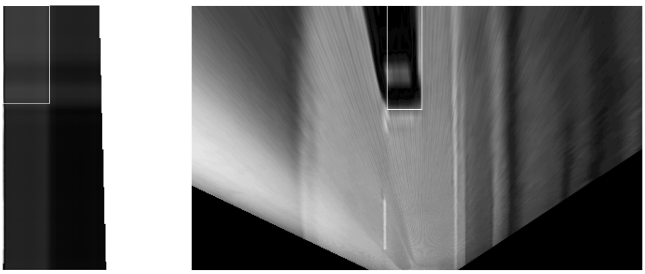


d. The common part exceeding the maximum permitted dispersion in the horizontal and vertical dispersion maps

Fig. 4. Formation of dispersion map

When obstacles exist, the obstacle region has higher values in the dispersion map and looks relatively bright, as in (Fig. 4.d). If locations that have road dispersion values exceeding the maximum permitted dispersion are discovered using this, it is possible to obtain obstacle candidate regions. The method to find a obstacle candidate region is the following algorithm 1 and the result is shown below (Fig. 5).

[Algorithm 1]
① In the horizontal and vertical dispersion maps, regions where the dispersion value exceeds the road dispersion maximum permitted value are extracted.
② Out of the regions that were extracted from the horizontal and vertical dispersion maps, common regions are set up as obstacle candidate regions.



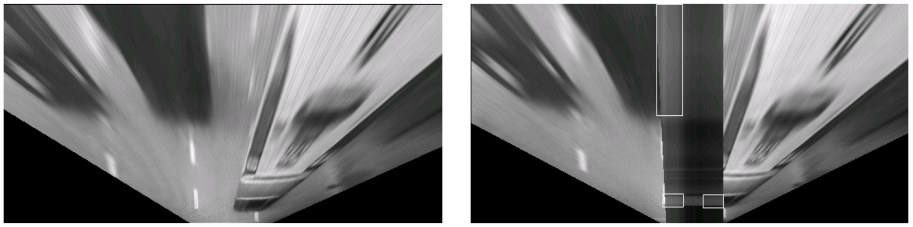
a. Common region in the horizontal and vertical dispersion maps

b. Obstacle candidate region

Fig. 5. Detection of obstacle candidate regions

2.3 Obstacle Detection Using Characteristic Points

In the event of one obstacle, dispersion map gives a good result. But, in the event of multiple obstacles, it is difficult to determine accurate positions of obstacles, because locations of obstacle candidate regions appear to be even in regions of no obstacle as in below (Fig. 6) when obstacle regions are detected by using dispersion map alone.



a. Input image

b. Obstacle candidate region

Fig. 6. Problems when there are multiple obstacles

This is a general problem that occurs in calculating dispersion, and should be interpreted as that there are multiple obstacle candidates. To solve this problem, the following algorithm 2 is proposed.

[Algorithm 2]
① Characteristic points are detected in the interested region using a FAST corner detector.
② When a characteristic point overlaps with the location of an obstacle candidate region detected by algorithm 1, it is interpreted as an obstacle.

In this paper, characteristic points were detected using a FAST corner detector [8] which is able to detect characteristic points and very accurate, and the result is shown below (Fig. 7).

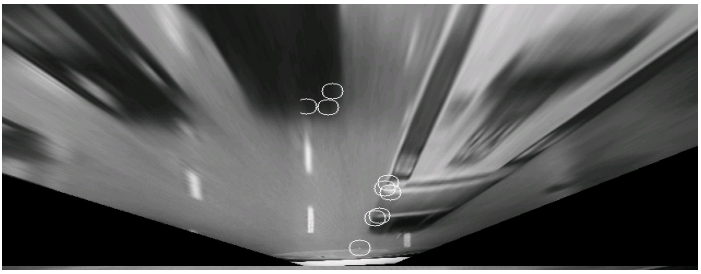


Fig. 7. Detection of characteristic points in an interested region

Locations of detected characteristic points are mostly concentrated on obstacle inside the lane and may be distributed even in other region because of noise effect in image. In the case that locations of characteristic points are distributed in the obstacle candidate region, the candidate region is set up as an obstacle region, and the result is shown below (Fig. 8).

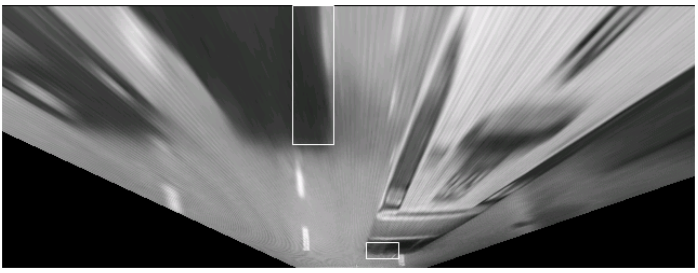
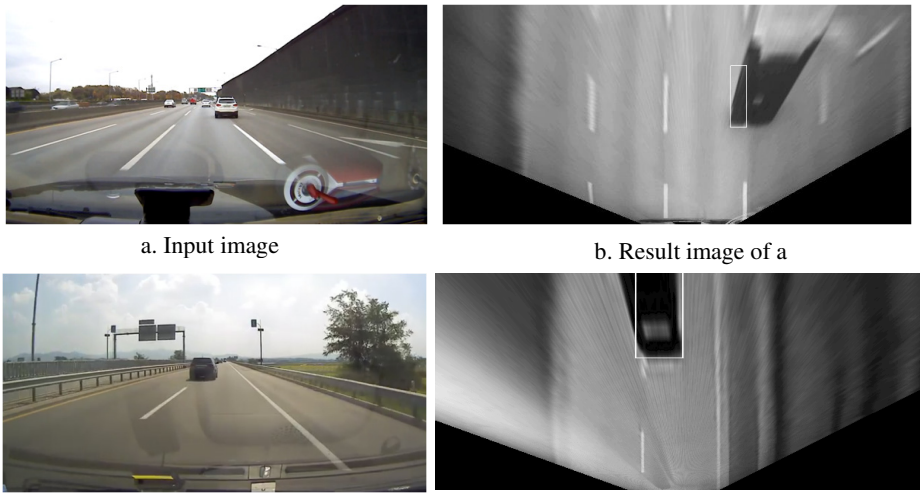


Fig. 8. Obstacle region detection

3 Experiment Result

The proposed method was tested in an experiment conducted on the high way and local road images, using input images obtained from smart-phone fixed on the smart-phone holder installed in the car front window.



a. Input image b. Result image of a

c. Input image d. Result image of c

Fig. 9. Result of obstacle detection

In this paper, after inverse perspective transformation was applied and the lane was detected, the region inside lane was designated as an interested region and, using this, the obstacle in the car forward direction could be detected.

The method proposed in this paper was developed as OpenCV for convenience, and the smart-phone used in experiment was SHV-E250L of Samsung Electronics. The image size was 640*480 and the processing speed per frame was 80~90ms. In <Table 1>, algorithm that detects an automobile using edge information under the environment of using a single camera instead of smart-phone [6] and algorithm based on movement information [7] were compared with the method proposed in this paper in terms of processing speed, detection rate, and image resolution.

Table 1. Comparison of previous methods and proposed method

	Method 1[6]	Method 2[7]	Proposed Method
Processing Speed	35ms	30ms	80~90ms
Detection Rate	90%	94%	95%
Resolution	640*480	320*240	640*480

4 Conclusion and Future Research Directions

This paper proposed a method to detect obstacles in the car front lane using a smartphone. Under the smart-phone environment, inside of lanes detected by inverse perspective transformation with no camera parameter is set up as an interested region, obstacle candidate regions are detected using dispersion map of the interested region, characteristic points are extracted from the interested region using a FAST corner detector, and out of candidate regions the region where an obstacle exists could be detected. The proposed method uses only the region inside lane to detect cars in the front direction, and it was possible to improve the processing speed by simplifying obstacle extraction algorithm.

In future researches, not only the method to detect obstacles only in the car front direction, but also the method to detect obstacles in the region outside lane will be

References

1. ETRI Industry Analysis Study Team Senior Researcher. Trend of Intelligent Automobile Safe Driving Technology, National IT Industry Promotion Agency IT Planning Series (August 2012)
2. Claudio, R.J.: A Lane Departure Warning System Using Lateral Offset with Uncalibrated Camera. In: IEEE Conference on Intelligent Transportation Systems, pp. 348–353 (September 2005)
3. Korea Electronics Technology Institute. Automobile Safe Driving System Industry Trend (December 2010)
4. Bertozzi, M., Broggi, A.: Real-time lane and obstacle detection on the GOLD system. In: Proceedings of the 1996 IEEE Intelligent Vehicles Symposium. IEEE (1996)
5. Heo, H., Han, G.-T.: A Robust Real-Time Lane Detection for Sloping Roads. KIPS Tr. Software and Data Eng. 2(6) (June 2013)
6. Baek, Y.-M., Lee, G.-G., Kim, W.-Y.: Nearby Vehicle Detection in the Adjacent Lane using In-vehicle Front View Camera. Journal of Korea Multimedia Society 15(8), 996–1003 (2012)
7. Kim, G., Cho, J.-S.: Vision-based vehicle detection and inter-vehicle distance estimation. In: 2012 12th International Conference on Control, Automation and Systems (ICCAS). IEEE (2012)
8. Rosten, E., Drummond, T.W.: Machine learning for high-speed corner detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 430–443. Springer, Heidelberg (2006)

Improved Depth Map Generation Using Motion Vector and the Vanishing Point from a Moving Camera Monocular Image*

Su-Min Jung and Taeg-Keun WhangBo

Gachon University, 1342 Seongnamdaero, Sujeoung-gu,
Seoung-si, Gyeonggi-do, Korea

Abstract. As a clue often used to acquire 3D information from a monocular image, there is a vanishing point. This article compares the inclines among valid straight lines at searching for the vanishing point of a monocular image and calculates the proximity degree and eliminates unnecessary information in generating a node so as to elevate accuracy to estimate the vanishing point of Hough Transform and draw an initial depth-map through the vanishing point calculated. Moreover, to obtain a more accurate depth of the initial depth-map, the study uses the difference of vector values according to the distance between the camera and the object based on the motion vector information within the monocular image with a camera moving and suggests a method to calculate a more accurate depth of the monocular image.

Keywords: Vanishing point, Motion Vector, Depth-map.

1 Introduction

Multimedia contents technologies that are in the spotlight recently, have been changing from letter-based to graphics and 3D images, which is more realistic and 3-dimensional, and these changes are based on human visual-recognition function. Many researches that are trying to express 3D images escaping from former 2-dimensional images are in progress worldwide, and researches trying to express natural scenery, animation and virtual reality space in the 3-dimensional image are receiving attention, as 3D television gains high interest. Under the influence of 3D technology, informatization, and infrastructure, people are consuming much information and contents, with visual media, such as movies and dramas, gaining the highest demand. Thus, demand for 3D contents which are former media with enhanced visualization, is rising explosively. However, 3D contents is very insufficient quantitatively compared to demand.

* This research is supported by Ministry of Culture, Sports and Tourism(MCST) and Korea Creative Content Agency(KOCCA) in the Culture Technology(CT) Research & Development Program [R2012030006].

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_92,

For this reason, recently in both Korea and foreign countries, researches trying to satisfy the consumer demand by converting 2D contents to 3D contents are in progress. However, as these researches demand manual labor, they have the disadvantage of taking a lot of time and money. Therefore, study on depth map generation with minimum manual labor while using monocular image is desperately needed.

The purpose of this paper is to calculate depth information for each area of one image frame that was filmed through motion of camera, in monocular images, not multi or stereo images. Algorithm suggested in this paper processes the motion estimation process in real time and introduces the full motion estimation method suitable for hardware realization, to generate relative depth information based on movement in the 2D monocular image series with camera movement. In addition, through researches on the acquisition method of vanishing point [1][2], where line and another line meets when an extension line is drawn to an object in image, straight line Hough-Transform extraction [3] which is usually used to search vanishing point was improved. The goal was to obtain additional information, through this, by generating the entire depth map of the image before using motion vector.

2 Suggested Method of Depth Map Creation

In this paper, after searching for vanishing point using the improved Hough-Transform, a rough depth map is drawn based on this, and relative depth for each pixel is obtained using motion vector in the image with camera movement. This paper suggest a method that completes the overall depth map by giving inequivalent values to relative depth values through the information of vanishing point.

2.1 Suggested Vanishing Point Search Method

To generate a depth map of 3D information using geometric characteristics of a monocular image, this paper employs the method that uses vanishing point [4][5], which is a typical method of estimating depth, one of geometric properties.

This paper tries to estimate location of the vanishing point in image more accurately, by using Hough-Transform efficient in straight line detection. The suggested vanishing point estimating algorithm uses Hough-Transform to generate noise within acquired effective straight lines, and uses removal of vertical and horizontal lines within effective straight lines, removal of effective straight lines of same slope, and intersection-point prevention between super close straight lines to eliminate unnecessary straight lines that hinder estimation of vanishing point from cluster of intersection points.

System structure map is shown in Fig. 1.

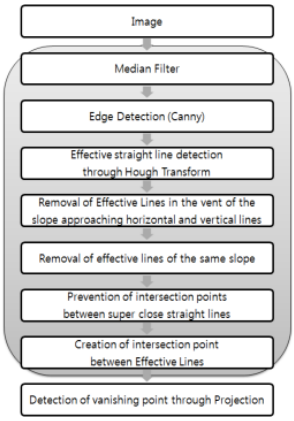


Fig. 1. System structure map

2.1.1 Removal of Vertical and Horizontal Lines

To obtain the vanishing point using straight lines acquired through Hough-Transform in a monocular image, straight lines unnecessary in vanishing point estimation are deleted. The first method is to delete vertical and horizontal lines that are unrelated to the vanishing point in an ordinary image.

$$l_{i\theta} = \begin{cases} (0+\epsilon)^\circ < l_{i\theta} < (90-\epsilon)^\circ \rightarrow a = true \\ else \qquad\qquad\qquad\quad \rightarrow a = false \end{cases} \tag{1}$$

$li\theta$ of (Eq. 1) stands for the slope of the ith straight line among the effective straight lines acquired through Hough-Transform. Formation of intersection point is prevented by deleting the rest straight lines except for effective straight lines in the slope range between (0+ϵ)° and (90-ϵ)° which are not vertical or horizontal lines.

Fig.2 shows the image of effective straight lines obtained from the original image, and the image after removal of vertical and horizontal lines.

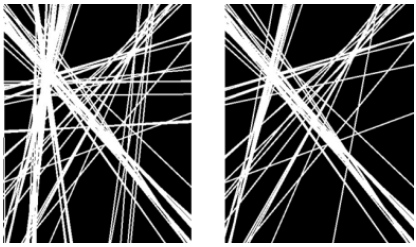


Fig. 2. Before deleting vertical and horizontal lines (left), after deleting (right)

2.1.2 Prevention of Intersection Point between Super Close Straight Lines

As in Fig. 3, in the case where straight lines heading to the vanishing point are very close and have very similar slopes, many intersection points between straight lines are created and the cluster degree in the no vanishing point area is increased creating error in estimating vanishing point.

Therefore, formation of intersection point is prevented when the slope and distance between straight lines are below certain thresholds. However, as the distance between two unparalleled lines cannot be calculated, only the slope difference is used as shown in (Eq. 2) but the accuracy of intersection point cluster is increased by adding sample number.

$$\begin{cases} |l_{i\theta} - l_{j\theta}| \geq threshold \rightarrow b = true \\ |l_{i\theta} - l_{j\theta}| < threshold \rightarrow b = false \end{cases} \tag{3}$$

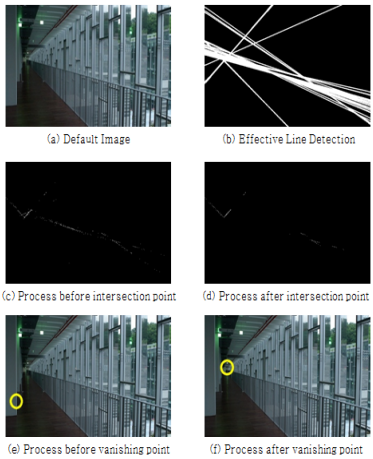


Fig. 3. Prevention of intersection point in super close straight lines

When the slope difference between two straight lines, $li\theta$, $lj\theta$ is below threshold (b=false), intersection point is not created.

2.1.3 Estimation of Vanishing Point from the Intersection Point

Vanishing point is obtained in the intersection point image resulting from prevention of intersection point and removal of straight lines that hinder intersection point cluster from effective straight lines obtained through Hough-Transform. The vanishing point is estimated through the column and row projection method.

Intersection point image is a binary image, and an intersection point has a value of 1. And, after calculating cumulative values of each row and column in the x,y-axis, as in Fig. 4, the x,y-coordinates corresponding to the maximum value are searched and corresponded to the vanishing point (i, j).

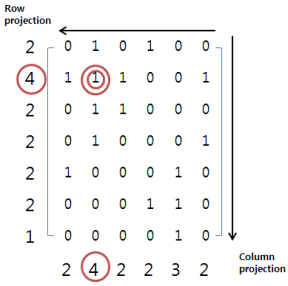


Fig. 4. x, y-axis projection method

2.2 Suggested Method of Depth Map Generation Using Motion Vector

The corresponding algorithm is based on the assumption that parallel camera movement gives equal influence to inside of the image, but motion of each area inside image frame with respect to camera's direction and degree of movement is not constant depending on the distance to camera.

In this paper, after preprocessing with respect to camera movement image, the motion vector value of each pixel is calculated through Optical Flow, and relative depth is calculated, by applying each different algorithm depending on occurrence of camera's back-and-forth movement, to complete the depth map.

The corresponding algorithm flow chart is shown in Fig. 5.

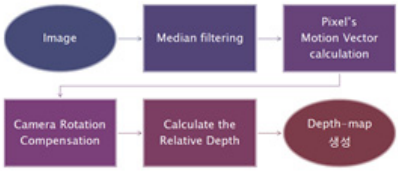


Fig. 5. Algorithm flow chart

2.2. 1 Calculation of Motion Vector for Each Pixel

To calculate motion vector for each pixel, the Pyramid Lucas-Kanade and Horn-Schunck Optical Flow algorithm is used[6][7].

Optical flow equation gives brightness change by movement and gives the same brightness value for the space-time movement trajectory of the same object. If x_1 and x_2 are pixels of the movement trajectory and $s_c(x_1, x_2, t)$ is the brightness distribution in the continuous space-time, brightness is constant on the movement trajectory, and the brightness change of the movement trajectory is given in (Eq. 3)

$$\frac{ds_c(x_1, x_2, t)}{dt} = 0 \tag{3}$$

If $v_1 = dx_1/dt$ and $v_2 = x_2/dt$ are velocity vectors in the continuous spatial coordinates, changes of pixels located on the movement trajectory and motion vectors satisfy (Eq. 4).

$$\frac{\partial s_c(\mathbf{x};t)}{\partial x_1} v_1(\mathbf{x},t) + \frac{\partial s_c(\mathbf{x};t)}{\partial x_2} v_2(\mathbf{x},t) + \frac{\partial s_c(\mathbf{x};t)}{\partial t} = 0 \tag{4}$$

The experiment image assumes only parallel movement without back-and-forth movement.

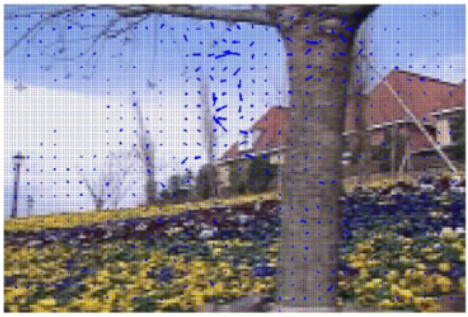


Fig. 6. Calculation of motion vector of each pixel

2.2. 2 Calculation of Relative Depth Information Using Motion Vector Value

Motion vector (v_x, v_y) for each block is calculated from motion estimation, camera movement components T_x, T_y, T_z and f that have constant values inside the frame are global variables, and Z, x, y, v_x, v_y are local variables that change according to the pixel location. In (Eq. 5), depth information can be easily acquired when the camera movement components and motion vector are known.

$$Z = \frac{1}{v_x}(-fT_X + xT_Z) + T_Z = \frac{1}{v_y}(-fT_Y + yT_Z) + T_Z \tag{5}$$

As global variables T_x, T_y, T_z are relative value with respect to f, exact actual values of f, T_x, T_y, T_z cannot be known. That is, for each pixel inside the frame, $\frac{fT_x}{Z}, \frac{fT_y}{Z}, \frac{T_z}{Z}$, the global variable rate for Z, can only be calculated.

In image restoration study, f is generally assumed as 1 in many cases for equation simplification. Ratio of relative depths between different pixels is calculated as in (Eq. 6).

$$\frac{1}{v_x}\left(-\frac{fT_X}{Z} + x\frac{T_Z}{Z}\right) + \frac{T_Z}{Z} = \frac{1}{v_y}\left(-\frac{fT_Y}{Z} + y\frac{T_Z}{Z}\right) + \frac{T_Z}{Z} = 1 \tag{6}$$

When parameters expressing relative depth information are defined as $r_x = \frac{fT_x}{Z}$, $r_y = \frac{fT_y}{Z}$, $r_z = \frac{T_z}{Z}$, (Eq. 7) is acquired as

$$\frac{1}{v_x}(-r_x + xr_z) + r_z = \frac{1}{v_y}(-r_y + yr_z) + r_z = 1 \tag{7}$$

From (Eq. 6) and (Eq. 7), it is possible to see that the ratio of depth over camera movement has a value inversely proportional to motion vector (v_x, v_y).

In other words, it means that if close radius and long radius, as assumed in advance, have inconsistent motion vector value for camera movement and if motion vector can be obtained, its relative depth information can be obtained.

This paper calculates r_x as 0 for an image with no back-and-forth movement, and uses average values of r_x, r_y as the relative depth information.

3 Result and Future Research Directions

This paper suggested the method that estimates vanishing point more accurately by deleting straight lines that hinder intersection point cluster after using Hough-Transform which searches vanishing point to generate the depth map, an important clue in the 3D image spatial structure restoration, and the method that adds the relative depth information using motion vector inside the camera movement image.

Result for each stage of images applied to experiment in accordance with vanishing point estimation algorithm is shown in Fig. 7, 8, 9, 10.

(a), (b), (c), (d) of Fig. 7 shows each experiment image, respectively.

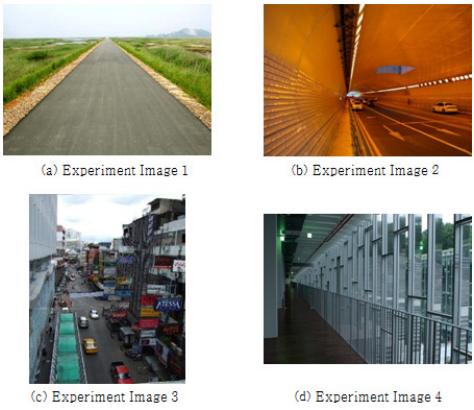


Fig. 7. Experiment images

(a), (b), (c), (d) of Fig. 8 shows the effective straight line extraction image of each experiment image.

(a), (b), (c), (d) of Fig. 9 shows the intersection point image of each experiment image. In the case of (d), even though formation of intersection point between super close straight lines was prevented, it is possible to verify that a large amount of intersection points were created in the right lower end.

(a), (b), (c), (d) of Fig. 10 shows the vanishing point estimation image for each experiment image.

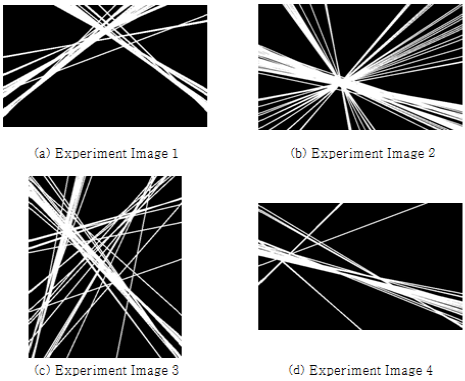


Fig. 8. Result of the effective straight line extraction after preprocess for each experiment image

Fig. 9. Intersection point extraction result for each experiment image

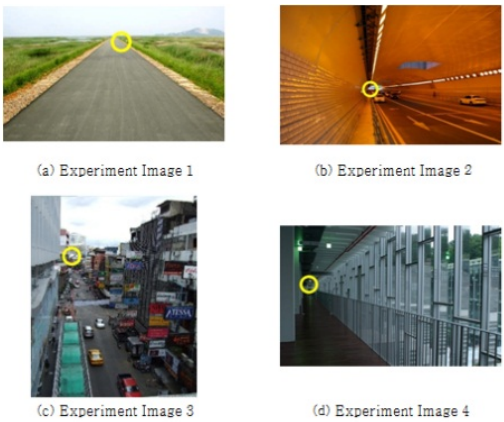


Fig. 10. Result of the vanishing point estimation for each experiment image

In the case of (a), the vanishing point was estimated to be leaned rightward compared to the expected vanishing point, in the vanishing point estimation stage. The reason is the error in Column Projection, which was caused by a limitation in the method.

In the case of (c), vanishing point estimation using only Hough-Transformation failed as various objects and straight lines having many vectors existed, but estimation succeeded by applying the corresponding algorithm.

According to the experiment results, vanishing point estimation showed big enhancement when the suggested algorithm was applied, with 127 out of 132 samples being able to be estimated.

However, there are drawbacks of not being able to find the vanishing point in images of two vanishing points, unclear image, and image without clear object. Also, cases of estimating location slightly different from the actual vanishing point occurred because of limitations of the vertical and horizontal Projection method that estimates the vanishing point in the intersection point image.

The experiment result of depth estimation using motion vector in camera motion images is shown in Fig. 11.

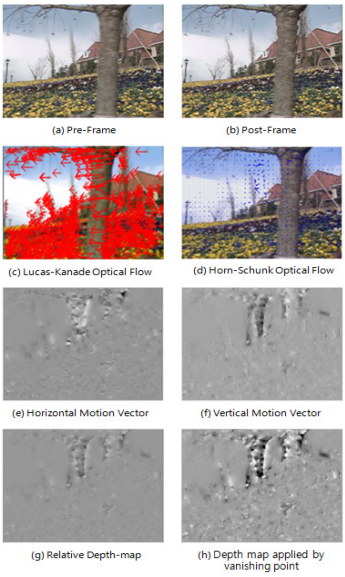


Fig. 11. Result of depth estimation using motion vector

(a), (b) are pre and post frame, respectively, used in the experiment, and are images where camera were moving rightward in parallel. (c), (d) show the estimated values of motion vector for each pixel using two methods. Each motion vector is calculated independently in the vertical and horizontal directions, as in (e), (f), and these values are used to make the relative depth map such as (g). Then, the final depth map (h) is completed by giving the linear weighted value for each pixel in accordance with the vanishing point location.

Future studies will include process of decreasing noise in calculating motion vector, acquisition of motion vector through Block Matching algorithm, and acquisition of inside-outside parameters in camera without calibration to solve non-linear problem depending on camera rotation.

Currently, the study on parameter acquisition using Fundamental Matrix in the process of SfM (Structure from Motion) is in progress. Through this, it is expected from the study in the future to obtain the relative depth information of image not only in the image with the camera parallel movement but also in the image with the camera rotation.

References

1. Kogecka, J., Zhang, W.: Efficient Computation of Vanishing Points. In: ICRA 2002, vol. 1, pp. 223–228 (2002)
2. Shufelt, J.A.: Performance Evaluation and Analysis of Vanishing Point Detection Techniques. In: Proc. ARPA Image Understanding Workshop, pp. 1,113-1,132 (1996)
3. Matas, J., Galambos, C., Kittler, J.: Progressive Probabilistic Hough Transform for Line Detection. In: Computer Vision and Pattern Recognition, CVPR (1999)
4. Ban, K.J., Kim, J.C., Kim, E.K.: An Object Representation System Using Virtual Space Coordinates. In: KIMICS, vol. 8(4) (August 2010)
5. Battiato, S., Curti, S., La Cascia, M., Scordato, E., Tortora, M.: Depth-Map Generation by Image Classification. In: Proceedings of SPIE Electronic Imaging 2004, Three-Dimensional Image Capture and Applications VI, San Jose, California, USA, vol. 5302-13 (2004)
6. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Proc. DARPA Image Understanding Workshop, pp. 121–130 (1981)
7. Horn, B.K.P., Schunck, B.G.: Determining optical flow. Artif. Intell. 23, 185-2-3 (1981)

Design and Implementation of Customized Encryption Platform for Data Security in Open Software Environment

Jae-Sung Shim[1] and Seok-Cheon Park[2,3,*]

[1] Department of Computer Science, Gachon University, Republic of Korea
11sjs28@naver.com
[2] Department of Computer Engineering, Gachon University, Republic of Korea
scpark@gachon.ac.kr
[3] College of IT, Gachon University,
Bokjeong-dong, Sujeong-gu, Seongnam-si, Gyeonggi-do, South Korea

Abstract. In this paper, data types were defined in an open software environment for the design of an encryption platform for open software data security. Then the most appropriate encryption algorithm for such defined data types was analyzed and a data-type-conversion module was designed for the use of such encryption algorithm module as a platform. In addition, modules wherein the user could select an encryption algorithm that was easy to use were designed, and an encryption platform was realized by integrating such modules.

In this paper, a motion compliance test of the designed and realized data encryption algorithm was performed, and the results confirmed that the designed algorithm was working precisely. In addition, the data-type-conversion module and the algorithm that selected the encryption algorithm were applied to the application and confirmed that they were working precisely.

Keywords: Encryption Platform, Data Security, Open Software, ECC.

1 Introduction

The open software market is rapidly growing due to the recent expansion of open software platforms, vitalization of the community, etc. However, there are frequent cases wherein the applications that were developed based on an open software environment were developed without considering the environment security. As such, security problems are continuously increasing. These problems were brought on by the fact that lots of time and resources should be invested to apply security to such applications [1,2].

Moreover, because it is hard to figure out the application types suitable for separate use with an encryption algorithm, developers are actually finding it difficult to apply security to such applications [3,4].

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_93,

Therefore, in this study, data types were analyzed in an open software environment for open software data security, each encryption algorithm that is suitable for use in each open software environment was investigated and developed separately, and an encryption platform was designed and realized to allow encryption of data when a developer develops an application in a particular environment. Furthermore, the detailed modules of each encryption algorithm and each data encryption platform were separately tested, and the tests confirmed that the modules of each encryption algorithm were working in normal conditions, which was verified by applying such encryption platforms to the actual applications.

This paper is organized as follows. The encryption algorithms and security platforms are introduced in Chapter 1 and then analyzed in Chapter 2, and the data types are defined and an encryption algorithm is designed in Chapter 3. A data encryption platform is realized and tested in an open software environment based on the design details in Chapter 4, and the conclusion and future development directions are presented in Chapter 5.

2 Related Studies

2.1 Encryption Algorithm

The term encryption algorithm refers, in the narrow sense, to an algorithm that is used when encoding a plain text and decoding such a coded message, and in the broad sense, to all the algorithms used in cryptography. Encryption algorithms are currently being popularized in services that use confidentiality, integrity and authentication, among the existing security services. Furthermore, they are being used for non-repudiation, besides the aforementioned three services [5,6].

When using an encryption algorithm, a key should be used to change a plain text to and from a coded message, and the keys are separated depending on their exposure or absence of it. Therefore, encryption algorithms are divided into private-key encryption algorithms and public-key encryption algorithms. The characteristics of the symmetric key algorithm and the public key encryption algorithm are analyzed in Table 1 [7].

2.2 JCA(Java Crypto Architecture)

JCA is an important part of a platform. It contains a series of APIs for digital signatures, structured into service providers; digesting messages; issuing and verifying certificates; encoding, generating and managing keys; generating safe random numbers; etc. Such APIs enable developers to simply integrate security into application codes. Their structure is designed based on the following guidelines.

- Independence of realization: A security service is available from the Java platform. The service providers are plugged into the Java platform through a standard interface. In some cases, the security functions of applications are dependent on multiple numbers of independent service providers.

Table 1. Comparison of the symmetric encryption algorithm and the public-key encryption algorithm

	Symmetric Key Encryption Algorithm	Public Key Encryption Algorithm
Basic Characteristics	Simple calculations such as XOR and MOD	Comprehensive mathematical equations such as prime factorization and oval curves
Relationship of the Encryption Keys	Encoding key = Decoding key	Encoding key ≠ Decoding key
Encoding Key	Confidential	Disclosed
Decoding Key	Confidential	Confidential
Encoding and Decoding Speed	Fast	Slow
Key Distribution	Difficult	Easy
Key Algorithms	DES, AES, SEED	RSA, ECC

- Interoperability of realization: This means that applications are not bound to a particular provider, and providers are likewise not bound to a particular application.
- Expandability of the algorithm: The Java platform includes a number of transfer providers who realize basic security services that are currently being popularized on a large scale. The Java platform supports the installation of the custom provider that realizes services.

3 Design of the Data Encryption Platform in an Open Software Environment

3.1 Construction of the Platform System

The proposed platform divides the data in an open software environment into three types: the terminal storage type, the normal transmission type and the real-time transmission type. It can encode and decode appropriately to data types by designing in DES, RSA and ECC algorithms depending on each data type. Ordinary developers use this platform to apply data security to the open software applications that they are developing. They do not directly include the source codes but use them for the platform type.

3.2 Data-Type-Conversion Design in an Open Software Environment

The data-type-conversion use-case diagram based on an open software program is constructed as shown in Figure 2. The diagram consists of the user and four use-cases: string to array, array to string, string to big-integer, and big-integer to string.

In addition, the diagram was constructed to inform the flows between classes when a user uses the encryption platform, and consists of a collaboration diagram for encoding data-type conversion and a collaboration diagram for decoding data-type conversion. Figure 1 shows the collaboration diagram for encoding data-type conversion.

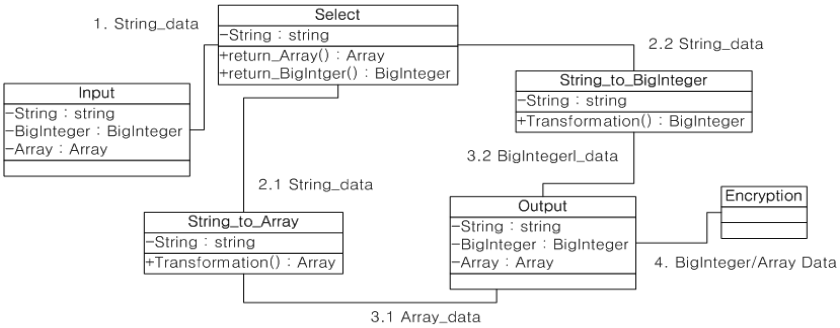


Fig. 1. Data-type-conversion encoding diagram

The ‘Input’ class sends the received data to the ‘Select’ class. The ‘Select’ class transmits the data to the class appropriate to the data types required for the encoding process. The DES encoding process sends the data to the 'String to Array’ class, and the RSA and ECC encoding processes, to the 'String to Big-integer’ class. The ‘String to Array' and the 'String to Big-integer’ classes convert the data into the data types required for the encoding process and send the converted data to the ‘Output’ class. The ‘Output’ class implements the encoding process by sending the data to the ‘Encryption' class that is appropriate for each encoding algorithm. Figure 2 shows the collaboration diagram for decoding data-type conversion.

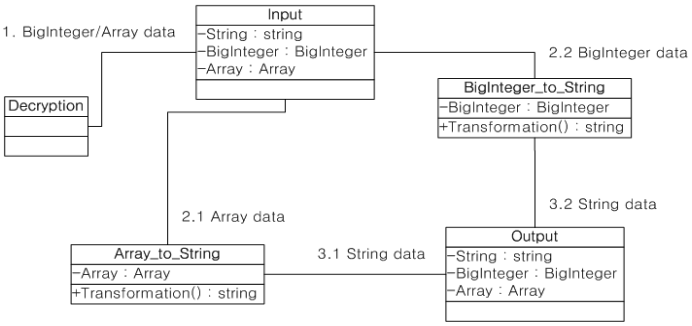


Fig. 2. Data-type-conversion decoding diagram

The ‘Input’ class selects the data (of the ‘Big-integer’ and ‘Array’ classes) processed through the decoding process for transmission to the decoding process as the original data of the ‘String’ class to the ‘Array to String’ class in the DES

decoding process, and to the ‘Big-integer to String’ class in the RSA and ECC decoding processes. The 'Array to String' and the 'Big-integer to String' classes convert the data into the String data type through the decoding process, and transmit the converted data to the 'Output' class.

4 Realization of the Data Encryption Platform in an Open Software Environment

4.1 A Block Diagram of the Data-Type-Conversion Module System in an Open Software Environment

The encoding process in the encryption platform can be described as follows. The ODEP performs data-type conversions of the variables received in the file and string types at the Int_Array data and the Big-integer data through the DT_ODEP (data transformation) module for encoding the data; the Int_Array data is transported to the DES_ODEP and the big-integer data, to the RSA_ODEP and ECC_ODEP modules; and those data are encoded respectively. The encoded data are stored at the terminal or transmitted to other terminals. Figure 3 shows the block diagram of the data encryption system.

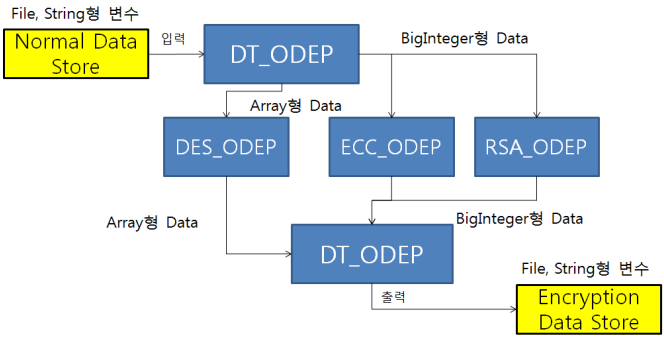


Fig. 3. Block diagram of the data encryption system

The decoding process in the encryption platform can be described as follows. The ODEP performs data-type conversions of the variables received in the file and string types at the Int_Array data and the big-integer data through the DT_ODEP (data transformation) module for decoding the data; the Int_Array data is transported to the DES_ODEP and the Big-integer data, to the RSA_ODEP and ECC_ODEP modules; and those data are decoded respectively. The data decoded to the original data are supplied as the data to be used by the user, i.e., to be presented on the output screen and/or saved. Figure 4 shows the block diagram of the decoding system.

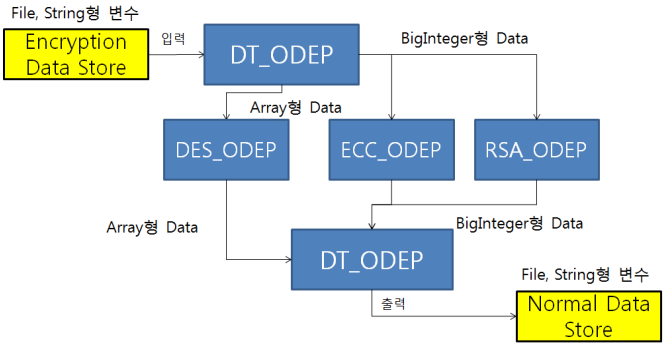


Fig. 4. Block diagram of the data decoding system

4.2 Execution Screen of the Data-Type-Conversion Module in an Open Software Environment

Figure 5 shows the execution screen of the data-type-conversion platform in an open software environment. The upper box shows that the data is converted from the string type to the big-integer type by the initialized values on top of the input 'Test123.' That is, the data type is changed to the big-integer coordinates (926026277631003472949 and 241877181190128230872115). The lower box shows that the aforementioned big-integer coordinate-type values were again converted to the data type from the original value of the Test123 type.

Fig. 5. Execution screen of the data-type-conversion platform in an open software environment

4.3 Test and Evaluation of the Data Encryption Platform in an Open Software Environment

The test was performed after designating the test IDs, and the results were evaluated as 'passing' or 'failing.' The functions that belonged to the page were classified by module and tested independently, and it was verified if there were any malfunctions or errors in their functions and if the output values were produced in normal

conditions according to the input values. At this time, the expected values and the resulting values from some tests were compared to verify if the outputs were normally processed.

The normal operations of each platform were investigated through this test process. The test results on the data-type-conversion module for encoding and decoding are described in Table 3.

Table 2. Data-type-conversion module for encoding and decoding

No.	Test ID	Test Type	Result
1	ECC_En_DataChange	Data-type conversion for encoding ECC data (string -> big-integer)	Pass
2	ECC_Den_DataChange	Data-type conversion for decoding ECC data (big-integer -> string)	Pass
3	RSA_En_DataChange	Data-type conversion for encoding RSA data (string -> big-integer)	Pass
4	RSA_Den_DataChange	Data-type conversion for decoding RSA data (big-integer -> string)	Pass
5	DES_En_DataChange	Data-type conversion for encoding DES data (string -> big-integer)	Pass
6	DES_Den_DataChange	Data-type conversion for decoding DES data (big-integer -> string)	Pass

The test results on the encryption platform by function are presented in Table 4.

Table 3. Realization of the encryption platform

No.	Test ID	Test Type	Results
1	Know Application Information	Source codes with the identified application information	Pass
2	Know Encryption Algorithm	Source codes with the indentified encryption algorithms	Pass
3	Default in Mobile	Encryption algorithms with the default value suitable for mobile environments	Pass
4	App Application in Java	Java application for application information analysis	Pass
5	App Application in Android	Android application for application information analysis	Pass

5 Conclusion

The open software developments that recently became an issue are underway and focus on particular core open software such as cloud and Android. However, in the

case of security, UI, etc., minor parts were applied for the developments. Such problems can become more serious, such as exposure of personal information to damage, critical errors in programs, etc., as ordinary developers use open software applications without security when manufacturing their own open software applications.

Therefore, in this paper, the data types in an open software environment were identified as terminal-storage-type data, normal-transmission-type data and real-time transmission-type data, and an encryption algorithm that suits an open software environment was designed to enable the design of an encryption platform for data security in open software environments. In addition, a data-type-conversion module was designed to convert the data types to be used in each encryption algorithm module.

Furthermore, a Java realization environment was constructed for use in the Windows and Android operating systems to realize the data encryption platform in the open software environment designed in this paper. Also, a data-type-conversion module that converts data types to make them applicable to each encryption algorithm was realized to allow the use of the realized encryption algorithm module as a platform.

Accordingly, motion compliance tests of the detailed modules were performed to investigate the performance of the data encryption platform in the open software environment that was designed and realized in this paper, and other tests were performed while applying the realized data encryption platform to actual environments.

References

1. Jeon, Y.-S., Kim, T.-Y.: An Empirical Analysis on Open Source Software Promoting Factors. Korea Economic Research Institute 17(1), 151–181 (2008)
2. Kim, K.-Y., Kang, D.-H.: Smartphone Security Technology in an Open Mobile Environment. Korea Institute of Information Security and Cryptology Review 19(5), 21–28 (2009)
3. Lee, J.-H.: The Development of Smart Mobile Information Security. Korea Information Society Development Institute 22(13), 17–33 (2010)
4. Hwang, J.-Y., Choi, D.-W., Chung, Y.-H.: An Efficient Encryption Technique for Cloud-Computing in Mobile Environments. Journal of the Institute of Signal Processing and Systems 12(4), 298–302 (2011)
5. Yu, Y.-G.: Current Situation of Technique in Digital Encryption. Electronic Information Center (2006)
6. Kwak, M.-S.: Technical security in the protection of the Information (Encryption Techniques). Seoul National University College of Medicine (2005)
7. Kim, J.-H., Chae, C.-J., Choi, B.-S., Lee, J.-K.: Study on secure system on wireless internet. Korean Institute of Information Technology: Summer Conference, pp. 173–177 (2006)

An Effective Adoption of Disparity to Enhance Recognizing Three-Dimension Facial Expression

Kwangmu Shin and Kidong Chung

Dept. of Computer Engineering, Pusan National University,
Busan, Republic of Korea
{sin,kdchung}@pusan.ac.kr

Abstract. The facial expression recognition can be highly useful in various fields. two-dimensional information based approach don't work robustly in the unpredicted influences of illumination, head pose and posture. In this paper, we analyzed the subtle change of facial expression from the perspective of disparity in stereo vision system. This disparity factor means three-dimensional information. Next, we present a effective and plain way to improve facial expression recognition rates. As a results, we were able to reach conclusion that this method can be effective role to improve recognition rates of the existing facial expression recognition system.

1 Introductions

The facial expression recognition system is able to classify kind of facial expression which source material, that is, facial expressions are acquired with regular criterion. To mention in detail, the facial expression recognition system requires a number of preprocessing steps which attempts to detect and locate characteristic facial regions, extract facial expression features, and model facial gestures using anatomic information about the face [1].

This system can be highly useful for computer games, on-line education, Human Computer Interaction (HCI), psychological studies, security, synthetic face animation, robotics, virtual reality and so on [2]. But most previous researches related with the facial expression recognition are based on two-dimensional information of intensity [3][4]. In two-dimensional information based approachs, features information just only exploits plain property of face without enough utilizing of knowledge of the depth information. These methods using two-dimensional information are computationally fast and simple but don't work robustly in the unpredicted influences of illumination, head pose and posture. In other words, they have different recognition rates according to external environments.

So recently researches to utilize efficiently three-dimensional information are being performed. They provide methods to overcome efficiently above-mentioned such limitations [5][6][7][8][9].

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_94,

There are realistic problems in three-dimensional model based facial expression recognition. These methods need additional three-dimensional modelling processes or scanning equipments. Those cause increase of cost or processes in facial expression recognition system. So we present method which can be applied to more generally environment in facial expression recognition. That is, economics, practicality and flexibility are considered. The flexibility means combination with other facial expression recognition systems. Finally, we analyze the subtle change of facial expression from the perspective of disparity in stereo vision system. This disparity points three-dimensional information considering non-plain property of face. And we present a way to improve facial expression recognition rates which it is based on analysis results of experiments.

Additionally, we were faced with a difficult problem in the verification of experimentation. Acquisition of verified facial expression database is much more difficult as compared to acquiring images. Because the facial expression database consists of stereo image. Recently, Yin et al. at Binghamton University have constructed a three-dimensional facial expression database and four-dimensional facial expression database for facial behavior research [10]. We use BU-4DFE database. But this database don't provides stereo image. So we extracted effectively stereo image in BU-4DFE database. Details of the database are presented in Section 4.

The main contribution of this work was to utilize three-dimensional disparity information of stereo vision which can improve effectively facial expression recognition rates.

This paper are organized as follows. Section 2 reviews briefly the three-dimensional model based facial expression recognition system and the disparity of stereo vision system. Analysis processes of subtle disparity change in facial expression is presented in Section 3. Experiments and analaysis results are presented in Section 4. Finally, concluding remarks and future works were descripted in Section 5.

2 Facial Expression Recognition and Stereo Vision

In this Section, we mention briefly to the three-dimensional model based facial expression recognition systems. For reference, three-dimensional information can be used in several different ways. Next, we explain the disparity of stereo vision system. The dispairty is important information which it is the basis of facial expression recognition in this paper.

2.1 The Facial Expression Recognition

Fig. 1 describes the general facial expression recognition system.

Firstly it acquires enough training images and testing images. Labelling feature points are perfomed in the both images and extract features. Generally labelling step is manual. Learning using PCA method etc. is perfomed in the training images. Finally, the facial expression recognition result is drawn through classifier using SVM method etc. with learned training images and testing images.

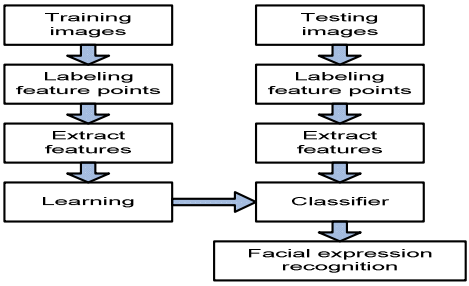


Fig. 1. The general facial expression recognition system

2.2 The Disparity of Stereo Vision

The distance between two cameras exists in stereo vision because of physical position difference of two cameras. A disparity is founded referencing a picture of other view. The stereo matching criterion to find disparity are Sum of Absolute Difference (SAD), Sum of Square Difference (SSD), MeanAbsolute Difference (MAD), Mean Square Difference (MSD) etc. In this paper, we use SAD approach and fixed block size.

Eq. (1) is SAD which it is used to extract disparity of stereo image.

$$M(p,q) = \sum_{i=1}^{n}\sum_{j=1}^{n}\left|I_c(i,j) - I_r(i+p, j+q)\right| \tag{1}$$

Given an n x n block, a matching criteria, M(p, q), measures the dissimilarity of a block in the current frame, Ic, and a block in the reference frame, Ir, shifted by (p, q). In other words, (p, q) is disparity of right images corresponding to the left image. Fig. 2 describes the geometric structure of stereo vision and the principle of stereo matching using two cameras.

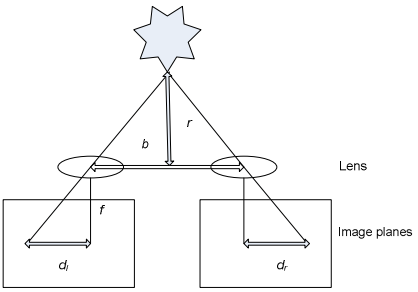


Fig. 2. The geometry of stereo vision

The details about Fig. 2 are as follows. b represents base line which is distance between two cameras. d represents disparity which it is dl - dr. f represents focal length. r represents distance between subject and camera. Eq. (2) can be made through the geometric structure of Fig. 2.

$$r = f\frac{b}{d} = f\frac{b}{d_l - d_r} \tag{2}$$

3 The Subtle Disparity Change

Fig. 3 describes the preprocesses required to analyze disparity map changes between two facial expressions.

In the first extraction step of stereo image, we could extract stereo image using MATLAB VRML tool. The viewpoint of stereo image is convergent with fixed angle. In the second extraction with fixed size, we could extract with fixed resolution of 200 x 250 pixels manually. The reason is more precisely to acquire disparity map. Additionally, The manual step of extraction does not matter because our ultimate goal isn't to automate face detection. And in the disparity map creation, we could acquire disparity map effectively using the public stereo matching MATLAB code of Wim Abbeloos. Finally, we could acquire disparity map changes between two facial expressions as each sub-region. The sub-regions consists of eyes, nose, mouth and total region. In other words, this means the absolute change of average pixel value as each sub-region between two disparity map. By default, is is compared with netural expression.

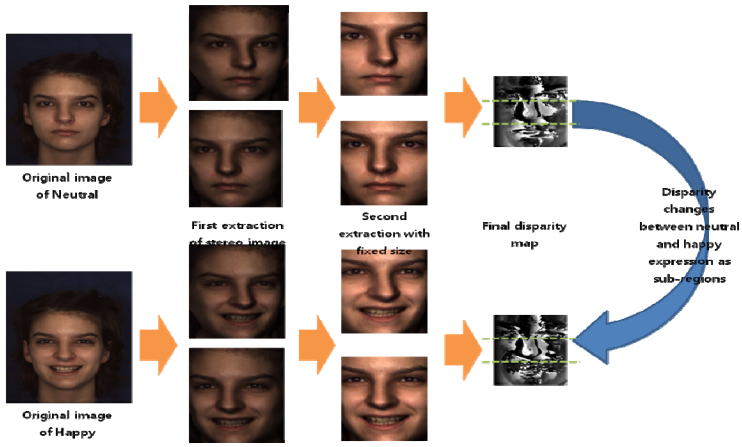
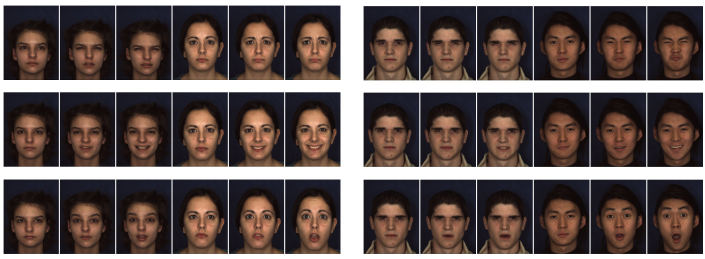


Fig. 3. The disparity map changes between two facial expressions

4 Experiments and Analysises

4.1 Facial Expression Database

Facial expression is naturally a dynamic facial behavior. The four-dimensional facial representation is believed to be the best reflection of this nature. So four-dimensional facial expression data captures the dynamics of time-varying three-dimensional facial surfaces, making it possible to analyze the dynamic facial behavior in a three-dimensional spatio-temporal domain. Fig. 4 describes the part of four-dimensional facial expression database. In this Figure, they show three types of expression and two types of intensity in facial expression.



(a) Female subjects (b) Male subjects

Fig. 4. The part of four-dimensional facial expression database. (Angry, Happy, Surprise)

And Table 1 describes the summary of four-dimensional facial expression database.

Table 1. The Summary of four-dimensional facial expression database

Items	Number
Subjects	101
Expressions	6
Model sequences	606
Texture videos	606

The resulting database consists of 58 female and 43 male subjects, with a variety of ethnic / racial ancestries, including Asian, Black, Hispanic / Latino and White.
In this paper, we use four-dimensional facial expression database to know more precisely the change of facial expression in terms of successive images.

4.2 Analysis Results

In this Sub-section, we described analysis results about facial expression. Table 2 describes The characteristics of the facial expression data used in the experiments.

Table 2. The characteristics of the facial expression data used in the experiment

Items	Number
Subjects	50
Expressions	3
Intensity	2

The number of subjects is ten which they consists female and male with same percentage. The types of facial expression are angry, happy and surprise (included netural expression). The intensity means the degree of facial expression. And we use only texture image. The existing four-dimensional facial expression database is considerably larger as mentioned in Sub-section 4.2. But we used the part of the database. Because the processes to extract additional images are very time-consuming. Counting totally collected images from the existing database, as follows.

{ 50 (subjects) x 3 (expressions) x 12 (extracted images with intensity) } + { 50 (subjects) x 3(expressions) x 3 (disparity map) } = 2250

The experimental processes for the analysis were discussed in Section 3 sufficiently. Table 3 describes the median values of disparity changes which is divided by each sub-region. In other words, it is simplified version about the entire experiments. We used only high intensity facial expression in this experiment.

Table 3. The median values of disparity changes

Area / Expression	Angry	Happy	Suprise
Mouth	7.5	12.5	17.5
Eyes	9	10	11.5
Total	11	11.7	12.15

Fig. 5 describes the flowchart for facial expression recognition enhancement using the disparity map changes. In the first step, we used the median value of mouth area to distinguish between angry and happy&surprise expression. And in this Figure, the threshold values can be changed with experiment environments.

As shown Fig. 5, we could find that this method can help to improve recognition rate of the existing facial expression recognition system although the experiment data is not enough and is not high recognition rates.

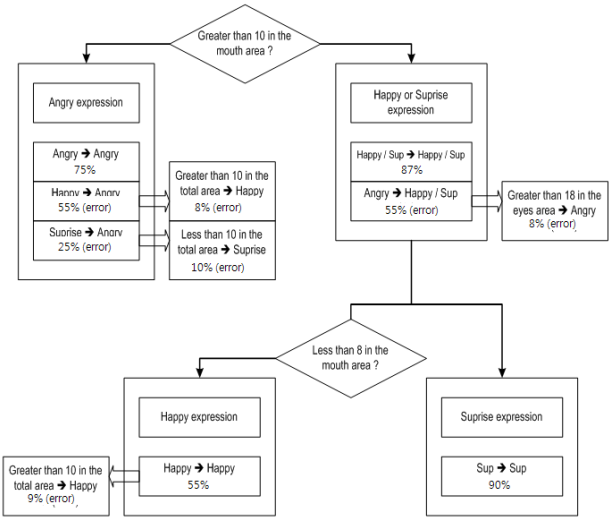


Fig. 5. The flowchart for facial expression recognition enhancement using the subtle disparity changes map

5 Conclusions

The facial expression recognition can be highly useful in various fields. But two-dimensional information based approach don't assure recognition rates robustly in the unpredicted influences of illumination, head pose and posture. In this paper, we analyzed the subtle change of facial expressions from the perspective of disparity in stereo vision system. Next, we presented a effective way to improve facial expression recognition rates which it is based on analysis results of experiments. As a results, we were able to find that this method can help to improve recognition rates of the existing facial expression recognition system. In the future, we will more precisely verify analysis results with more stereo images data. And recently we directly obtained image data by using stereo camera. The related experiments are underway. we are expecting that it is able to get similar results with this paper.

References

1. Wang, J., Yin, L., Wei, X., Sun, Y.: 3D Facial Expression Recognition Based on Primitive Surface Feature Distribution. In: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2006), New York (June 17-22, 2006)
2. Gong, B., Wang, Y., Liu, J., Tang, X.: Automatic Facial Expression Recognition on a Single 3D Face by Exploring Shape Deformation. In: ACM International Conference, MM 2009, Beijing (October 19-24, 2009)
3. Tian, Y., Kanade, T., Cohn, J.F.: Recognizing actions units for facial expression analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(2), 97–115 (2001)

4. Rosenblum, M., Yacoob, Y., Davis, L.S.: Human expression recognition from motion using a radial basis function network architecture. IEEE Transactions on Neural Networks 7(5), 1121–1138 (1996)
5. Mpiperis, I., Malassiotis, S., Petridis, V., Strintzis, M.: 3D facial expression recognition using swarm intelligence. In: IEEE International Conference of Acoustics, Speech and Signal Processing, ICASSP 2008, pp. 2133–2136 (2008)
6. Mpiperis, I., Malassiotis, S., Strintzis, M.: Bilinear Models for 3-D Face and Facial Expression Recognition. IEEE Transactions on Information Forensics and Security 3(3), 498–511 (2008)
7. Soyel, H., Demirel, H.: Facial expression recognition using 3d facial feature distances. In: Kamel, M.S., Campilho, A. (eds.) ICIAR 2007. LNCS, vol. 4633, pp. 831–838. Springer, Heidelberg (2007)
8. Tang, H., Huang, T.: 3D facial expression recognition based on automatically selected features. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 1–8 (2008)
9. Mpiperis, I., Malassiotis, S., Strintzis, M.: Bilinear Models for 3-D Face and Facial Expression Recognition. IEEE Transactions on Information Forensics and Security 3(3), 498–511 (2008)
10. Yin, L., Chen, X., Sun, Y., Worm, T., Reale, M.: A High-Resolution 3D Dynamic Facial Expression Database. In: The 8th International Conference on Automatic Face and Gesture Recognition (FGR 2008) (September 17-19, 2008)

Protecting Cloud-Based Home e-Healthcare with Cryptographic Scheme

Ndibanje Bruce[1], Hyun Ho Kim[1], Mangal Sain[2], and Hoon Jae Lee[2]

[1] Department of Ubiquitous IT, Graduate School of Dongseo University, Sasang-Gu, Busan 617-716, Korea
bruce.dongseo.korea@gmail.com, feei_@naver.com
[2] Division of Computer and Engineering Dongseo University, Sasang-Gu, Busan 617-716, Korea
mangalsain1@gmail.com, hjlee@dongseo.ac.kr

Abstract. The emergence of ubiquitous technologies gives rise to huge applications that enable the data accessibility anytime anywhere. Cloud-based home healthcare system is one of researching area of the cloud computing applications. As cloud computing allow the on-demand network to a shared pool of configurable computing resources, patients' data protection is the paramount requirement for the security and privacy to ensure the trustworthiness of the cloud-based home healthcare system. To this end, this paper proposes cryptographic scheme where a patient can encrypt his data before uploading the data to the cloud. To achieve this, we design and implement a healthcare monitoring system to collect patient data and send them to computer. In this experiment we use pulse sensor on arduino board for heart rate measurement. The data collected from the patient can be uploaded to cloud after encryption operations. Only the user with shared private key can decrypt the patient data. Thus, the patient would trust in the cloud infrastructure that supports critical applications for the healthcare system. In addition, the analysis of the proposed scheme ensure the honesty of the cloud provider, since the patient has the ability to control who has access to his data by issuing a cryptographic access credential to data users.

Keywords: cloud, healthcare, cryptographic, credential.

1 Introduction

The cloud offers the potential of easy access to electronic medical records in a medical setting. Quick access to a person‘s medical history could speed up treatment, help to avoid complications, and even saves lives. In addition, the cloud could make it easier for the patients to locate and keep track of their own medical history. However, on the other hand, patient also wants privacy and guarantees that their health information is secure. According to the Health insurance Portability and Accountability Act (HIPPA) regulation, the providers of IT services should first get the trust from consumers and minimize all areas of risks then eHealth cloud can be totally deployed [1].

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*, Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_95,

Different suggestions to accomplish privacy and access control in eHealth have been published [2-4]. Moreover, assurance of privacy and address privacy issues must be provided by the eHealth systems at different system levels: architectural design, access control, communication protocols, etc. Thus, it is commonly achieved in practice by means of a form of access control or authentication [5-9]. However, typical eHealth systems, especially in future, will be highly distributed and require interoperability of many subsystems. Even if health-care data is well protected and access control is perfectly employed, improperly designed communication protocols for such interoperability will cause information leakage and hence breach users' privacy. So far, security and privacy of communication protocols in eHealth systems is seldom studied in the literature.

The problem addressed in this paper is the confidentiality or trust of cloud providers by the customer. This paper considers a case of patients who wants to upload his data to cloud but because of untrustworthiness, we propose a way where the patient can encrypt his data before uploading it to cloud. Hence, the patient would trust in the cloud infrastructure that supports critical applications for the healthcare system.

The remainder of this paper is organized as follow; Section 2 illustrates the related work while. We develop our method in Section 3 and security analysis is given in Section 4 before concluding in Section 5.

2 Related Work

In order to ensure the security of e-health systems, several different schemes have been proposed. In the following, we provide an overview of some existing security implementations for e-health systems. The patient centred access to secure systems online has been presented in [10]. The authors claim that, initially; it aims to permit patients and health care providers to access health information, even the sensitive data. Their access scheme combines role-based access control, mandatory access control and discretionary access control. The implementation is a patient-centred and centralized approach that stores all the data on a single server.

Different countries have developed and implemented their e-Healthcare System such as electronic Health Card (eHC) system [11-12] in German where each patient has an eHC smartcard as described in the compulsory health insurance system. The main function of eHC is to store the administrative data (for billing with the health insurance), with embedded encryption operations of patient’s records to be saved on HER servers and to give access rights when the data is needed.

Figure 1 shows an advanced model, which not only involves more parties (e.g., health insurances), but also includes some technical means to enforce data security and privacy of EHRs.

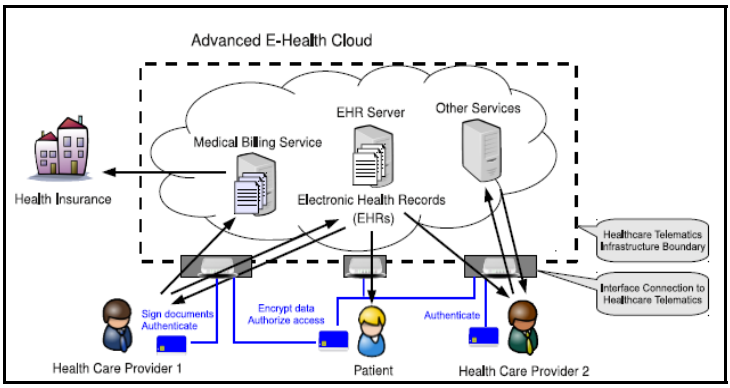


Fig. 1. Advanced E-Health Cloud model

3 Data Collection and Cryptographic Techniques

This section describes the main work of our paper. Details are sketched to show how we designed and implemented a system to collect data from patient and store the data to computer using sensor pulse, arduino board and wireless sensor node and finally how a patient can encrypt this data before upload it to cloud.

3.1 Data Collection Design for eHealthcare System

This subsection describes the architectural design and implementation of the healthcare data collection system using the sensor pulse to measure the heart rate, arduino sensor node to support the wireless technology. Figure 2 shows the process of the pulse sensor set up operations. Partially, we recall the work we did in our previous research [13] and incorporate the cryptographic scheme for the reason that we apply our Healthcare Monitoring Application to cloud computing area. To send the data

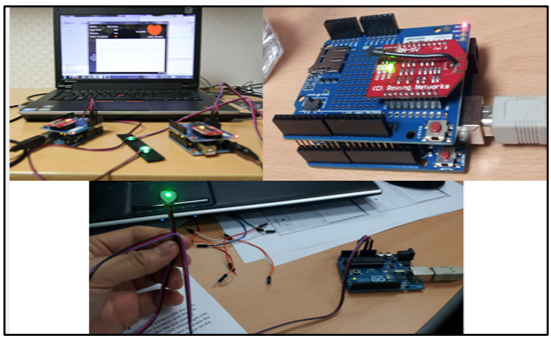


Fig. 2. Pulse sensor and Arduino set up operations

wirelessly, the sensor pulse is connected to the wireless sensor node which uses RN-XV module base on 802.11x protocol micro controller made by Roving Networks. The sensor pulse is first connected to the arduino and then all the set are connected to the computer via an usb cable. In this paper, we focus on the pulse sensor which collects raw data of the heart rate and transmits received data through Wi-Fi connection to the computer.

When the sensor is sensing the heart rate, it sends raw data without any operations of analyzing. To make the raw data understandable by doctors or nurses, we developed software which interpret the raw data to a graph platform. The description of the algorithm is given in Figure 3. It has been observed that changes in heart rate occur before, during, or following behavior such as posture changes, walking and running. Therefore, it is often very important to record heart rate along with posture and behavior, for continuously monitoring a patient's cardiovascular regulatory system during their daily life activity. The algorithm of HR analysis with activity monitoring is shown in Figure 3. Analysis of HR data is done on server for HR status and activity monitoring of patient. After calculating HR parameters, the algorithm goes for their classification for activities monitoring and HR status. If the heart rate is between 60 and 110 bpm and the patient is in rest position then can classify as a normal condition.

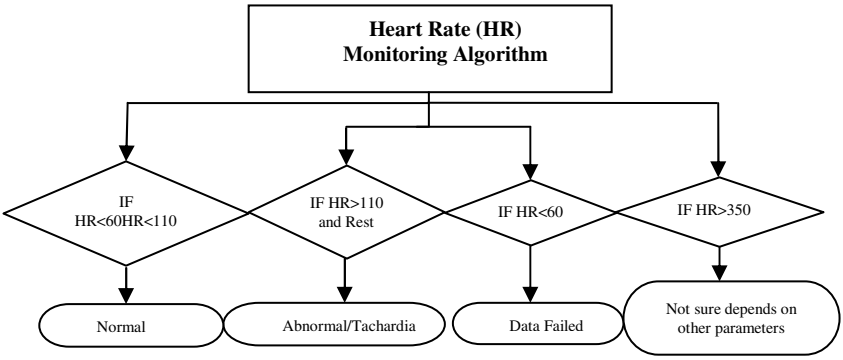


Fig. 3. Analysis of Heart Rate

If heart rate is greater than 110 bpm and patient position is rest then can classify for abnormal. For more precision, abnormal or regular HR, while resting if the HR is between 60 and 100 bpm the person is an adult otherwise if HR goes to 110 he is a baby. Therefore, it is necessary to know about the patient activities during measurement of HR parameter. During moving activity of the person, an HR analysis depends on other parameters also such as blood pressure, temp etc.

The measurement of the heart rate is done by a sensor pulse connected to Arduino as interface between computer and the sensor pulse. And the wireless communication is done by the wireless module connected to arduino wireless shield. The sensor pulse sends raw data and the HR algorithm translates the raw data using HR calculating the code. The result of the HR is given in a graph platform where the doctor or nurse can

easily ready it. Figure 4 gives the result of the HR with 2 scenarios where the HR is between 60 and 110 for the first and HR is below 60 for the second.

For the left graphic, the patient doesn't present any problem from the status which is normal we can see that the HR is 97, and then the alarm condition is green which means no disease. On the right side the status is "fail" because the HR is under 60 in this case the doctor should take further decision for the patient.

Fig. 4. HR result with normal and fail status

3.2 Cryptographic Design for Data Security and Privacy

This subsection presents the method to encrypt the patient's data before sending the data to cloud. We use the data obtained using the method to collect data illustrated in subsection A. Following the semi-trusted situation between consumer and cloud provider, we believe that the cloud providers are honest but curious. Thus any attacker can (un) intentionally initiates a misuse case. Considering those untrustworthily situation, in order to establish trust in the cloud infrastructure to support critical applications (such as healthcare systems), security and privacy should be built in to assure trustworthiness. To do so, cryptographic techniques to build in security and privacy are needed as regards the specific security and privacy challenges for the home healthcare system in the cloud. Figure 5 is an overview of an eHealthcare Information System where a patient can encrypts his HR result and then uploads the ciphertext to cloud, after an authorized person can decrypt the data after downloading.

The patient using his secret key K encrypts his HR result, let us call $<M>$ the data obtained. From his from his device he uploads the ciphertext to the cloud.

$$E_K(M) = C(M) \tag{1}$$

The content of the ciphertext $C(M)$ such as attributes, shared secrets keys and accessibility conditions will define the rights of the authorized person to decrypt the data. If any attacker can access the encrypted data from the cloud he will not be able to read the content because restrictions access embedded into the cipher. Furthermore, shared secret keys are required to access the data.

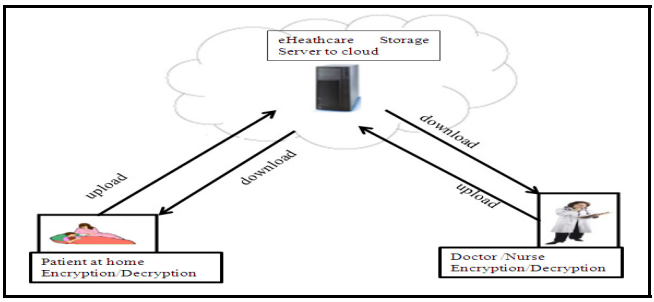


Fig. 5. Encryption and Decryption processes

Ciphertext contents: patient and Doctor attributes. The patient attributes are:Nonce_patient (*Np*), Patient_ID (*Idp*), and TimeValid_patient (*Tp*).The Doctor attributes are: Nonce_Doctor *(NDr)*, Doctor_ID *(IdDr)* and Time_Doctor *(TDr).Ciphertext contents: embedded conditions* OR-AND-NOT. The embedded conditions play the important role to dispatch the rights for viewing the information incorporated into the ciphertext. In addition, the ciphertext *C (M)* can be accessible by everyone. Decryption is only possible *iff* the attribute set of the secret key satisfies the access policy specified in the ciphertext. If the data is encrypted with embedded conditions such as "[Doctor Staff] OR [Practioner AND Family member]", then it can be decrypted with a key containing the attributes [Doctor, Nurse, Wife], but not by a key containing the attributes [Lab Personnel, Financial department]. The cases in the following illustrate the states among many cases.

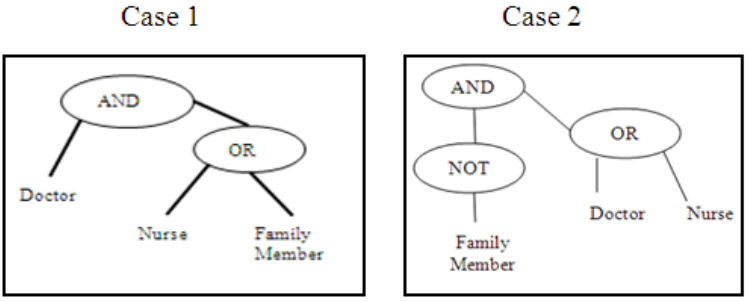


Fig. 6. Embedded Conditions with restricted rights to Family Member

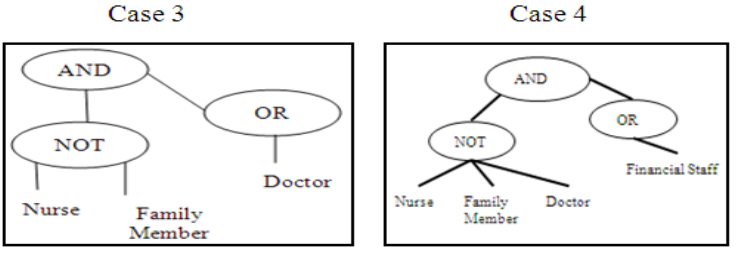
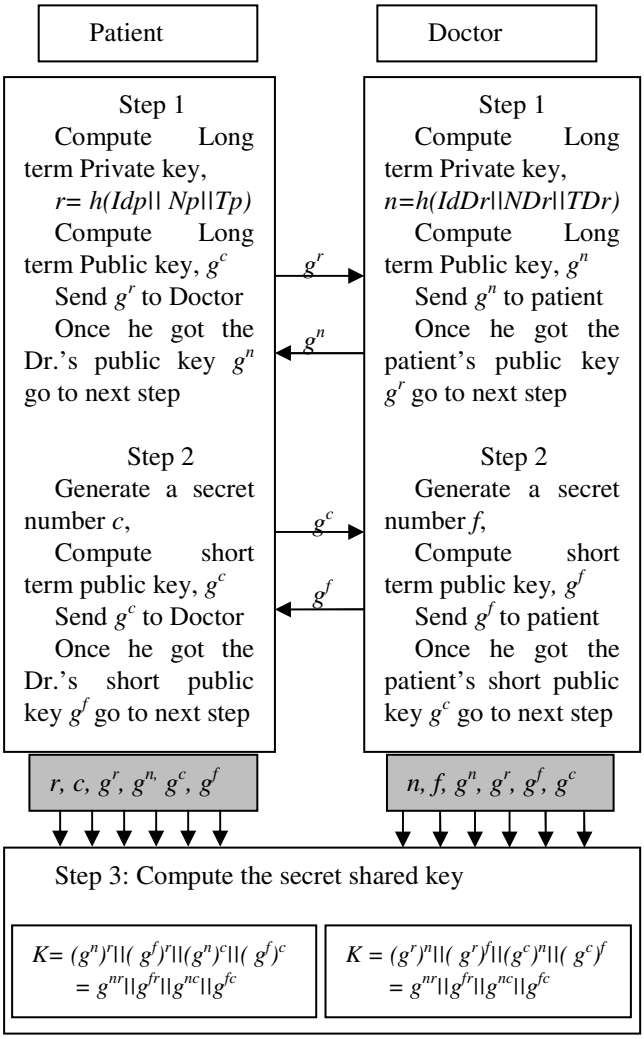


Fig. 7. Embedded Conditions with Access Rights to Doctor and F Staff

As abovementioned, the embedded conditions in the proposed cryptographic scheme define who has the rights to decrypt the patient's records data otherwise the file will be opened in encrypted form. In the system of eHealthcare, we assume that the ciphertext *C (M)* can be accessible by everyone and also the encryption/decryption operations are done into both ways. The accessibility rights from Figure 6 and Figure 7 can be explained like this:

- Case 1: The privileges are granted to all entities, once downloaded the patient's data, they can open it following the cipher embedded conditions
- Case 2: Doctor and Nurse have the rights to decipher the data while family member does not have.



- Case 3: The embedded condition *"NOT"* does not allow the Nurse and Family Member to read the patients' records data. Only Doctor can communicate with his patients.
- Case 4: The only Financial Staff can read the patients data(Data relating to the bills and payment process)

Key Exchange Computation Procedures. The ciphertext C (M) stored by data owner contains the keys and they are computed into the system. They play vital role in our proposed method such as: mutual authentication, data access rights, encryptions and decryption. This subsection describes the fundamental operations to compute the keys.

Once the shared keys computations between patient and Doctor are done, the encryption process of patient data can start whenever the user wants to upload it to the cloud. From the aforementioned operations, we can see that, the keys contain all IDs to identify each entity involved into the exchange and communication flow. In this paper we consider only those attributes and for the future work we will take in consideration others in case of experiment and implementation purposes.

3.3 Access Credential Architectural Principle

The keypoint architectural principle behind the proposed cryptographic scheme is the ability of patient to issue access credentials to Doctor and then he would be trust the cloud provider for his data security and privacy. A credential is a statement that specifies what access rights its holder has with respect to very specific data. The credential is cryptographically signed by its issuer. The possessor of the credential may present it to the issuer to gain access to the data. For the purpose of this work we are using the Key Note Trust Management system [14] which provides us with the necessary credential functionality. In this paper we describe two types of credentials where the patient issue the credentials according to the aforementioned cases (we take Case 1 and Case 3); the holder will have access to all of the data owner's patient records. The credential contains the public keys of the two parties along with the cryptographic signature (shared secret key) verifying the validity of the credential. The specific credential also has an expiration date, invalidating it past that date. Finally the credential has an extra field specifying the type of application the credential is supposed to be used for, in this case cloud computing. Example 1 is a sample credential for allowing the entities in case 1 to read patient records and Example 2 is a sample credential corresponding to case 3 where nurse and family member are not allowed to ready patient's data.

```
Authorizer: Patient_ID
    Local-Constants:
      Patient-ID_KEY="SecretKey||rsa-base64: MIGJb..."
      Doctor-ID_KEY= "SecretKey||rsa-base64: MIGJb..."
      Nurse-ID_Key= "SecretKey||rsa-base64: MIGJb... "
      Family Member-ID_Key= "SecretKey||rsa-base64: MIGJb... "
    Conditions:
    ((app_domain == "CLOUD_COMPUTING") &&
    (Medical Data == "Checkup_Results") &&
    (Permissions == "Access_Read_Only") &&
    (Timevalid <= "20130630")) -> "permit_if_not_Invalid";
    Licensees: Doctor-ID_OR_Nurse-ID_AND_Family member
Signature: "sig-shared-secret-key: QU6..."
```

Example 1: Credential for allowing the entities in case 1

```
Authorizer: Patient_ID
    Local-Constants:
      Patient-ID_KEY="SecretKey||rsa-base64: MIGJb... "
      Doctor-ID_KEY= "SecretKey||rsa-base64: MIGJb..."
      Nurse-ID_Key= "SecretKey||rsa-base64: KI&^%4... "
      Family Member-ID_Key= "SecretKey||rsa-base64: KI&^%4... "
    Conditions:
    ((app_domain == "CLOUD_COMPUTING") &&
    (Medical Data == "EGC-Glucose-Measurements") &&
    (Permissions == "Access_Read_Prescription") &&
    (Timevalid <= "20130625_26")) -> "permit_if_not_Invalid";
    Licensees: AND_Doctor-ID_NOT_Nurse-ID_OR_Family member-ID
Signature: "sig-shared-secret-key: QU6..."
```

Example 2: Credential for allowing the entities in case 3

4 Security Analysis

The scheme is secure to chosen ciphertext-only attack: Data transmissions from patient to cloud as well as from cloud to patient are done with proper encryption. The processes are the same under Chosen Ciphertext Attack (CCA) [15] based on the modification of stored ciphertext to cloud. Such adversaries are permitted to see the ciphertexts for messages of their choice, and (in the public-key setting) to generate ciphertexts on their own. However, the adversary never gets to see the decryptions of any messages. For the reason that the credentials contain hashed secrets parameters $r = h\ (Idp||\ Np||Tp)$ and $n=h\ (IdDr||NDr||TDr)$ to verify the authenticity of the attacker. Nonetheless, for some applications, the adversary will need a stronger definition in which he gets (limited) access to the decryption machinery as well.

The scheme is resistant to the eavesdropping attack: The aim of an eavesdropping attacker is to have the access to the private and sensitive patient's medical data. This attack may be happened between the involved entities during communication and exchange message. To access the data at the health cloud server, an attacker needs to have sufficient attributes to complete the access authentication protocol process *(step1, step2 and step 3)*. Here the shared secret key is made by multiple secrets parameters based on the attributes set. For the non-privacy dataset, he may get access and allowed in our scheme. But he cannot modify the data due to the verification bindings. However, for the patient sensitive data, a unique random number is embedded into the ciphertexts (*Np*) and *(NDr)*. Without knowing that secret number, it is impossible to access the data. Therefore, any attacker cannot successfully launch the eavesdropping even thought sophisticated applications can do that.

The scheme ensures message integrity, non-repudiation, and source authentication: We use the patient's secret key and the session identity to generate the signature playing the session key role $K = (g^r)^n || (g^r)^f || (g^c)^n || (g^c)^f$. The data receiver can verify the signature by using the public parameters of the sender "g^c". This verification ensures the corresponding source authentication. The scheme computes secrets keys "*n*" and "*r*" by computing their hash values of the concatenated message. Only the patient and Doctor know each other their secrets keys which include the same of their attributes such Ids. With others subkeys, the secret key are also used to generate the signature "*K*". Therefore the message integrity with non-repudiation can be provided by our proposed scheme.

5 Conclusion

In this paper, we presented a healthcare data collection system where we designed and implemented the system. We have shown the HR result obtained using the sensor node. In addition we described the cryptographic scheme for protecting the HR data where a patient can encrypt his data before uploading and storing to cloud provider. The aim of our proposed method it is to revoke the patients' semi-trusted assumption to cloud provider by giving the patient the ability to issue an access credential with embedded rights to the authorized person to decrypt the patient's record medical data.

First, we defined the attributes of the involved entities in this paper and we described the contents of the ciphertext *C (M)* uploaded to cloud. Moreover we presented the concept of credential based on the key note trust management which provides the credentials functions. By the end, we made a security analysis regarding known attack to cipher and cloud storage, hence the proposed scheme have been founded efficient and resilient to various kinds of attacks

Acknowledgments. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology. And it also supported by the BB21 project of Busan Metropolitan City.

References

1. Osterhaus, L.C.: Cloud Computing and Health Information. U of I SLIS Journal, Iowa Research, University of Iowa's Institutional Repository (November 2010)
2. Matyas, V.: Protecting doctors' identity in drug prescription analysis. Health Informatics Journal, 205–209 (1998)
3. Ateniese, G., de Medeiros, B.: Anonymous e-prescriptions. In: The Proceedings of the ACM Workshop on Privacy in the Electronic Society, pp. 19–31. ACM Press (2002)
4. De Decker, B., Layouni, M., Vangheluwe, H., Verslype, K.: A privacy-preserving eHealth protocol compliant with the Belgian healthcare system. In: Mjølsnes, S.F., Mauw, S., Katsikas, S.K. (eds.) EuroPKI 2008. LNCS, vol. 5057, pp. 118–133. Springer, Heidelberg (2008)
5. Anderson, R.: A security policy model for clinical information systems. In: The Proceedings of the 17th IEEE Symposium on Security and Privacy, pp. 30–44. IEEE CS (1996)
6. Reid, J., Cheong, I., Henricksen, M., Smith, J.: A novel use of rBAC to protect privacy in distributed health care information systems. In: Safavi-Naini, R., Seberry, J. (eds.) ACISP 2003. LNCS, vol. 2727, pp. 403–415. Springer, Heidelberg (2003)
7. Evered, M., Bögeholz, S.: A case study in access control requirements for a health information system. In: Proceedings of the. 2nd Australian Information Security Workshop. Conferences in Research and Practice in Information Technology, vol. 32, pp. 53–61. Australian Computer Society (2004)
8. Hung, P.C.K.: Towards a privacy access control model for e-healthcare services. In: Proceedings of the 3rd Annual Conference on Privacy, Security and Trust, October 12-14 (2005)
9. Masys, D.R., Baker, D.B.: Patient-Centered Access to Secure Systems Online (PCASSO): A Secure Approach to Clinical Data Access via the World Wide Web (1997)
10. Gematik. Einführung der Gesundheitskarte - Gesamtarchitektur, Version 1.7.0 (August 2009)
11. Gematik. Einfüuhrung der Gesundheitskarte - Netzwerkspezi_kation, Version 2.0.0 (August 2009)
12. Kemis, H., Bruce, N., Ping, W., Lee, H.J., Gook, L.B., Antonio, T.: Healthcare Monitoring Application for Ubiquitous Sensor Network: Design and Implementation based Pulse Sensor with Arduino. In: The 6th International Conference on New Trends in Information Science and Service Science (NISS 2012), Taipei, Taiwan, October 23-25 (2012)
13. Blaze, M., Feigenbaum, J., Ioannidis, J., Keromytis, A.: The role of trust management in distributed systems security. In: Vitek, J. (ed.) Secure Internet Programming. LNCS, vol. 1603, pp. 185–210. Springer, Heidelberg (1999)
14. Blaze, M., Feigenbaum, J., Ioannidis, J., Keromytis, A.D.: The Key Note Trust Management System Version 2. RFC 2704 (September 1999)

Partitioning-Based Selection of Aggregator Nodes in Wireless Sensor Networks

Aziz Nasridinov, Wuin Jang, and Young-Ho Park*

Department of Multimedia Science, Sookmyung Women's University,
Cheongpa-ro 47-gil 100, Yongsan-Ku, Seoul, 140-742, Korea
{aziz,yhpark}@sm.ac.kr, pinky0715@nate.com

Abstract. Since the communication is a main source of energy consumption in WSN, it is preferably to jointly collect sensor data from different sensor nodes, combine it based on specific variables, and forward combined data to the base station. In a typical data aggregator process, the aggregator nodes perform the data aggregation process. However, aggregator nodes can reside in the region, where the attributes of the aggregator node can change easily. In this paper, we study partitioning-based selection aggregator nodes in the WSNs. We argue that selecting an aggregator node from the region, where the attributes of the aggregator node can change easily is not efficient. Thus, in our method, we first propose to partition the sensor networks into several regions, and select an aggregator node only from those regions that are defined to be unchanged. We further formulate the selection process of an aggregator node as a top-*k* query problem, where we efficiently solve the problem by using a modified Sort-Filter-Skyline (SFS) algorithm. Through the experiments, we will demonstrate that the proposed method outperforms the existing methods by up to several times.

Keywords: data aggregation, aggregator node selection, top-*k*.

1 Introduction

Recently, wireless sensor networks (WSNs) have been widely used in many fields of healthcare, wildlife monitoring, meteorological hazards, natural disaster, and military target tracking and surveillance. A WSN consists of sensor nodes with limited battery power, computing capability and memory. Since the communication is a main source of energy consumption in WSN [1], it is better to jointly collect sensor data from different sensor nodes, combine it based on specific variables, and forward combined data to the base station. This process is called as a *data aggregation process*. One of the advantages of the data aggregation process is when the base station initiates the query on the WSN, rather than sending each sensor node's data to the base station, one of the sensor nodes performs the data aggregator process [2]. Thus, the data aggregation process reduces redundant data transmissions and communication overhead, and improves the network life-time.

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_96,

In a typical data aggregator process, the aggregator nodes perform the data aggregation process. The aggregator nodes receive the sensor data results from sensor nodes in a specific region and perform computations on the data to produce a collective view of the observed physical phenomenon. This computation could be simply finding a mean value of the observations in that nodes area [3]. This aggregation result is then transmitted to the base station instead of having individual sensor nodes send their result to the base station. Careful selection of the aggregator nodes in the data aggregation process results in reducing large amounts of communication traffic in the WSNs. However, network conditions change continuously due to sharing of resources, computation load, and congestion on network nodes and links, which makes the selection of the aggregator nodes difficult [2]. Moreover, the selected aggregator nodes can reside in the region, where the attributes of the aggregator node can change easily. For example, in the seismic region or in the region where the weather changes frequently, a sensor node can change its geographical position or energy consumption level.

In this paper, we study partitioning-based selection aggregator nodes in the WSNs. We argue that selecting an aggregator node from the region, where the attributes of the aggregator node can change easily is not efficient. Thus, in our method, we first propose to partition the sensor networks into several regions, and select an aggregator node only from those regions that are defined to be unchanged. We further formulate the selection process of an aggregator node as a top-k query problem, where we efficiently solve the problem by using a modified Sort-Filter-Skyline (SFS) [4] algorithm. Through the experiments, we will demonstrate that the proposed method outperforms the existing methods by up to several times. We also provide an analysis of the major factors that impact the performance of previous approaches.

The remainder of this paper is organized as follows. Section 2 discusses the related work. Section 3 describes our proposed method. Section 4 presents performance evaluation. Section 5 highlights conclusions and future work.

2 Related Study

Data aggregation protocols can be categorized into two types such as tree-based data aggregation protocols and cluster-based data aggregation protocols. In this section, we will discuss only representative methods.

In tree-based data aggregation protocols, the aggregator node is determined, and the data is transformed to the base station through the determined data aggregator nodes. Madden et al. [5] proposed the TAG (Tiny Aggregation) service for aggregation in low-power, distributed, wireless environments. TAG enables users to express simple, declarative queries by borrowing an idea from the aggregation operators in database query language, and have them distributed and executed efficiently in networks of low-power, wireless sensors. Lindsey et al. [6] proposed PEGASIS (Power-Efficient Gathering in Sensor Information Systems), which is near-optimal chain-based protocol that minimizes energy. The main idea of the PEGASIS is to form a chain among the sensor nodes and have each sensor node communicate

only with a close neighboring node and takes turns forwarding the data to the base station.

In cluster-based data aggregation protocols, sensor nodes are divided into clusters. In each cluster, a cluster head is selected. Cluster head aggregates the sensor data locally and forwards the aggregation result to the base station. Heinzelman et al. [7] proposed LEACH (low-energy adaptive clustering hierarchy). LEACH includes a new, distributed cluster formation technique that enables self-organization of large numbers of nodes, algorithms for adapting clusters and rotating cluster head positions to evenly distribute the energy load among all the nodes, and techniques to enable distributed signal processing to save communication resources. Younis and Fahmy [8] proposed HEED (hybrid energy-efficient distributed), that periodically selects cluster heads according to a hybrid of the node residual energy and a secondary parameter, such as node proximity to its neighbors or node degree. HEED terminates in $O(1)$ iterations, incurs low message overhead, and achieves fairly uniform cluster head distribution across the network. The disadvantages of tree-based and cluster based data aggregation protocols are that they do not consider geographical position of the sensor nodes when determining aggregator nodes. Since the selected aggregator nodes can reside in the region, where the attributes of the aggregator node can change easily, it can lead to the re-selection of aggregator node frequently, which is time-consuming in WSNs.

3 Partitioning-Based Selection of Aggregator Nodes

In this paper, we study partitioning-based selection aggregator nodes in the WSNs. In this section, we describe our proposed method. The proposed approach mainly contains two steps such as partitioning step and skylining step. Figure 1 demonstrates partitioning-based selection aggregator nodes. The sensor nodes are represented as points in the 2-dimensional space, with the coordinates of each point indicating the values of the sensor nodes in two attributes, such as power consumption and communication cost.

We assume that the proposed approach considers aggregator node selection in a medium-scale sensor network. In a typical medium-scale sensor network, the number of nodes can reach 200–300 sensor nodes, where the potential candidate to be an aggregator node does not exceed 30–40 sensor nodes, according to the size of each cluster. However, it is important to mention that in a large-scale sensor network, potential candidates to be an aggregator node can be huge, which means that there is a need to build an index before applying our approach.

The partitioning steps are shown in Figure 1 (a), where the universe (i.e. WSN) consists of thirteen sensor nodes. In partitioning step, we partition the sensor network into p subspaces using grid-based partition technique. The skylining step is shown in Figure 1 (b), where we distinguish frequently changing region and safe regions. We determine the frequently changing region among the partitioned regions using a *change analyzing function.* The change analyzing function calculates various

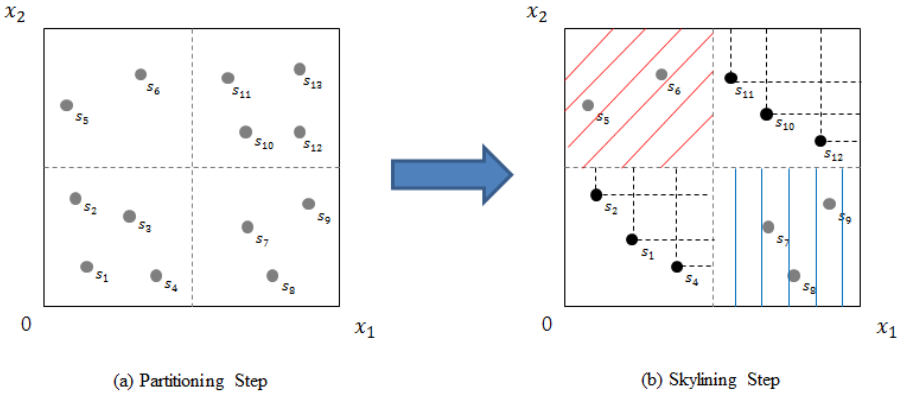


Fig. 1. Partitioning-based selection of aggregator nodes

attributes of the sensor network, such as density of nodes, geographic location and seismicity of the region. In Figure 1 (b), red and blue striped regions are determined as frequently changing region. We do not select aggregator nodes from these regions. Once we determine the frequently changing region, we then select an aggregator node only from those regions that are not defined to be unchanged. In the aggregator node selection step, we immediately perform a skyline query on the sensor nodes only in the regions that defined to be unchanged. This enables to extract the sensor nodes that are potential candidates to become an aggregator node. Our approach selects a set of aggregator nodes according to their attributes, such as such as distance from the base station, power consumption, battery life and communication cost. Thus, we can reduce large amounts of communication traffic by sending only the aggregated data through selected aggregator nodes, instead of individual sensor data, to the base station.

4 Performance Evaluation

In this section, we present performance evaluation of our approach. The aim of the experiment is to compare the computation time of the proposed approach with the method of traditional clustering data aggregation protocol. We called this method as NIP [2] by taking author's initials.

4.1 Experimental Setup

Experiments were carried out on a 2.4GHz Pentium processor with 512MB of RAM running Windows XP Professional. For implementation of our proposed approach, we used C++ programming language. We used syntactic data which consists of 100K data records. The following experiments are carried out.

4.2 Experimental Results

We compare node selection time. Graphs in Figures 2 demonstrate this comparison. In Figure 2, *x*-axis represents aggregation node selection time in milliseconds and *y*-axis represents *d* dimensions in universe. *d* dimensions can be interpreted as the sensor node attributes such as distance from the base station, power consumption, battery life, and communication cost.

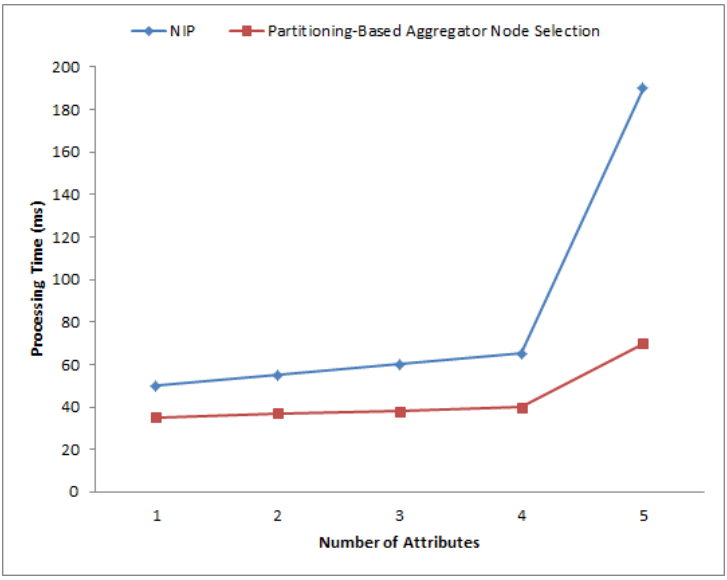


Fig. 2. A comparison of node selection time

From the graph in Figure 2, we can observe that the proposed method outperforms the traditional clustering data aggregation protocol (NIP), by up to 2 times. Traditional clustering data aggregation protocol selects aggregator nodes that can reside in the region, where the attributes of the aggregator node can change easily. For example, in the seismic region or in the region where the weather changes frequently, a sensor node can change its geographical position or energy consumption level. On the other hand, in our method, we first propose to partition the sensor networks into several regions, and select an aggregator node only from those regions that are defined to be unchanged. Thus, we can select aggregator nodes more accurately.

5 Conclusion

In this paper, we have proposed partitioning-based selection aggregator nodes in the WSNs. We insist that selecting an aggregator node from the region, where the attributes of the aggregator node can change easily is not efficient. Thus, in our method, we first proposed to partition the sensor networks into several regions, and

select an aggregator node only from those regions that are defined to be unchanged. We further formulated the selection process of an aggregator node as a top-k query problem, where we efficiently solve the problem by using a modified Sort-Filter-Skyline (SFS) [4] algorithm. Through the experiments, we will demonstrate that the proposed method outperforms the existing methods by up to several times. We also provide an analysis of the major factors that impact the performance of previous approaches.

Acknowledgement. This work was supported by the IT R&D program of MKE/KEIT. [10041854, Development of a smart home service platform with real-time danger prediction and prevention for safety residential environments].

References

1. Pottie, G.J., Kaiser, W.J.: Wireless integrated network sensor. Communications of the ACM 43(5), 51–58 (2000)
2. Nasridinov, A., Ihm, S.Y., Park, Y.H.: Skyline-Based Aggregator Node Selection in Wireless Sensor Networks. International Journal of Distributed Sensor Networks 2013, 1–7 (2013)
3. Eggen, S.: Security in wireless sensor networks. Master Thesis, University of Oslo (2008)
4. Chomicki, J., Godfrey, P., Gryz, J., Liang, D.: Skyline with presorting. In: Proceedings of the 9th International Conference on Data Engineering, pp. 717–719 (2003)
5. Madden, S., Franklin, M.J., Hellerstein, J.M., Hong, W.: TAG: a tiny aggregation service for ad-hoc sensor networks. Proceedings of the 5th Symposium on Operating Systems Design and Implementation (OSDI 2002), 131–146 (2002)
6. Lindsey, S., Raghavendra, C., Sivalingam, K.M.: Data gathering algorithms in sensor networks using energy metrics. IEEE Transactions on Parallel and Distributed Systems 13(9), 924–935 (2002)
7. Heinzelman, W.B., Chandrakasan, A.P., Balakrishnan, H.: An application-specific protocol architecture for wireless microsensor networks. IEEE Transactions on Wireless Communications 1(4), 660–670 (2002)
8. Younis, O., Fahmy, S.: HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. IEEE Transactions on Mobile Computing 3(4), 366–379 (2004)

A Propose on a Varied Dual Match Method for Subsequence Queries

Sun-Young Ihm, Wuin Jang, and Young-Ho Park

Department of Multimedia Science, Sookmyung Women's University,
Cheongpa-ro 47-gil 100, Yongsan-Ku, Seoul, 140-742, South Korea
{sunnyihm,yhpark}@sm.ac.kr, pinky0715@nate.com

Abstract. Time-series data has been widely used in many databases applications, such as data mining and data ware-housing. One of the main tasks in handling time-series data is to find subsequences matches similar to a given query sequence. The state-of-the-art methods to find subsequences matches in time-series data leave many remained data sequence not compared for determining candidates and thus, produce many false alarms. In this paper, we propose a varied dual match method for subsequence queries. In our proposed method, we compare the remained data sequence for determining candidates, and calculate the distance between data subsequence and query sequence. Hence, our proposed method reduces the number of candidates and false alarms. Through the experiments with synthetic data and query sets, we show that our proposed method reduces the number of candidates by up to several times comparing to the state-of-the-art methods.

Keywords: Time-series data, subsequence matching, Dual Match.

1 Introduction

Time-series data has been widely used in many databases applications, such as data mining and data ware-housing [1, 2]. Examples include stock prices, exchange rates, and weather data [3, 4]. In stock databases, we want to find similar patterns from the past stock price data so that these patterns resemble last week's stock price pattern of a company. In this paper, we consider a problem of *subsequence matching* [3, 5] in time-series data, where the lengths of data and query sequences are different. Specifically, when D is defined as the Euclidean distance [3, 5, 6], we define ϵ-match as follows: two sequence Q and S are in ϵ-match if $D(Q, S)$ is less than or equal to a given tolerance ϵ. The outline of subsequence matching methods is as follows. First, in the index building phase, each data sequence is transformed into the frequency domain by using a transformation method, and it is regarded as an f-dimensional point. This point is indexed. In the query processing phase, a query sequence given by a user is similarly transformed to an f-dimensional point, and a range query is constructed using the point and the given tolerance ϵ.

The problem of efficiently locating matches of a given time sequence in time-series has been studied by many researchers [3, 5, 7, 8]. For example, I-adaptive [7] divides

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_97,

data sequence into sliding windows and the query sequence into disjoint windows. It constructs minimum bounding rectangles (MBRs). However storing MBRs causes false alarms. Dual Match [5] and E-Dual Match [8], which use different window construction approach from I-adaptive, improve the performance in subsequence matching comparing to I-adaptive. Dual Match also divides data sequence into sliding windows and the query sequence into disjoint windows. It does so by introducing a duality in constructing the windows. However, Dual Match still produces many false alarms by filtering points through comparing only one query window with its corresponding data window. On the other hand, E-Dual Match handles all possible query windows for determining candidates. Hence, E-Dual Match reduces the false alarms, and improves performance comparing to Dual Match. However, Dual Match and E-Dual Match leave many remained data sequence not compared for determining candidates and thus, produce many false alarms (i.e. candidates that do not qualify to be an answer to the query).

In this paper, we propose a varied dual match method for subsequence queries. In our proposed method, we compare the remained data sequence for determining candidates, and calculate the distance between data subsequence and query sequence. Hence, our proposed method reduces the number of candidates and false alarms. Through the experiments with synthetic data and query sets, we show that our proposed method reduces the number of candidates by up to several times comparing to the state-of-the-art methods.

The remainder of this paper is organized as follows. Section 2 discusses related work. Section 3 describes our proposed approach. Section 4 presents performance evaluation. Section 5 highlights conclusions.

2 Related Work

There are many papers [3, 5, 7, 8] that proposed methods for efficiently locating matches for a given time sequence. For example, Faloutsos et al. [7] have proposed the subsequence matching method, called I-adaptive. I-adaptive consists of index building and subsequence matching algorithms. First, in the index building phase, each data sequence is transformed into the frequency domain by using a transformation method, and it is regarded as an f-dimensional point. This point is indexed. In the query processing phase, a query sequence given by a user is similarly transformed to an f-dimensional points, and a range query is constructed using the point and the given tolerance t. This approach guarantees that there are no false dismissals, but may cause false alarms because it uses f features instead of the whole data sequence. Thus, for each candidate subsequence obtained, the actual data sequence is read from the disk, and the distance from the query sequence is computed. The candidate is discarded if it is a false alarm.

Dual Match [5] also divides data sequence into sliding windows and the query sequence into disjoint windows. It does so by introducing a duality in constructing the windows. Dual Match can reduce the number of points to store comparing to I-adaptive. By storing individual points directly, Dual Match exploits point-filtering

effect, and accordingly, reduces false alarms and improves performance significantly. However, it has the problem of having a smaller allowable window size half that of I-adaptive given the minimum query length. A smaller window increases false alarms due to window size effect. Ihm et al. [8] proposed an efficient subsequence matching method, which is called the Efficient Duality-based Subsequence Matching (simply, E-Dual Match). E-Dual Match handles all possible query windows for determining candidates. Hence, E-Dual Match reduces the false alarms, and improves performance comparing to Dual Match. However, Dual Match and E-Dual Match leave many remained data sequence not compared for determining candidates and thus, produce many false alarms

3 Proposed Method

In this paper, we propose a varied dual match method for subsequence queries. In this section, we first explain subsequent matching. Then we describe the construction steps of the proposed method.

3.1 Subsequent Matching

One of the main data mining tasks in time-series data is to efficiently find subsequences similar to a given query sequence. In this paper, we study a problem of *subsequence matching*, where the lengths of data and query sequences are different. We formally describe the problem of subsequence matching. Given two time sequence S and Q which may have different length, the goal is to find all similar subsequence pairs between two time sequences with a specified threshold. We describe some relevant definitions for the targeted problem as follows.

Definition 1 *(Time Sequence):* A time sequence S, $S(s_1, s_2, s_3, \dots, s_{l\,(S)})$, is an ordered set of real values, where s_i is the i^{th} element of S, and $l(S)$ is sequence length of S.

Definition 2 *(Time Subsequence):* A time subsequence is an ordered sequence. $S_{i,j} = s_i, s_{i+1}, s_{i+2}, \dots, s_j$ denotes the time subsequence of a time sequence S, which contains the elements of S in positions i through j, and the length of $S_{i,j}$ is $(l)S_{i,j} = i + j - 1$.

Definition 3 *(Similar Time Sequence):* Two time S and Q are called similar if and only if $W(S, Q) \geq \Delta$, where $W()$ is a function for calculating similarity between S and Q, and Δ is a specified threshold value.

Definition 4 *(Similar Subsequence):* Given two time sequence S and Q, the sequences S' and Q' are called a similar subsequences of S and Q if and only if S' and Q' are similar and they subsequences of S and Q, respectively.

Definition 5 *(Distance of Similar Subsequence):* If S and Q have similar subsequences $S_{i,j}$ and $Q_{x,y}$, then the distance between $S_{i,j}$ and $Q_{x,y}$ is $D(S_{i,j}, Q_{x,y}) = |i - x|$.

Notice that any kind of measure could be used for the similarity function $W()$ in Definition 3. Based on the above definitions, the shape or trend of two subsequences should be very close if they are similar subsequences.

3.2 Construction of Varied Dual Match

The proposed method has four steps. Index building is the first step of the proposed method. It constructs R*-tree index using the same method as Dual Match. First, we divide the data sequence into disjoint windows and then, transform each window into low-dimensional data with *f*-points using DFT (Discrete Fourier Transform) [5]. Next, we construct the R*-tree index with the transformed points and store these points as a file. The data values in this file are stored sequentially. In the window filtering step, we search R*-tree index and file, compare all windows in the data subsequence with all windows in the query sequence. If the result of comparison is small than tolerance *t* defined by user, then we consider this data subsequence as candidate. In the window filtering step, we search R*-tree index and file, compare all windows in the data subsequence with all windows in the query sequence. If the result of comparison is smaller than tolerance *t* defined by user, then we consider this data subsequence as candidate. A real distance between data subsequence and query sequence in the candidate set is calculated using Euclidean distance.

4 Performance Evaluation

In this section, we first explain the experimental setup and data used in the experiment in Section 5.1. Then, we show the advantage of the proposed methods through the results of the experiments in Section 5.2.

4.1 Experimental Setup and Data

We compare the number of candidates of our proposed method with the existing methods, such as Dual Match [5] and E-Dual Match [8]. We perform experiments using synthetic dataset, which is used in [5]. The synthetic data set, called MIX-DATA, consists of 1,055,525 entries. We experiment the proposed method with seven queries used in Dual Match [5], where the query size is 384. Table 1 shows each tolerance *t* for seven queries of MIX-DATA. We implement Dual Match, E-Dual Match and proposed method in C language. We conduct the experiments on an Intel i5-760 quad core processor running at 2.80GHz in Linux PC with 16GB of main memory.

4.2 Experiment Results

Table 2 summarizes the experiments and the parameters used for the comparison of the number of candidates.

Table 1. The tolerance *t* of queries

Number of Queries	Tolerance *t* for MIX-DATA
1	1.091062
2	1.80611
3	15.680434
4	27.911256
5	1596.712525
6	27.784614
7	16069.48438

Table 2. Summary of parameters for Experiment 1

Number of Queries	Number of candidates of Dual Match	Number of candidates of E-Dual Match	Number of candidates of Varied Dual Match
1	21403	4451	480
2	52046	25624	855
3	786	303	169
4	5600	100	35
5	5058	273	72
6	6806	4295	2256
7	17845	854	63

Figure 1 shows the result of comparison of the number of candidates of Dual Match, E-Dual Match and Varied Dual Match using MIX-DATA, which is the synthetic dataset. We use seven queries for the experiment. In the graph of Figure 3, *x* axis represents the number of queries, and *y* represents the number of candidates. Varied Dual Match reduces 3.01~293.25 times over the number of candidates of Dual Match and 1.78~30.41 times over the number of candidates of E-Dual Match.

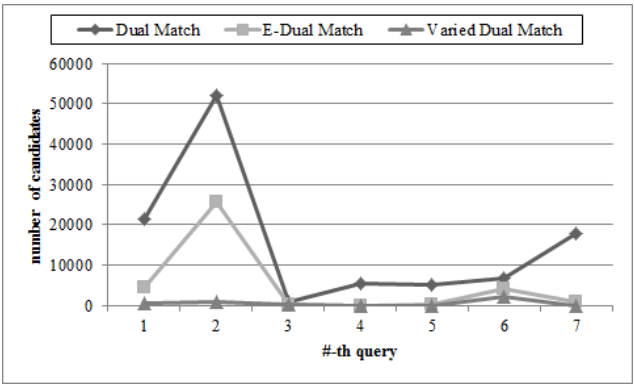


Fig. 1. The comparison of the number of candidates using MIX-DATA

5 Conclusion

In this paper, we have proposed a varied dual match method for subsequence queries. In our proposed method, we compared the remained data sequence for determining candidates, and calculate the distance between data subsequence and query sequence. Hence, our proposed method reduces the number of candidates, false alarms, and significantly improves the performance of subsequence matching. Through the experiments with synthetic data and query sets, we showed that our proposed method reduces the number of candidates by up to several times comparing to the existing methods.

Acknowledgement. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012003797).

References

[1] Ding, H., Trajcevski, G., Scheuermann, P., Wang, X., Keogh, E.: Querying and mining of time series data: experimental comparison of representations and distance measures. Proceedings of the VLDB Endowment 1(2), 1542–1552 (2008)
[2] Rafiei, D., Mendelzon, A.: Similarity-Based Queries for Time Series Data. ACM SIGMOD Record 26(2), 13–25 (1997)
[3] Agrawal, R., Faloutsos, C., Swami, A.: An Efficient similarity search in sequence databases. In: Lomet, D.B. (ed.) FODO 1993. LNCS, vol. 730, Springer, Heidelberg (1993)
[4] Agrawal, R., Lin, K.I., Sawhney, H.S., Shim, K.: Fast similarity search in the presence of noise, scaling, and translation in time-series databases. In: Proceedings of the 21th International Conference on Very Large Data Bases, pp. 490–501 (1995)
[5] Moon, Y.S., Whang, K.Y., Loh, W.K.: Duality-based subsequence matching in time-series databases. In: Proceedings of the 17th International Conference on Data Engineering, pp. 263–272 (2001)
[6] Chan, K.P., Fu, A.W.C.: Efficient time series matching by wavelets. In: Proceedings of the 15th International Conference on Data Engineering, pp. 126–133 (1999)
[7] Faloutsos, C., Ranganathan, M., Manolopoulos, M.: Fast subsequence matching in time-series databases. In: Proceedings of the 1994 ACM SIGMOD International Conference on Management of Data, pp. 419–429 (1994)
[8] Ihm, S.Y., Nasridinov, A., Park, Y.H.: Efficient Duality-Based Subsequent Matching on Time-Series Data in Green Computing. Submitted to Journal of Supercomputing (2013)

A Survey on Density-Based Clustering Algorithms

Woong-Kee Loh[1] and Young-Ho Park[2,*]

[1] Department of Multimedia, Sungkyul University
[2] Department of Multimedia Science,
Sookmyung Women's University
woong@sungkyul.ac.kr,
yhpark@sm.ac.kr

Abstract. Density-based clustering forms the clusters of densely gathered objects separated by sparse regions. In this paper, we survey the previous and recent density-based clustering algorithms. DBSCAN [6], OPTICS [1], and DENCLUE [5, 6] are previous representative density-based clustering algorithms. Several recent algorithms such as PDBSCAN [8], CUDA-DClust [3], and GSCAN [7] have been proposed to improve the performance of DBSCAN. They make the most of multi-core CPUs and GPUs.

1 Introduction

Density-based clustering forms the clusters of densely gathered objects separated by sparse regions; it has the advantage that it can discover the clusters of arbitrary shapes and easily filter out noise objects. DBSCAN [6], OPTICS [1], and DENCLUE [5, 6] are widely used density-based clustering algorithms. OPTICS is an extension to DBSCAN that solves the problem of parameter selection, and DENCLUE 2.0 [5] is an upgrade of DENCLUE [6] that improves its performance. The previous density-based algorithms are explained in detail in Section 2.

As the dataset size increases dramatically, the performance of density-based algorithms has become more and more important. PDBSCAN [8] is a distributed parallel algorithm which performs DBSCAN for a massive dataset using multiple computers connected via network. CUDA-DClust [3] improves the performance of DBSCAN using a Graphics Processing Unit (GPU). A GPU contains many execution units called *cores* which perform simple operations such as distance computations in a massively parallel manner. CUDA-DClust* [3] is an extension to CUDA-DClust that uses the indexing technique. GSCAN [7] improved the performance of CUDA-DClust by splitting the whole dataset space into a number of grid cells and reducing the number of distance computations using the grid structure. The recent density-based algorithms are explained in detail in Section 3.

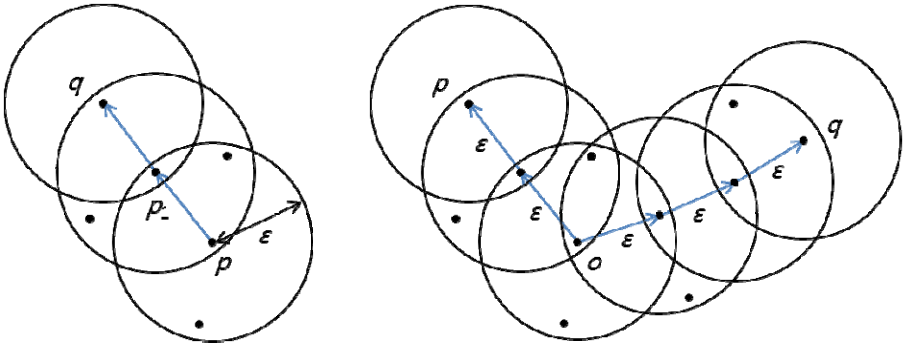
* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_98, © Springer-Verlag Berlin Heidelberg 2014

2 Previous Density-Based Algorithms

DBSCAN proposes the *density-connectivity* relationship for two objects, and defines a cluster as a maximum set of density-connected objects. DBSCAN has a weakness that it is difficult to discover parameters to obtain an optimal clustering result; it requires much time to discover those parameters. OPTICS is an extension of DBSCAN that solves the problem of parameter discovery. OPTICS performs clustering with various parameters simultaneously, and creates ordered clustering results. This enables to locate an optimal clustering result easily. DENCLUE defines an *influence function* for each object. For a unique object p, it defines a *density function* as the sum of the influence function values for all objects. The object whose density function value is the local maxima is defined as a *density attracter*. Density attracters are used to create clusters. DENCLUE performs faster than DBSCAN. However, in DENCLUE, the quality of clustering result is greatly influenced by parameter selection. DENCLUE 2.0 [5] is an upgrade of DENCLUE that improves the performance.

We briefly explain about DBSCAN. In DBSCAN, the density-connected relationship is defined using two input parameters such as ϵ and *MinPts*. For an object p, a set of all objects existing in an ϵ-range from p is called as ϵ-neighbor $N_\epsilon(p)$. If $N_\epsilon(p)$ contains *MinPts* or more objects, p is called a *core object*, and q is defined to be *directly density-reachable* from p. For two objects p and q, if there exist $p_1, \ldots, p_{m-1}$ objects and p_{i+1} is directly density-reachable from $p_i (0 \leq i < m, p_0 = p, p_m = q)$, p is defined to be *density-reachable* from q. For two objects p and q, if there exists an object o such that both p and q are density-reachable from o, p and q are defined to be *density-connected*. Note that the (directly) density reachability relationship among two objects p and q is uni-directional and density-connectivity is bi-directional. Figure 1 shows an example with *MinPts* = 4. Figure 1(a) shows that q is density-reachable from p. Figure 1(b) shows that p and q are density-connected with each other.



(a) q is density-reachable from p. (b) p and q are density-connected with each other.

Fig. 1. Density-reachability and density-connectivity in DBSCAN

Based on density-connectivity relationship between two objects, DBSCAN for a given set of objects D can be summarized as Algorithm 1. At first, the state of all objects set to be *undecided*. For each *undecided* object p, DBSCAN computes ϵ-neighbor $N_{\epsilon}(p)$ to determine whether p is a core object or a noise object. The time complexity of DBSCAN is $O(n^2)$, where n is the number of objects in D. If spatial index structure that contains all objects is constructed in advance, then for each object p, it is possible to compute ϵ-neighbor in $O(log\, n)$ time. Thus, the time complexity of DBSCAN becomes $O(n \log n)$.

Algorithm 1 DBSCAN.

```
1: Set the state of each object o in 𝒟 to be undecided;
2: while there exist undecided objects in 𝒟 do
3:    Choose any undecided object p ∈ 𝒟 and compute N_ε(p);
4:    if |N_ε(p)| ≥ MinPts then
5:       Form a new cluster C and insert p in C;
6:       Seed ← N_ε(p) − p;
7:       while there are undecided or noise objects in Seed do
8:          for each undecided or noise object q ∈ Seed do
9:             Insert q in C and compute N_ε(q);
10:            if |N_ε(q)| ≥ MinPts then
11:               Seed ← Seed ∪ N_ε(q) − q;
12:            end if
13:         end for
14:      end while
15:   else
16:      Set the state of p as noise;
17:   end if
18: end while
```

3 Recent Extensions of DBSCAN

Several algorithms have been proposed to improve the performance of DBSCAN. Xu et al. [7] proposed a distributed parallel algorithm called PDBSCAN, which performs DBSCAN for a massive dataset. PDBSCAN uses multiple computers connected via network. At first, PDBSCAN divides the entire dataset into N disjoint region S_i $(1 \leq i \leq N, S_i \cap_{i \neq j} S_j = \emptyset)$, and then transmits the objects in each region S_i to the slave computer C_i. Separate clustering is performed for objects in each slave computer C_i,. Once the separate clustering is completed, all clustering results are transmitted to the master computer, where the cluster merging process is performed. Brecheisen et al. [2] proposed a clustering algorithm dealing with complex objects such as trees, vectors, and graphs whose distance computation is highly costly. At first, the proposed algorithm performs approximate clustering based on low-cost lower-bounding distance. Each approximate cluster is transmitted to the several slave

computers. If there are large approximate clusters, they are further divided into several sub-clusters and transmitted to different slave computers. For objects assigned to each slave computer, distributed clustering is performed based on exact distance. Finally, clustering results obtained from sub-clusters are transmitted to the master computer, where a merging process is performed.

Bohm et al. [3] proposed CUDA-DClust that improves the performance of DBSCAN using a Graphics Processing Unit (GPU). While DBSCAN forms a cluster for a single undecided object p ($\in D$) at a time, CUDA-DClust assigns a single undecided object p to a GPU thread block and forms multiple clusters simultaneously by executing Algorithm 1. Each cluster created in a block is called as a *chain*. In order to speed up computing ϵ-neighbor $N_{\epsilon}(p)$ of p in each block, CUDA-DClust performs multiple distance computations from p to a different object in each thread simultaneously. If the collision occurs between the chains, i.e., an object is inserted into two different chains, it is recorded in a separate data structure called the *collision matrix*. In the final stage, collision chains are merged into a single cluster. Through experiments, Bohm et al. [3] demonstrated that CUDA-DClust outperforms DBSCAN by up to 15 times. However, CUDA-DClust has a disadvantage that most of distance computations from an undecided object p to all the other objects in D to compute $N_{\epsilon}(p)$ is unnecessary. Moreover, intermediate clustering results (collision matrix, etc.) are stored in the off-chip device memory of the GPU, which requires high cost of memory access.

CUDA-DClust* [3] is an extension of CUDA-DClust that uses the indexing technique. CUDA-DClust* uses a simple indexing technique in order to compute $N_{\epsilon}(p)$ for a single object p efficiently. It extracts candidate objects and computes the distance only for those objects. Thus, CUDA-DClust* decreases the number of calculations dramatically. CUDA-DClust* outperforms CUDA-DClust by up to 11.9 times. However, as in CUDA-DClust, CUDA-DClust* stores the intermediate clustering results in the costly off-chip device memory of the GPU. It has the following disadvantages too. There is an additional cost to build an index for a given dataset. Moreover, for scanning the index, it can hardly use the parallel features of the GPU. Generally, an index is constructed with multiple levels. The nodes to be scanned in lower levels are decided according to the scan result in upper levels. While only a small number of GPU threads are used to perform a scanning in upper levels, the other threads are put in the waiting state. In CUDA-DClust*, the index scanning is performed using a single thread, which causes performance degradation.

GSCAN [7] is an extension of CUDA-DClust, which improves the performance of DBSCAN. As shown in Algorithm 1, in order to compute the ϵ-neighbor $N_{\epsilon}(p)$ for each undecided object p, DBSCAN and CUDA-DClust computes the distance between p and all the other objects in the dataset D. However, as shown in Figure 2, most of the objects in D are located beyond ϵ range from p, thus it is unnecessary to perform the distance computation with them. GSCAN improves the performance by reducing unnecessary distance computations. As an experimental result, GSCAN outperformed CUDA-DClust and DBSCAN by up to 13.9 and 32.6 times, respectively.

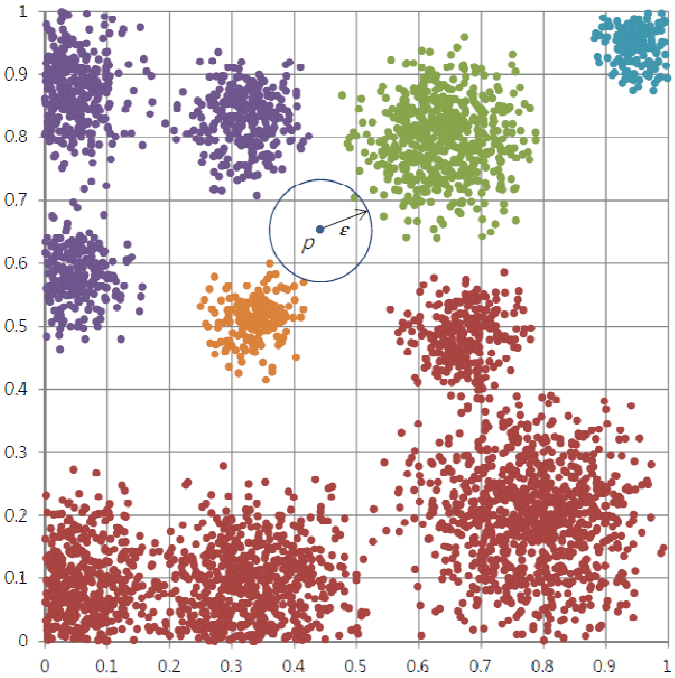


Fig. 2. Data objects divided into grid cells [7]

GSCAN reduces unnecessary distance computations as follows. At first, GSCAN divides the entire data space into grid cells as shown in Figure 2. Then, for each cell in the data grid, a list of objects that belongs to the cell is generated. In order to compute ϵ-neighbor $N_{\epsilon}(p)$ for each undecided object p, unlike DBSCAN and CUDA-DClust, GSCAN performs distance computations only with the objects contained in the cells located in ϵ range from p. As shown in Figure 2, GSCAN performs distance computations only with the objects in 9 cells around p. Since GSCAN computes the distance with much less number of objects, it achieves much better performance.

GSCAN finds the cells located in ϵ range from p as follows. Since there are many grid cells in the dataset, it is inefficient to compute the distance of each cell from p. GSCAN considers only a set of cells C_p around the cell C_p' containing p. Let c_i (≥ 0) be the coordinate of i-th dimension of C_p', then the coordinates of i-th dimension of the cells C_p are $c_i - 1, c_i, c_i + 1$. Thus, the number of cells C_p in the d'-dimensional grid is not more than $3^{d'}$. The x/y coordinates of the cells in Figure 2 are in [0, 9] range. The x-coordinate for C_p that contains objects p is 4, and thus the x-coordinates of the cells C_p' is 3, 4, and 5. A necessary condition is that the size for every dimension of a cell should not be less than ϵ; otherwise, some objects in $N_{\epsilon}(p)$ may not be contained in the cells C_p. Since the size for every dimension of the cells in Figure 2 is 0.1, ϵ should not exceed 0.1.

4 Conclusions

Recently, many advanced algorithms that use multi-computers and multi-core processors to process large datasets have replaced the traditional density-based algorithms that use a single thread on a single machine. In such algorithms, a few parallelism issues should be considered. Specifically, efficient methods to divide and deploy the entire dataset and to minimize the data and control transmission should be proposed.

References

1. Ankerst, M., Breunig, M.M., Kriegel, H.-P., Sander, J.: OPTICS: Ordering Points To Identify the Clustering Structure. In: Proc. of Int'l Conf. on Management of Data, ACM SIGMOD, Philadelphia, Pennsylvania, USA, pp. 49–60 (June 1999)
2. Brecheisen, S., Kriegel, H.-P., Pfeifle, M.: Parallel Density-Based Clustering of Complex Objects. In: Ng, W.-K., Kitsuregawa, M., Li, J., Chang, K. (eds.) PAKDD 2006. LNCS (LNAI), vol. 3918, pp. 179–188. Springer, Heidelberg (2006)
3. Bohm, C., Noll, R., Plant, C., Wackersreuther, B.: Density-based Clustering using Graphics Processors. In: Proc. Conf. on Information and knowledge management (CIKM), Hong Kong, China, pp. 661–670 (November 2009)
4. Ester, M., Kriegel, H.-P., Sander, J., Xu, X.: A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In: Proc. Int'l Conf. on Knowledge Discovery and Data Mining (KDD), Portland, Oregon, USA, pp. 226–231 (1996)
5. Hinneburg, A., Gabriel, H.-H.: DENCLUE 2.0: Fast Clustering Based on Kernel Density Estimation. In: Berthold, M., Shawe-Taylor, J., Lavrač, N. (eds.) IDA 2007. LNCS, vol. 4723, pp. 70–80. Springer, Heidelberg (2007)
6. Hinneburg, A., Keim, D.A.: An Efficient Approach to Clustering in Large Multimedia Databases with Noise. In: Proc. Int'l Conf. on Knowledge Discovery and Data Mining (KDD), New York, USA, pp. 58–65 (August 1998)
7. Loh, W.-K., Moon, Y.-S., Park, Y.-H.: Fast Density-based Clustering Using Graphics Processing Units. IEICE Trans. Information and Systems (2014) (accepted to appear)
8. Xu, X., Jäger, J., Kriegel, H.-P.: A Fast Parallel Clustering Algorithm for Large Spatial Databases. Data Mining and Knowledge Discovery (DMKD) 3(3), 263–290 (1999)

A Bandwidth Allocation Mechanism in Mobile P2P Streaming in the Wireless LAN

Geun-Hyung Kim

Department of Digital Media Engineering, Dong-Eui University,
Department of Visual Information Engineering, Dong-Eui University,
995 Eomgwang-Ro, BusanJin-Gu, Busan 614-714, Korea
geunkim@deu.ac.kr

Abstract. The proliferation of mobile smart device with multimedia capabilities accelerates the IP-based live video streaming or video on demand (VoD) in the wireless networks. In addition, the IEEE 802.11 standard wireless networks have been developing to handle glowing demands of new applications like video streaming. Mobile peer to peer (P2P) streaming architecture has been considered as a promising way to distribute media streams over the wireless networks because of its scalability and efficiency. In the IEEE 802.11 wireless network, the AP or node handles packets stored in the queue in the same way. Therefore, this operation causes the performance degradation of mobile P2P streaming. In this paper, we investigate the effect of overlay topology on the overall throughput and propose the adaptive bandwidth allocation mechanism for mobile P2P streaming system.

Keywords: mobile P2P streaming, adaptive bandwidth allocation, IEEE 802.11 wireless LAN.

1 Introduction

With the proliferation of mobile smart device with multimedia capabilities and the wide deployment of IEEE 802.11 wireless local area network (WLAN), the demand on the IP-based mobile live video streaming or VoD in the WLAN increases greatly. The mobile video traffic has been already dominating network usage and it will take up 90 % of mobile traffic in this year in accordance with the Cisco's prediction[1]. Therefore, the investigation on the efficient dissemination of video stream in the wireless networks is an important issue.

In order to deliver the video content efficiently, several delivery architectures for live video streaming, such as IP multicast, client/server architecture, content delivery network (CDN) or P2P architecture, have been discussed in the wired domain. Among these architectures, P2P architecture has been proved tremendously effective in large scale video streaming services. In P2P streaming architecture, peers send the data chunks, that they received, to others. Cooperative video sharing among peers reduces the load of centralized streaming servers. P2P streaming architecture has attracted lots of researchers' attention

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_99,

in recent years due to its scalability, lower infrastructure costs, and traffic load distribution in the network. Current P2P streaming systems have been developed for the case, where users cooperate at their PC, which is connected to the ISP via reliable and broadband access network. P2P streaming systems tend to use the network bandwidth very aggressively and have no or little preference to exchange the data chunks among nearby peers[2][3]. However, users want to use their streaming applications with any devices in any place with the advent of ubiquitous environment. Thus, new researches for applying P2P streaming in the wired domain into the wireless networks is required. Mobile P2P streaming architecture should provide sufficient quality of experience (QoE) and efficient wireless resource utilization.

In the P2P streaming system, peers are connected with each other in overlay networks to disseminate the data chunks in which the video stream data is split up. In general, peers in the P2P streaming system, peers are classified into seeders and downloaders. The seeders are peers with a good capacity in terms of bandwidth and take care of the initial content distribution. The downloaders obtain the data chunk to the seeders or other downloaders. The downloaders send the data chunk to other downloaders too[4]. Peers keep state information related to the connection that they have. The state information includes the buffer maps, that indicate the chunks a peer currently holds, and the list of unchocked peers. Peers periodically select the set of unchoked peers to which they send the data chunks. Since the available bandwidth on each peer is nondeterministic, it is hard to guarantee the upload and download bandwidth constantly. In addition, rate adaptation for IEEE 802.11 networks are applied to determine the optimal data transmission rate most appropriate for current wireless channel conditions.

In this paper, we investigate the effect of rate adaptation on overall service quality and propose the bandwidth allocation mechanism for serving the corresponding peers by considering the data transmission rate. To the best of our knowledge, this work is the first to consider more realistic scenario to achieve high performance on the mobile P2P streaming.

The remainder of this paper is organized as follows. In section 2, we briefly discuss the related work. In section 3, we describe the mobile P2P streaming architecture and the problem that we have investigated. We present the experiment results in section 4 to evaluate the our algorithm. Finally, the conclusion is drawn in section 5.

2 Related Work

In spite of the fact that a large amount of scientific studies are performed, there is relatively little effort regarding the mobile P2P streaming. Venot *et al.*[5] implemented on-demand P2P streaming system on JXTA overlay. Their system can not play the video contents progressively. It can only play them only after the content downloading is finished. Jinfeng *et al.*[6] proposed mobile P2P live

streaming system in which the mobile devices have multiple network interfaces. In their system, seed peers receive streaming data via cellular network and share the data with other peers in the overlay networks consisting of other wireless networks. Their study mainly focused on the energy consumption on the mobile devices. Leung *et al.*[7] developed a protocol for the P2P dissemination of multimedia content to mobile devices on the cellular network. Similar to previous works, only few peers obtain video descriptions from base stations, and then they share stream data to other peers using a free broadcast channel. At that time, the researches on the mobile P2P streaming are based on the cellular networks. After the advent of smart device and the wide deployment of wireless LAN, Peltotalo *et al.*[8] proposed P2P streaming system using the partial RTP stream concept on a bluetooth-based overlay network which groups peer into clusters according to their proximity using RTT value between peers as a criteria for the cluster selection. Previous works focused on how to remove bottlenecks in cellular networks and how to make P2P networking on JXME, in the case when mobile devices had other network interfaces except cellular network. Eittenberger *et al.*[3] presented the architecture of mobile P2P streaming prototype for the Android, implemented P2P streaming application for Android compatible devices, and evaluated the feasibility on a rather small scale. We*et al.*[9] extended the Goalbit P2P streaming software[4] to IEEE 802.11 *N* wireless network and proposed an adaptive unchoking algorithm to solve the problem caused by existing unchoking algorithm in the wireless LAN. Mobile P2P streaming in the wireless LAN encounters different network dynamics compared to fixed P2P streaming in the wired networks.

3 Problem Statement

Nowaday, IEEE 802.11 based Wireless Local Area Network (WLAN) has been widely deployed to provide ubiquitous network access and it will be prominent technology to be used for mobile P2P streaming service. Next generation of wireless LAN technologies also endeavor to provide broadband download capacity. However, the bandwidth of wireless LAN changes dynamically depending on the symmetric allocation and the mobility of peers. The fluctuating link bandwidth negatively affects the performance of mobile P2P streaming. As discussed in our previous work[9], the transmission delay increases when a peer moves away from the access point (AP), since the available bandwidth on the peer decreases.

In mobile P2P streaming architecture, the transmission control protocol (TCP) is used to exchange the chunk data among peers. In general, a peer sends chunk data to multiple peers and receives chunk data from multiple peers. The TCP throughput on mobile node is not guaranteed in the infrastructure wireless LAN, when several nodes coexist. Especially the throughput of a node is influenced by other nodes though it does not move and the other nodes move. In order to investigate the effect of peers' position on the throughput in the mobile P2P streaming, we consider the two cases: 1) one peer sends chunk data to two peers which are at the same distance from the AP and 2) one peer sends chunk data

to two peers which are at the different distance from the AP. In addition, we adopt automatic rate fallback (ARF) and two propagation models, LogDistance and Nakagami to investigate the effect of propagation model and rate control mechanism in the mobile P2P streaming. In the first case, the distance between the sender node and the AP is 10 m. The distance between and the receiver node and AP changes 15m, to 40m , 65m, 90, and 115 m.

Fig. 1 shows the TCP throughput at the recipient, when two nodes are at the same distance from the AP. The throughput of two nodes are similar and decreases when the distance from the AP increases. Nakagami propagation model brought more channel errors than LogDistance propagation model. Therefore, the throughput of LogDistance is higher than that of Nakagami.

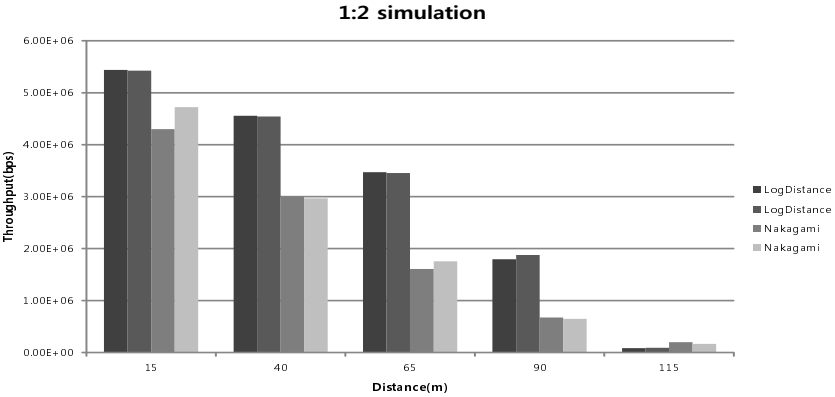


Fig. 1. The TCP throughput when two nodes are the same position

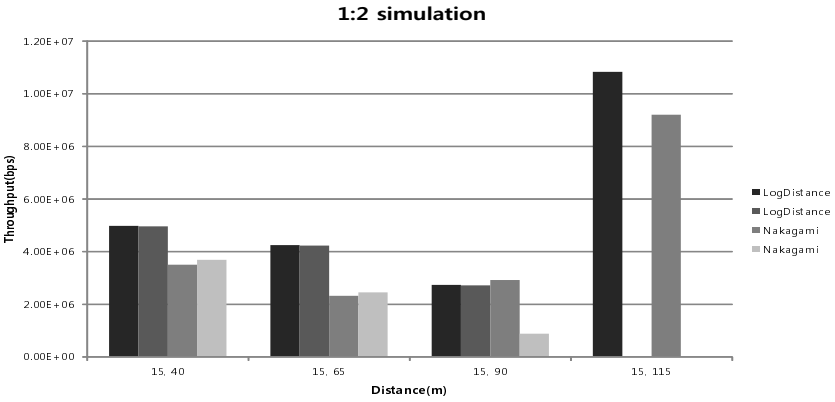


Fig. 2. The TCP throughput when two nodes are at the same position

In the second scenario, the two recipients are at the different positions. The first recipient is at 15 m from the AP. The position of second recipient changes from 40 m to 65 m, 90 m, and 115 m. Unlike the result in Fig. 1, the throughput of the first recipient decreases gradually.

4 Proposed Bandwidth Allocation Mechanism

As described in section 3, the total transmission throughput is degraded when peers send the chunk data to multiple peers that are at the different locations, e.g., the different distance from AP. In the P2P streaming system, peers select the group of peers by unchoking algorithm among the list of peers in the response messages from the tracker. After selecting the group of peers, peers send control messages and exchange chunk data with each other. When peers send chunk data to the unchoked peers in the previous P2P streaming system, peers send the data to unchocked peers in the group by the same rate. However, the link capacities to recipient peers are not the same and time-varying. This characteristic in the wireless LAN brings about the inefficient content distribution and performance degradation of overall mobile P2P streaming system.

The main idea of proposed algorithm is that peer allocates adaptively its upload bandwidth to corresponding peers according to the link capacity to mitigate the performance degradation of overall P2P streaming. In the proposed algorithm, the link capacity is estimated using the average chunk data delivery delay. Fig. 4 shows the measured available bandwidth and chunk delivery delay[9].

In the proposed algorithm, peers allocate adaptively bandwidth based on the link capacity estimated by measured delay. Fig. 4 shows the result of our proposed algorithm that is different from that of Fig 2. Overall throughput in Fig. 4

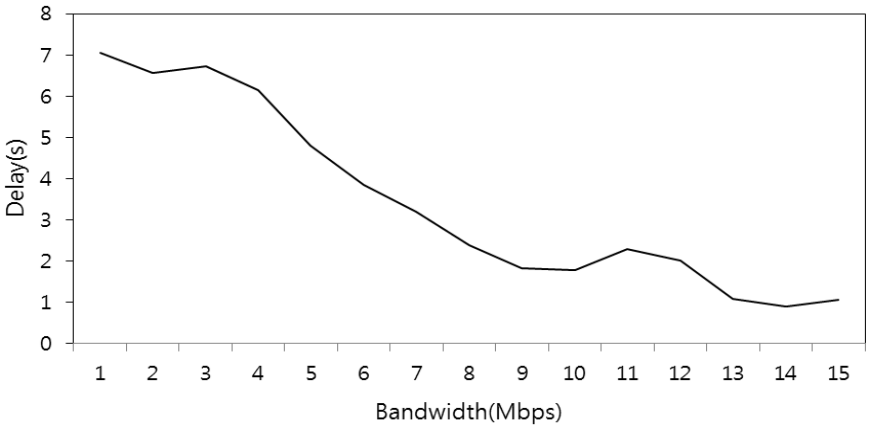


Fig. 3. Chunk delivery delay vs. an available bandwidth

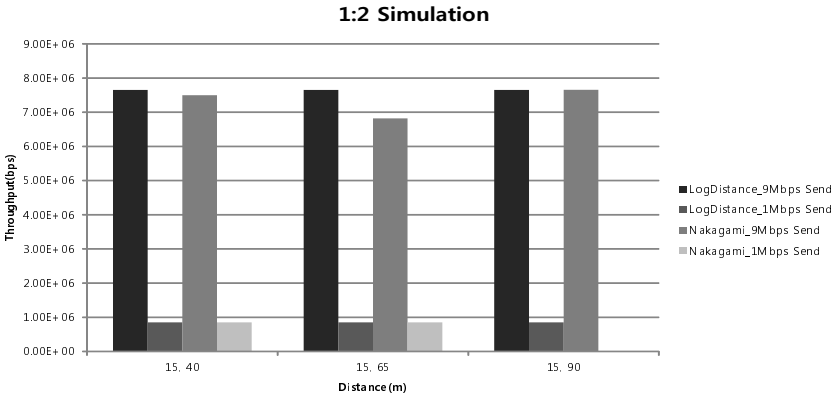


Fig. 4. The TCP throughput in an adaptive bandwidth allocation mechanism

is higher than that in Fig. 2. Adaptive bandwidth allocation results in efficient utilization of link capacity in wireless LAN and higher throughput in mobile P2P streaming.

5 Conclusion

Mobile P2P streaming architecture has been considered as a promising way to distribute media streams over the wireless networks because of its scalability and efficiency. In the wireless LAN, AP or node forwards packets stored in the queue sequentially, though packet are classified into several QoS classes that define the priority among traffic classes. Therefore, this kind of operation causes the performance degradation in mobile P2P streaming. In this paper, we investigate the effect of overlay topology on the overall throughput and propose the adaptive bandwidth allocation mechanism for mobile P2P streaming system. In our adaptive bandwidth allocation algorithm, peers allocate their upload bandwidth according to the distance of peers in the swarm. As the future work, we will extend our algorithm to AP in order for AP to allocate downlink bandwidth according to the distance of recipient peers.

Acknowledgments. This research was supported by Basic Science Research Program through the National Research Found (NRF) funded by Ministry of Education, Science and Technology (NRF-2010-0025069).

References

1. Talbot, D.: Broadcast Video will soon be packed into Smartphone Signals (May 2013), http://www.technologyreview.com/news/513311/broadcast-video-will-soon-be-packed-into-smartphone-signals/

2. Ciullo, D., Garcia, M., Horvath, A., Leonardi, E., Mellia, M., Rossi, D., Telek, M., Veglia, P.: Network Awareness of P2P Live Streaming Applications: A Measurement Study. IEEE Transactions on Multimedia 12(1), 54–63 (2010)
3. Eittenberger, P.M., Herbst, M., Krieger, U.R.: RapidStream: P2P Streaming on Android. In: 19th International Packet Video Workshop (2012)
4. Bertinat, M.E., De Vera, D., Padula, D., Robledo, P., Rodriguez-Bocca, P., Romero, P., Rubino, G.: Goalbit: The First Free and Open Source Peer-to-Peer Streaming Network. In: Proceedings of LANC 2009 (2009)
5. Venot, S., Yan, L.: On-demand mobile peer-to-peer streaming over the JXTA overlay. In: UBICOMM 2007, pp. 131–136 (2007)
6. Jinfeng, Z., Jianwei, N., Rui, H., Jianping, H., Limin, S.: P2P-Leveraged Mobile Live Streaming. In: AINAW 2007, pp. 195–200 (2007)
7. Leung, M.F., Gary Chan, S.H.: Broadcast Based Peer-to-Peer Collaborative Video Streaming Among Mobiles. IEEE Trans. on Broadcasting 53(1) (March 2007)
8. Peltotalo, J., Harju, J., Vaatamomen, L., Bouazizi, I., Curcio, I.D.D.: RTSP-based Mobile Peer-to-Peer Streaming System. International Journal of Digital Multimedia Broadcasting (2010)
9. Choi, H.-H., Kim, G.-H.: An Adaptive Unchoking Algorithm for Efficient Mobile P2P Streaming in Wireless LAN. In: Han, Y.-H., Park, D.-S., Jia, W., Yeo, S.-S. (eds.) Ubiquitous Information Technologies and Applications. LNEE, vol. 214, pp. 869–877. Springer, Heidelberg (2013)

An Indoor Location Tracking System Based on Wireless Sensor Networks and Marker-Based Fingerprinting Algorithm

Youn-Sik Hong[1] and Hye-Gyeong Jeon[2]

[1] Dept. of Computer Science and Eng., Incheon National Univ., Incheon, Korea
[2] AndTech Korea Co. Ltd., Incheon, Korea
yshong@incheon.ac.kr, hkjeon@andtechkor.co.kr

Abstract. In this paper, a method of moving control of an automatic guided vehicle (AGV) through marker recognition is presented. The existing AGV location tracking system using infrared wireless sensor nodes and landmarks have faced at two critical problems. First of all, since there are many windows in the special circumstance such as a crematorium, they are going to let in too much sunlight in the main hall which is the moving path of AGVs. Refraction and/or reflection of sunlight disturb the correct recognition of landmarks. The second one is that a crematorium has a narrow indoor environment compared to typical industrial fields. Particularly when it changes its direction to enter a designated furnace the information provided by the landmarks cannot be utilized to estimate its location because the rotating space is too narrow to get them. To overcome such difficulties that cannot access data in a wireless sensor network, a relative distance from a marker to an AGV will be used as a fingerprinting for location estimation. Compared to the conventional method which adopts received signal strength (RSS) as fingerprinting, our proposed method results in highly reliable estimation.

Keywords: Location-Tracking, Infrared Sensor, Automatic Guided Vehicle (AGV), Marker Recognition, Fingerprinting Algorithm.

1 Introduction

Unmanned moving vehicles can be classified into railway guided vehicle (RGV) and automatic guided vehicle (AGV). A RGV moves along the fixed track like railroad. An AGV, on the other hand, moves itself depending on navigation information obtained by using magnetics, laser beams, ultrasonic waves and gyroscope-induction, etc. Nowadays with easier installation and lower operational costs AGVs have been gained more popularity in the areas of factory automation.

In an old fashioned crematorium of our nation, a manned operating vehicle is used to transport a dead body to a designated furnace (also known as *cinerator*). However, with the trend of modernization of crematorium facilities as well as changes of the people's attitude to funeral ceremonies, it should be required to develop an AGV to

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_100,

take over such an unmanned vehicle. By doing so, it can increase efficiency and reduce costs by helping to automate such transportation process.

An AGV itself can transport a dead body safely from the loading place to its rightful destination, *i.e.*, a designated furnace, or vice versa as shown in Fig.1. Notice that it can endure high temperature (more than 300 degrees) near the furnace. The floor in a crematorium is typically made of marbles and thus very slippery. Even though its speed is very slow (less than 1Km/hour), it is necessary to have a precise control for an AGV movement not to slip depending on its weight, where loading dead body means more weight.

The entire moving path is divided into two distinct sub-paths: straight-line section and rotating section. In the straight line section, it can move forward until it arrives at the start location of the rotating section. At that location, it starts changing its direction towards the designated furnace. A well-known method of embedding guided magnetic bars in the floor is impractical in a modern crematorium because the floor is made of expensive materials, i.e., marbles. In addition, it has no flexibility. We presented an improved method of a moving control of an AGV using infrared wireless sensor nodes and landmarks [6]. With this approach it emits infrared lays to the landmarks attached to the ceiling to obtain a moving direction.

However, there are two critical problems. Since there are many windows in a crematorium, they are going to let in too much sunlight in the main hall which is the moving path of AGVs. Refractions and/or reflections of sunlight disturb the correct recognition of landmarks. Thus, the first problem is that there exists both dead zone and overlapped zone in the moving path of AGVs.

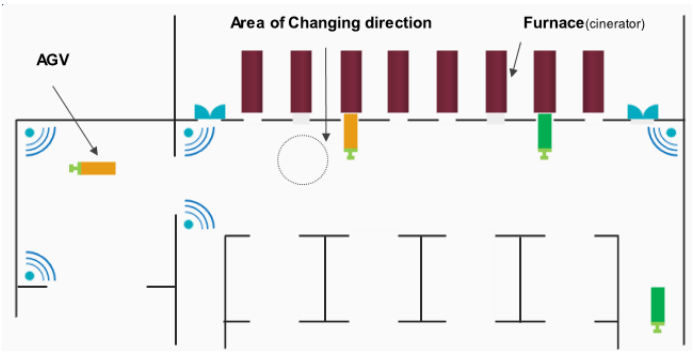


Fig. 1. A funeral cortege headed for a designated furnace using AGVs in a crematorium

The second one is that a crematorium has a narrow indoor environment compared to typical industrial fields. Particularly when an AVG changes its direction to enter a designated furnace the information provided by guided sensors like landmarks cannot be utilized to estimate its location because the rotating space is too narrow to get them. To resolve this, the presentation approach uses a modified-location fingerprint algorithm, which uses distance-to-marker information gathered at multiple locations in order to estimate an AGV's coordinates. By recognizing markers attached to the

wall in front of an AGV, its location can be estimated by clustering the Euclidean distance between measured fingerprint vector and location fingerprint vector.

Assume that the AVG at the time *t1* is located farthest away from the marker and at that time the image size recognized by its vision sensor is l_{t1}. In addition, the relative distance between the AGV and the marker is *d1*. At *t2* (*>t1*), it approaches the marker closely (*d2<d1*), and thus the image size l_{t2} recognized is larger than l_{t1}. In other words, the image size to be recognized becomes inversely proportional to the relative distance.

Received signal strength (RSS) may be considered as the simplest and cheapest method amongst the wireless distance estimation techniques, since it does not require additional, costly hardware for distance measurements. However, in practice, the radio signals are highly variable and unstable under the influence environment noises, obstacles, interference and the types of antenna. The signal strength is too sensitive to the harsh and dynamic environments, indoor environments due to multipath fading and interference which causes uncertainties in the radio communications amongst the wireless nodes. This condition influences the positioning accuracy. With the image size recognized as a fingerprint instead of RSS gives more reliable position estimation. We call it a relative distance fingerprinting.

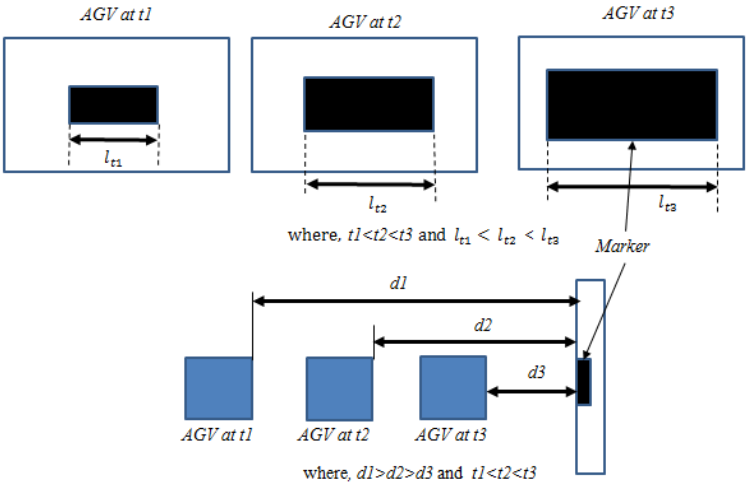


Fig. 2. The image size recognized with respect to each AGV move

Based on the starting and the ending location which is already known, the wheel moving trajectories of an AVG are computed first and then it will be controlled to move along the computed trajectory. Therefore, before it reaches the starting position of the rotating section, it can reduce its speed and then be prepared for changing its direction through marker recognition. By doing so, there is a very little chance that overrunning the target position will be happened.

The rest of this paper is organized as follows: Section 2 summaries related works. We describe a marker detection algorithm using object contour extraction in Section 3. Section 4 presents simulation results to verify the validity of the proposed methods. Finally Section 5 concludes our works

2 Related Works

2.1 Infrared Sensors Based Indoor Location Tracking System

Matthai et al [2] proposed an approach to generate maps of RFID tags with mobile robots. They presented a sensor model that allows us to compute the likelihood of tag directions given the relative pose of the tag with respect to the robot. However, their position error is too high even though the average speed of 0.225m/s of the test robot are faster than our AGV. The method [3] of recognizing locations of a mobile robot by using barcodes as landmarks was proposed. However, this method has the weakness that the location estimation error may vary from centimeters to meters due to limitation of the performance of the recognizing instrument.

2.2 Landmark

For the successful navigation, an AGV has to estimate its current location with respect to its immediate surroundings. An artificial landmark is a very simple and powerful tool for self-localization in indoor environments. The landmark used here is composed of a 4x4 array of rectangles [7]. The square in each corner of the landmark is used for indicating its direction. The ten squares represented as digits from 0 to 10 are used for calculating its ID.

2.3 Fingerprinting Algorithm

The radio-frequency fingerprinting requires multiple radio transmitters and receivers, deployed to provide overlapping coverage of an area, such as a floor in an office building [4, 5]. The sensor nodes, located at fixed infrastructure in order to model a physical space in terms of radio signal strength values by periodically broadcasting radio packets periodically to the air. The fingerprinting approach consists of a data collection process in which, the signal strengths at each fixed sensor node are measured and stored. Following the setup process, a mobile sensor node may be localized using the signal strengths measured at each fixed node. For localizing a mobile sensor node, the algorithm searches the matrix collected in the data collection phase, and find the fingerprint that best matches the signal strength observed. That is, the mobile node compares the observed signal energy with the recorded, and picks the location that minimizes the Euclidean distance as the location estimate. This technique is also referred to as nearest neighbor method.

3 A Position Estimation Using Relative Distance Fingerprinting

3.1 System Overview

The overall AGV system consists of three modules: PLC (Programmable Logic Control) module, main control unit and sensor modules. The PLC module controls the left and the right motor of the AGV to move or rotate its corresponding wheel. The primary function of the guided sensors detects landmarks attached to the ceiling. In addition, to avoid the collision with obstacles, the AGV is mounted with additional sensors like PSD (position sensitive detector) sensors and photo sensors for a precise guide with the approach to a designated furnace. Its current location will be reported to the application systems via wireless LAN.

3.2 Marker Identification

Note that since the marker is a planar object, we set the z coordinate to 0. That is, the marker is placed on the XY plane, centered at the origin. There are different ways to detect a marker in an image. The most simple and fast way is to do the following procedures:

Image Binarization: Binarize the image
The first step is to convert a colored image into a binary one. The simplest way to binarize an image is to convert it to a gray-scale image I_g, in which every pixel can take 256 possible values. Then the algorithm processes I_g in order to produce a binary image I_b, in which every pixel has only a value of 0 or 1. A pixel with a value greater than a fixed threshold T takes a value of 1 in I_b, otherwise it takes a value of 0.

Component Labeling: Do a component labeling algorithm
We use a local threshold algorithm. The local threshold algorithm uses an N by N window centered at each pixel of I_g in order to compute the threshold for that pixel. The algorithm computes the average value w_{avg} of the window. The threshold is computed using the formula $c \times w_{avg}$, where, $0 \le c \le 1$. This simple threshold algorithm is by far more robust to illumination problems than other threshold techniques. One of the advantages of constant threshold is that it produces images with more uniform regions of black or white pixels, lowering the number of connected components. This will speed up the component labeling algorithm.

Marker Verification: For every component, check if is a marker
After extraction of contours, the image was segmented with respect to the rectangles extracted. Using contour tree, the inclusion relationship is checked to verity the correctness of the markers.

3.3 A Relation between the AGV's Velocity and the Execution Time of Marker Recognition

To avoid a collision with the wall, the marginal distance between the marker attached to the wall and the current location of an AGV should be necessary. To measure the distance, a relation between the AGV's velocity and the execution time of marker recognition should be built and then a rotating point (RP) is determined while considering its rotating radius. . In addition, let *r1* (*r2*) be the distance between the inner (*outer*) wheel of an AGV and the center. Assume that *r2* = *r1* + *d*, where *d* is a non-zero positive constant.

Finding a Rotation Point of an AGV

Through marker recognition we can decide that an AGV stands a bit aside from the starting location to rotate. We call that location *rotation point* (RP). In Fig.3, the RP is represented as the circle. Let t_p and l_p be the time to be elapsed from the RP to just begin a rotation and the distance from it to the starting location, respectively. Notice that l_p is calculated by $l_p = v_c \times t_p$, where, v_c is the AGV's velocity.

Avoiding a collision with the wall (or an obstacle) would impose restrictive conditions on the RP. Let L and l_α be the total moving distance of an AGV during rotation and the space to protect the collision, respectively. During rotation, the following constraint should be satisfied to avoid the collision.

$$L \geq l_p + r_1 + l_\alpha \tag{4}$$

Therefore, l_p should satisfy the constraint $l_p \leq L - r_1 - l_\alpha$. It varies with the AGV's velocity v_c.

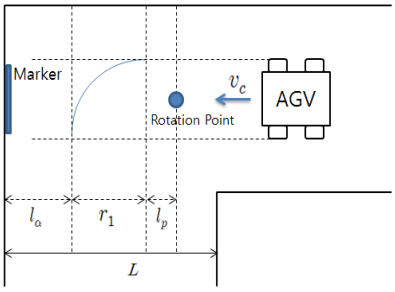


Fig. 3. The rotation point of an AGV

3.4 Location Estimation Using Relative Distance Fingerprinting

The fingerprinting approach consists of a data collection process in which, the image size recognized at each fixed reference point are measured and stored. The maximum distance between the marker and the camera of an AGV that still allows a successful detection strongly depends on the size of a marker. The location estimation using relative distance fingerprinting can be performed on a line-segment basis. As shown

in Fig. 4, given the set of *n* reference points, the whole line can be divided into *n-1* line segment. Each line segment the slope Δ can be obtained using two reference points and the fingerprints already measured. Notice that the slope of each line segment can be saved into the database together with the fingerprints.

Following the setup process, an AGV may be localized using the image sizes measured at each reference point. This will be done on a segment basis, where the segment can be determined as two consecutive fingerprints. The slope Δat *i*-th segment can be calculated as the following equation:

$$\Delta_{y_{i+1}=y_i} = \frac{fp_{i+1} - fp_i}{x_{i+1} - x_i}, \quad \Delta_{x_{i+1}=x_i} = \frac{fp_{i+1} - fp_i}{y_{i+1} - y_i} \tag{5}$$

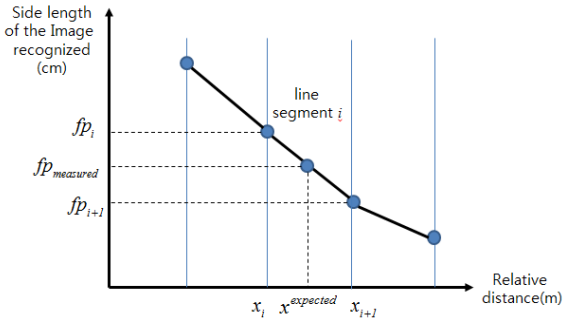


Fig. 4. A linear-wise location estimation of an AGV using relative distance fingerprinting

Given the image size measured $fp_{measured}$, the x-coordinate $x^{\exp ected}$ (*or* the y-coordinate $y^{\exp ected}$) of the estimated location for an AGV is obtained by using the following equation:

$$y^{\exp ected} = y_{\min} - \frac{fp_{\max} - fp_{measured}}{\Delta_{x_{i+1}=x_i}}, \quad x^{\exp ected} = x_{\min} - \frac{fp_{\max} - fp_{measured}}{\Delta_{y_{i+1}=y_i}} \tag{6}$$

where, $fp_{\min} = \max[\, fp_i, fp_{i+1}]$, $x_{\min} = \min[\, x_i, x_{i+1}]$, $y_{\min} = \min[\, y_i, y_{i+1}]$. Because an AGV is gradually closer to the marker, smaller x-coordinate (*or* y-coordinate) in the above equation is chosen as the reference value.

4 Experiments and Implementation

In the laboratory test, the mock-up AGV which embeds the ATmega128L as a microcontroller is used. Its CCD camera saves image as Motion JPEG format whose resolution is 160 by 120. The OpenCV libraries [8] are used for marker recognition. The number of reference points can be determined on both the AGV speed and the marker size.

4.1 Marker Size Versus the Maximum Recognizable Distance

A set of experiments to analyze the relation between the marker size and the maximum recognizable distance are performed. The single colored (red) square is used as the marker. Table 1 summarizes the marker size and the maximum distance to be successfully recognized. As you expected, as the marker size increases, so the recognition speed also becomes fast. In addition, with the 5cm increases of the marker size, the recognizable distance grows longer by 1m. In addition, marker recognition can be successfully performed within 1.3 seconds (average).

Table 1. The marker size and its maximum distance to be recognized

Marker size(m)	Maximum distance(m)	Processing time (average) (second)
10x10	1.5	1.6
15x15	2.0	1.3
20x20	3.0	1.2

4.2 Measurement of Relative Distance Fingerprinting

We have tested a validity of estimating the AGV's location and the remaining distance to the marker using relative distance fingerprinting. In this experiment, the marker size is set to 15cmx15cm. In addition, a general marker for pattern recognition is used. That means it can have various images inside the rectangle. Thus it gives AGV useful information. For example, the character *R* in the marker indicates that it should turn right on recognizing the marker. Table 2 summarizes the relation between the relative distance to the marker and the image size. Notice that the image size in the Table 2 means the image displayed on the screen of the mockup-AGV. As the relative distance becomes shorter, the image size increases.

Table 2. The relation of relative distance and its image size recognized

Relative distance(m)	Side-length of the image (cm)	Variation (cm)	percentage of the marker to the whole image
2.0	0.5	-	0.43
1.5	0.9	+0.4	1.39
1.0	1.5	+0.6	3.87
0.5	2.4	+0.9	9..91

4.3 Location Estimation Using Relative Distance Fingerprinting

The general marker whose size is 15cm x 15cm is used and every 50cm movement the mockup-AGV stores the image recognized. The maximum estimation error is 5cm

and the margin of the error is +0.025%. Typically, its estimated location is +1cm ~ +5cm ahead of the real location.

In addition, we have tested the accuracy of the AGV's velocity estimation using relative distance fingerprinting. At the time of 6.7 and 14.6 seconds, the relative distance between the AGV and the marker was measured at 2.25m and 1.35m, respectively. Thus, the AGV's velocity is calculated at 0.113m/second = 6.78m/minute. The real velocity is 6.75～6.80m/minute. So our estimation gives a high accuracy of the speed estimation.

4.4 An Improved Recognition Speed through Image Segmentation

To accelerate the processing time of the marker recognition, the original image is divided into a set of segments. A method of partial image extraction is applied to find the rectangle with the similar ratio of the original marker. The image extraction for the whole image took 170.6 milliseconds (average) to process. However, the partial extraction method took 22.2 (average) seconds only. Thus the processing time of the proposed method is 8-times faster than that of the full extraction methods.

5 Conclusion

For a more precise control of an AGV in a crematorium, we proposed a method of relative distance fingerprinting which uses the image size recognized as location estimation. The estimation can be done on a segment basis. This method will be used to assist the primary navigation method based on infrared wireless sensor nodes. These two distinct navigation methods are combined and integrated into the application systems which manage both moving control and location monitoring of the AGV. The main advantage of our systems is that they are highly flexible for on-demand delivery to any location. They also are quick to install, with less down-time for the crematorium.

Acknowledgements. This work was supported by the National Research Foundation of Korea (NRF) Grant 2013.

References

1. Yanying, G., Lo, A., Niemegeers, I.: A survey of indoor positioning systems for wireless personal networks. IEEE Communications Surveys & Tutorials, 13–32 (2009)
2. Matthai, P., Kenneth, F., Dieter, F., Dirk, H.: Mapping and Localization with RFID Technology. IRS-TR-03-014 (2003)
3. Huh, J., Chung, W.S., Nam, S.Y., Chung, W.K.: Mobile Robot Exploration in Indoor Environment Using Topological Structure with Invisible Barcodes. ETRI Journal 29(2), 189–200 (2007)
4. Kaemarungsi, K., Krishnamurthy, P.: Modeling of Indoor Positioning Systems Based on Location Fingerprinting. In: IEEE INFOCOM, pp. 1012–1022 (2004)

5. Bal, M., Xue, H., SHen, W., Ghenniwa, H.: A 3-D Indoor Location Tracking and Visualization Systems Based on Wireless Sensor Networks. IEEE Systems, Man and Cybernetics, 1584–1589 (2010)
6. Jeon, H., Hong, Y.: A Precision Control of an Automatic Guided Vehicle Using Double Landmark Recognition in a Crematorium. Applied Mechanics and Materials 278-280, 1537–1540 (2013)
7. Hagisonic web site, http://www.hagisonic.com/
8. OpenCV web site, http://opencv.org/

Design and Implementation of Customized Spatial Information Provider System for Chronic Disease Patients Based on PHR

Jae-Sung Shim[1] and Seok-Cheon Park[2,3,*]

[1] Department of Computer Science,
Gachon University, Republic of Korea
11sjs28@naver.com
[2] Department of Computer Engineering,
Gachon University, Republic of Korea
scpark@gachon.ac.kr
[3] College of IT, Gachon University, Bokjeong-dong, Sujeong-gu,
Seongnam-si, Gyeonggi-do, South Korea

Abstract. A system that provides tailored spatial information for chronic disease patients who require health self-management using Personal Health Record (PHR) information is proposed. PHRs, standard document structures and spatial information were analyzed for the system design; the structure of the messages were defined and based on the analysis results, exchanged in the customized spatial information provision system; and the exchange protocol and motion process of the customized spatial information provision system for the chronic disease patients were designed.

To realize the proposed system, Eclipse Juno, JDK, the Android platform, Java and JSP were used in the Windows 7 operating system. In addition, the message exchange structure based on the PHR standard document was defined to materialize a spatial information exchange protocol; and based on such protocol, a customized spatial information provision system that includes a service spatial information module, a user information management module and a tailored spatial information matching module was realized.

To verify the performance of the designed and realized system, tests were conducted on the exchange messages of the customized spatial information provision system and the mobility of the modules that provide the tailored spatial information for health self-management. The test results confirmed that the response messages were properly exchanged according to the requested messages, the service space where a user logged into the service using a smart device was recognized, and the matching modules for the service space management, user information management and customized spatial information were properly working.

Keywords: Spatial Information, Chronic Disease Patient, PHR, HL7.

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_101, © Springer-Verlag Berlin Heidelberg 2014

1 Introduction

As the number of chronic disease patients increases each year, paradigms in the medical service market are changing in favor of prevention services that improve lifestyles and promote sustainable health management by the recipients of medical services centered on the treatment and management of particular diseases[1,2].

In particular, chronic disease patients with hypertension and diabetes must be habituated to perform continuous health self-management to control their blood pressure and blood glucose level such as by controlling their diet (calories and salt) for each meal. Conventional medical services that target chronic disease patients use biosensors to monitor and record the patient's health status and to manage it through remote medical treatments and consultations. However, limitations were revealed in such services because they become unavailable when the patient goes out and acts in spaces such as restaurants, coffee shops and sports centers where it is difficult to conduct remote treatments and consultations [3].

To address such problems, a method is implemented that uses the spatial information generated and distributed in an indoor space such as a restaurant, coffee shop or sports center as tailored spatial information for health management.

Thus, in this study, to address the problems caused by the unavailability of health management services when chronic disease patients go out and act in spaces such as restaurants, coffee shops or sports centers, a customized spatial information provision system was designed and realized for various spaces that patients access such as restaurants, coffee shops, hospitals and sports centers, that utilizes the personal health record (PHR) information of the patients.

This paper is organized as follows. After the introduction in Chapter 1, the spatial information services for PHRs and chronic disease patients are analyzed in Chapter 2 (Related Studies). The message structures for the customized spatial information exchange based on the PHRs of the chronic disease patients are defined in Chapter 3, from which the system is realized and verified through a test in Chapter 4. Finally, the conclusion is presented in Chapter 5.

2 Related Studies

2.1 PHR(Personal Health Record)

A PHR is a tool that helps a person keep and manage all his health information or that of his or her family members throughout their lifetimes. Its development purpose differs from that of the EMR (Electronic Health Record) of medical institutions. The PHR is a tool that enables a person to manage his or her health record by strengthening his or her information rights, whereas the EMR improves access to hospital medical records and the efficiency of medical institutions in managing such records [4].

The capacity of the PHR to build a partnership between a medical service provider and a consumer is recognized, and the use of the PHR system is likely to reduce or

eliminate the need for repeated treatments or treatment processes and to cut costs and save time.

The PHR is also differentiated as a tool that can save legally obligatory records and records mixed with personal health records that can be identified as non-legal records. If the PHR is constructed and popularized, it will bring the advantages of perfecting health records, supporting U-Healthcare, educating and managing patients and expanding information exchanges among hospitals, among others [5].

2.2 Spatial Information Service for Chronic Disease Patients

The term spatial information includes all information available for the service that constructs and provides such information through a system that is conversed and integrated by all the information generated and distributed in space environments such as a diet list, treatment, exercise therapy (rehabilitation and training), contents, products and shop locations that are available in various spaces such as restaurants, hospitals, sports centers, cinemas and department stores, with information from other industries.

In this paper, smart devices that are being popularized fast and spatial information are used for the health self-management of chronic disease patients whose numbers are increasing each year, and a PHR-based customized spatial information provision system for chronic disease patients is proposed to enable them to perform effective health self-management when they access various spaces such as restaurants, hospitals and sports centers.

Diabetes management must adjust the diet of the patients by recommending the types and quantities of suitable foods and prohibiting foods that contain a large amount of sugar for normal blood sugar level maintenance, weight control, and prevention and delay of complications. In addition, as too much exercise may cause a hypoglycemic shock in the patient, it must be managed by determining the types and strengths of proper sports through consultations with a doctor. Table 1 shows the standard blood sugar level [6].

Table 1. Classification of diabetes

Status	Normal	Attention Required	Treatment Required
Empty stomach	<109	110-125	>126
One hour after a meal	<180	170-199	>200
Two hours after a meal	<140	140-170	>200

For hypertension patients, it is very important to skew the diet towards low-salt intake. To reduce the intake of sodium, which is the major cause of the rise of blood pressure, foods with higher sodium contents must be excluded when recommending the diet, and the intake of instant foods must be prohibited.

In addition, it is important to always remind the subject patients to reduce their sodium intake. Furthermore, too much exercise must be avoided as it can increase hypertension, the types and strengths of the proper exercises should be determined through consultations with a doctor, and the patient's activities must be managed so that they would not exceed such levels if he has a heart disease or complications such as nephritis. Table 2 shows a hypertension classification system based on the patient's blood pressure level [7].

Table 2. Classification of hypertension

Status	Normal	Attention Required	First-stage Hypertension	Second-stage Hypertension
Systolic Blood Pressure(mmHg)	<120	120-139	140-159	>160
Diastolic Blood Pressure(mmHg)	<80	80-89	90-99	>100

3 Design of the Customized Spatial Information Provision System for Chronic Disease Patients

3.1 System Outline

The proposed system provides tailored spatial information to chronic disease patients using the patient's PHR, such as the disease and gender of the patient, when the patient accesses various spaces such as restaurants, coffee shops, hospitals and sports center.

Chronic disease patients with hypertension, diabetes, etc. need to continuously manage their health such as by sticking to their diet and exercise therapy, depending on their disease. However, self-health management becomes difficult when the patient goes out. The proposed customized spatial information provision system that is based on the PHR recognizes the location information by using a QR code.

3.2 Definition of the Customized Spatial Information for Chronic Disease Patients

The term spatial information includes all information available for the service that constructs and provides the proposed system that in conversed and integrated with all information generated and distributed in space environments, such as the patient's diet list, treatment, exercise therapy (rehabilitation and training) and contents, and the products and shop locations in various spaces such as restaurants, hospitals, sports centers, cinemas and department stores, with information from other industries.

This paper restricted the spatial information that is provided for the health self-management of chronic disease patients to indoor environments and defined such spatial information by classifying them into two types: common spatial information or information that most spaces share, and distinctive spatial information that is

characteristically provided only in a particular space. Table 3 lists the common and distinctive spatial information.

Table 3. Common and distinctive spatial information

	Classification		Descriptions
Distinctive Spatial Information	Restaurants	Menu	Information on the foods sold in a restaurant
		Nutrient	Information on nutrient contents such as protein, sodium, sugar and saturated fat
		Price	Price information on each menu
	Hospitals	Department	Department information provided by a hospital, such as surgery and internal medicine information
		Administration	Information on names, volumes, etc. of the medicines prescribed for the patient
		Precautions	Information on the allergies, blood pressure, blood sugar level, etc. of the patient
	Exercise And Rehabilitation Centers	Exercise Therapy	Exercise titles, energy consumption, training regions, exercise methods
		Rehabilitation Therapy	Disease name, treatment regions, treatment levels, improvement status, etc.
		Precautions	Precaution information on using a space
Common Spatial Information	Business Names		Business names of spaces (OO restaurant, XX gymnasium, etc.)
	Business Types		Restaurant, hospital, sports center, etc.
	Business Address		Information on the location of a space
	Phone Number		Contact phone number of a space
	Business Representative		The representative for a space
	Spatial ID		Spatial identification information for a space

While common spatial information is general information that is not related to the PHR information of the patient and includes basic information on a given space, distinctive spatial information may differ from it in terms of the patient's PHR information and the output information produced according to the information selected and entered by the patient.

3.3 Motion Process of the Customized Spatial Information Service for Chronic Disease Patients

In the proposed system, if a chronic disease patient completes the user registration to the system in advance, he or she can receive customized spatial information on the corresponding service space without entering additional information separately when accessing various service spaces, because the system receives and manages the health status of the patient, including his or her disease, age, gender, allergies and vital signs.

The motion process of the entire system that provides the service upon the request of a non-registered chronic disease patient is designed as shown in Figure 1.

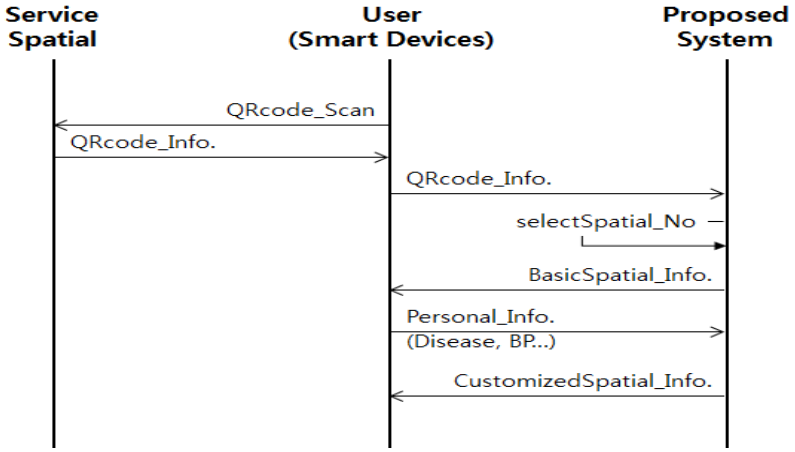


Fig. 1. Motion process of the customized spatial information service for a non-registered chronic disease patient

The motion process of the proposed system when a chronic disease patient is not registered as a user of the proposed system is as follows.

①The user scans the QR code in the service space using his or her smart device.

②The QR code information that is recognized by scanning the QR code is transmitted to the proposed system.

③A search will be executed on the registered service space through the information on the QR code.

④When the search is completed, the proposed system provides basic spatial information on the service space to the user.

⑤The user confirms the basic information, including his or her address, contact address, and the type of the service space, through the basic spatial information on the transmitted service space.

⑥The user enters his personal information, including his or her age, gender and disease, and transmits it to the system so that he or she will be provided customized spatial information.

⑦The proposed system provides customized information on the searched service space in accordance with the received user health information, including on the user's disease, gender, age and allergies.

In the proposed system, the registered users and non-registered users of the system receive the spatial information required for their health self-management in various space environments by using their PHR information, as shown in Figures 2 and 3

4 Realization of the Customized Spatial Information Provision System for Chronic Disease Patients

4.1 Module Structure of the Customized Spatial Information Provision System for Chronic Disease Patients

The PHR-based customized spatial information provision system for chronic disease patients that is proposed in this paper includes a service spatial information management module, a user information management module, and a customized spatial information matching and exchanging module for fabrication of the user PHR data and the data on the space where the user accessed the system, into tailored information through data matching in the proposed system, and for provision of such information to the patient. Figure 2 presents the structure diagram of the proposed system.

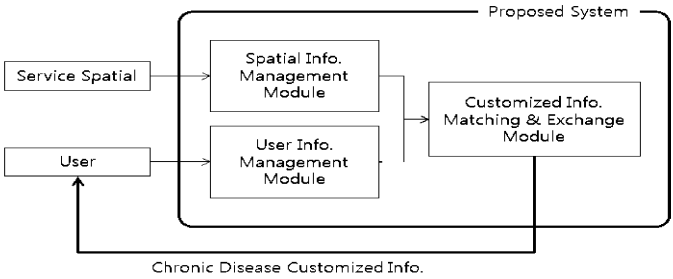


Fig. 2. Structure diagram of the proposed system

4.2 Realization of the Customized Spatial Information Provision System for Chronic Disease Patients

This chapter describes the test that was performed to verify if the system would work properly in a mobile environment (Galaxy Note II) where the design described in

Chapter 3 was realized. The test confirmed whether or not the system recognizes that a registered user is accessing Position 001 by scanning the QR code, and properly receives the customized spatial information according to the user's health status. Figure 3 shows the realization motion screen (for a registered user) of the proposed system.

Figure 3-① shows the Log-in Page, where a new user registers as a user by clicking the Subscribe button at the bottom if he or she wants to subscribe to the service and then entering his or her log-in ID and password (PW) and clicking the Confirm button.

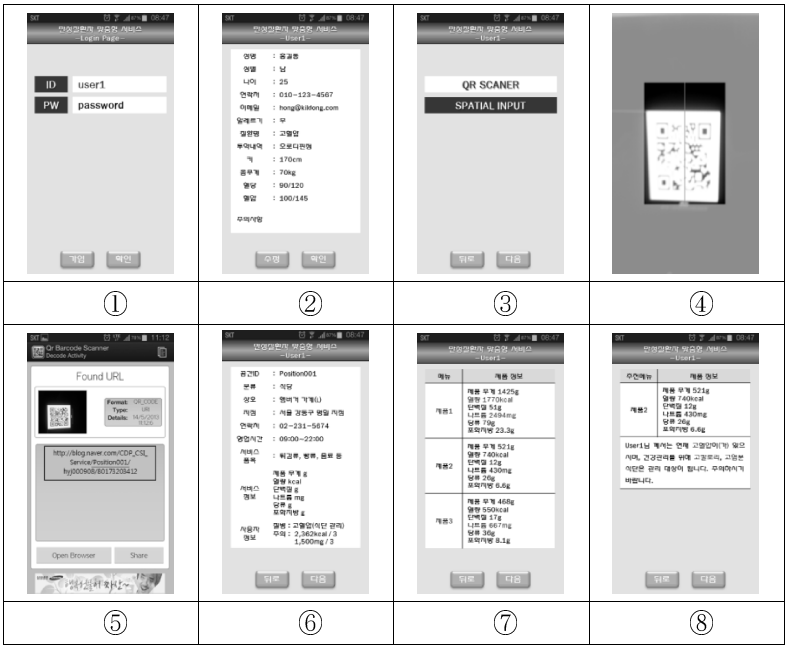


Fig. 3. Realization motion screen of the proposed system (for a registered user)

When the registered user enters correctly his or her ID and PW, the screen will show the page in Figure 3-②, where the user can confirm his or her registered basic information and can modify or delete any information by clicking the Modify button at the bottom. When the user finds the confirmed information correct, he or she goes to the next page by clicking the Next button.

Figure 3-③ shows the process whereby the user enters the Spatial ID so that the system would recognize the space where he or she is at. The process allow the user to either scan the QR code or directly enter the Spatial ID. If the user chooses to scan the QR code, the QR Barcode Scanner application is activated as shown in Figure 3-④ so that the user could scan the QR code.

When the QR code is successfully scanned, the page in Figure 3-⑤ that prompts the user to confirm the QR information on the space shown in the marked quadrangle is loaded, and the Spatial ID of the corresponding space is underlined.

Once the QR code is properly recognized, the system provides the basic information on the corresponding space (i.e., the Spatial ID, address, available service information, user precautions, etc.), as shown in Figure 3-⑥. The registered user can confirm his or her disease information and precautions in the user information section, but non-registered users have to enter the information other than their personal information.

After the basic information is confirmed, the detailed information on the corresponding space is presented as shown in Figure 3-⑦. As the space in the test environment is a restaurant, the menu (products) of the space is presented on the left side of the screen, and its spatial information appears on the right side of the product name, wherein the information in red indicates the products that the current user should use with care.

According to the information shown in Figures 3-② and -⑥, the user in the test case had hypertension and had to control his or her diet. In addition, the screen shows the daily recommended calories and sodium intake with a word of caution, and identifies product 1 (in red) as exceeding the recommended calorie and sodium level and product 3 as having too much sodium, in accordance with the diet information from the space where the user was situated.

Once the information on each product is confirmed, the user clicks the Next button at the bottom to check the information on the recommended menu, as shown in Figure 3-⑧. A menu will be recommended that has products that do not contain the user-restricted foods, depending on his or her health status.

5 Conclusion

Conventional medical services that target chronic disease patients use biosensors to monitor and record the patient's health status and manage his or her health through remote treatments and consultations. However, such conventional services have limits, because health management becomes unavailable when the patient goes out and act in spaces where it is difficult to conduct remote treatments and consultations, such as in restaurants, coffee shops and sport centers.

Therefore, in this paper, a PHR-based customized spatial information provision system for chronic disease patients, whose numbers are increasing yearly, was proposed, with which the patient can be provided the tailored spatial information that he or she requires to manage his or her health in an indoor space such as a restaurant or coffee shop, using his or her PHR information.

In this paper, customized spatial information exchanging protocols were designed for registered users of the customized spatial information provision system, non-registered users, service space registrations and spatial information exchanges. In

addition, based on the aforementioned protocols, the motion process of the customized spatial information provision system for chronic disease patients, which includes the service spatial information management function, the user information management function and the customized spatial information matching function, was designed.

The proposed system was realized using the Eclipse Juno, JDK, Android Platform, Java and JSP languages on the Windows 7 operating system. The realized system materialized the spatial information exchange protocols for the proposed system's registered users, non-registered users, service space registrations and spatial information exchanges. Based on such protocols, the service spatial information management module, the user information management module and the customized spatial information matching module were materialized.

In addition, for the mobility tests on the functions that provide the customized spatial information for health self-management, the proposed system was divided into the following functions: log-in, service space recognition through the QR code, user PHR information analysis and service spatial information provision according to such analysis. On the basis of the aforementioned classification, it was confirmed that the proposed system correctly recognizes the service space from the QR code when a user logs into the service through his or her smart device, and that the modules for service space management, user information management and customized spatial information matching worked properly in accordance with the diseases and requests of the user.

References

1. Shin, H.-M.: The future of health care services and digital hospital. MicroSoftware (2010)
2. Kang, S.-U., Lee, S.-H., Ko, Y.-S.: U-Healthcare coming of age. CEO Information, Samsung Economics Research Institute: SERI (2007)
3. Park, D.-K., Kim, J.-H., Kim, J.-K., Jung, E.-Y., Lee, Y.-H.: U-health Service Model for Managing Health of Chronic Patients in Multi-platform Environment. Journal of the Korea Contents Association 11(8), 23–32 (2011)
4. Park, Y.-M., Oh, Y.-H.: A Study on The Integration of Healthcare Information Systems based on SOA for PHR services. The Institute of Electronics Engineers of Korea - Telecommunications 48(2), 29–35 (2011)
5. Lim, C.-G., Rho, K.-T.: Design of Integrated Medical Information System based on XML. The Journal of The Institut of Internet, Broadcasting and Communication 10(2), 167–172 (2010)
6. Kang, G.-J., No, S.-Y., Ryu, H.-S., Lee, H.-S., Choi, S.-S.: Easy to learn Nutrition Assessment. In: SOOHAKSA (2011)
7. American College of Sports Medicine, ACSM's Resource Manual For Guidelines For Exercise Testing and Prescription 6th Edition, Lippincott Williams & Willims (2009)

A Scalable and Distributed Electrical Power Monitoring System Utilizing Cloud Computing

Ryousei Takano, Hidemoto Nakada, Toshiyuki Shimizu, and Tomohiro Kudoh

Information Technology Research Institute,
National Institute of Advanced Industrial Science and Technology (AIST)
1-1-1 Umezono, Tsukuba, Ibaraki 305-8568 Japan
takano-ryousei@aist.go.jp

Abstract. This paper proposes a scalable and distributed electrical power monitoring system utilizing cloud computing. This system collects power usage at measurement points geographically distributed over different locations, stores data on the cloud and provides a single unified view of power usage through a simple REST API. A system with 620 measurement points covering a server room and a clean room has been successfully installed at our campus, and we have operated it since the third quarter of 2011 to charge electricity bills and evaluate the power efficiency of data processing middleware. We have demonstrated that the proposed system can be smoothly scaled out based on the needs. This result provides an insight that cloud computing makes a power monitoring system elastic and cost-effective.

Keywords: Cloud computing, Electrical power monitoring system, Visualization System.

1 Introduction

The power consumption of data centers and networks becomes an issue of vital importance to IT industries. Especially in Japan, we faced planned power outages due to power shortages caused by the Great East Japan Earthquake on March 11, 2011. Even now the effort to save electricity continues. Real-time visualization of power consumption per segmented unit, such as a power distribution board and a breaker, is becoming essential to planning finer electricity savings. The measurement points may be distributed over several locations. For collecting and processing a large amount of distributed measured data, the total cost and scalability of a system have become key issues.

Utilizing cloud computing, we have shown that the total cost of a power monitoring system can be reduced, and the system can be easily scaled out. A system with 620 measurement points has been successfully installed at our campus. This system also provides a simple REST API with users. This API is able to utilize for several purposes such as expense billing and evaluation of green data center technologies.

The rest of the paper is organized as follows. Section 2 shows the design of the proposed power monitoring system. Section 3 describes the implementation,

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_102,

and Section 4 shows the demonstration of our campus system, and discussion of the lessons learned from the operation experience. Finally, Section 5 summarizes the paper.

2 AIST Electrical Power Monitoring System

2.1 Overview

We propose a scalable and distributed electrical power monitoring system utilizing cloud computing. The proposed system consists of four components: power measuring unit, data collecting unit, data store server and applications, as shown in Figure 1. A power measuring unit measures the electricity consumption from a power line. A data collecting unit gathers the above data from multiple power measuring units and pushes power usage data to a data store server on the cloud. A data store server is running on the cloud, e.g., Google App Engine [3], and it also provides a simple REST API to allow users to develop applications easily. An application retrieves and processes data, e.g., visualization and billing service.

The proposed system enables us to "start small and go big" by utilizing cloud computing. A data measuring unit has 4 sensors; a data collecting gathers data from up to 32 data measuring units. Therefore, a single data collecting unit supports to observe 128 measurement points and it can be incrementally installed on demand. A data store server can also scale up according to the workload. Moreover, the proposed system provides a single unified view of electricity consumption with users even if measurement points are distributed over several locations because the data store server is theoretically centralized.

2.2 REST API

To communicate among system components, we employ a simple power monitoring REST protocol with JSON data formats. Figure 2 shows the relationship between the REST protocol and the system components. This protocol provides users with update, get, and query operations, and it also provides set and get operations to configure a data collecting unit.

Table 1 shows the REST API list. The application identifies a resource using the URI, such as `http://example.appspot.com/$path`. It operates some requests using the HTTP methods, including GET and PUT. The message exchanged is represented as a JSON object. The detail of JSON data format shown in Appendix A. For instance, an application can obtain power usage of each measurement points, i.e., probes, for the last minute by accessing `http://example.appspot.com/latest` with the GET method as follows:

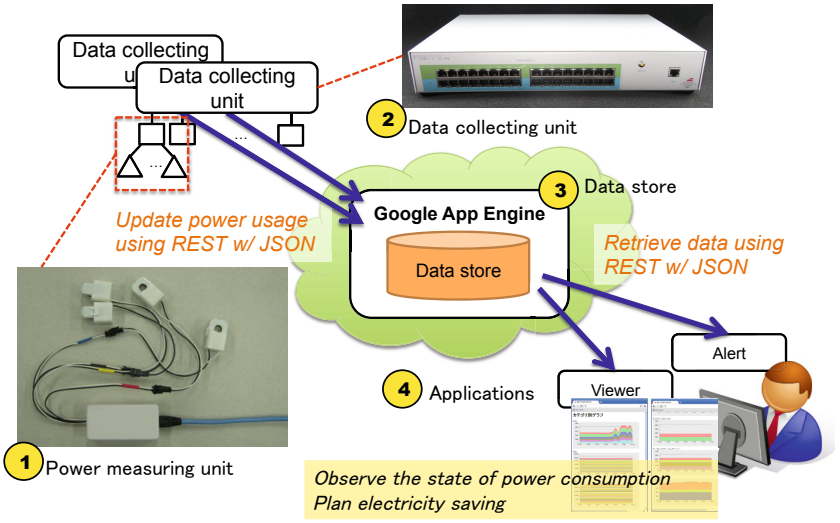


Fig. 1. An overview of a scalable and distributed electrical power monitoring system. This system consists of four components: 1) power measuring unit, 2) data collecting unit, 3) data store server, and 4) applications

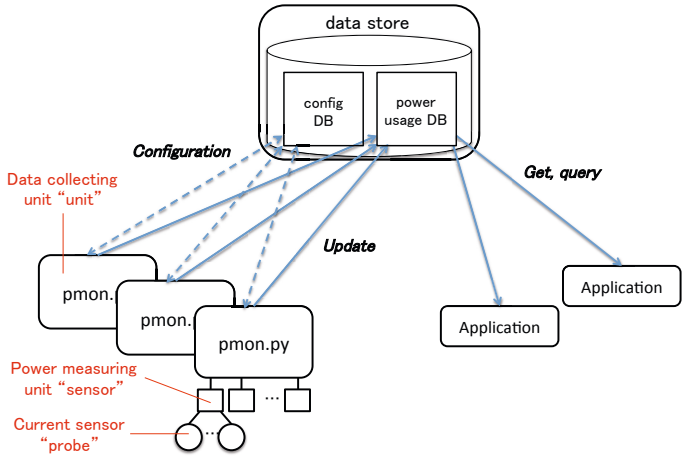


Fig. 2. Power monitoring protocol

```
{
    "power": {
        "aist.a02.2-12.1F.0:01141-1.p0": [1844.0], # single-phase 100V x 4 probes
        "aist.a02.2-12.1F.0:01141-1.p1": [82.0],
        "aist.a02.2-12.1F.0:01141-1.p2": [341.0],
        "aist.a02.2-12.1F.0:01141-1.p3": [55.0],
        "aist.a02.2-12.1F.0:01141-2.p0": [831.0],  # three-phase 200V x 2 probes
        "aist.a02.2-12.1F.0:01141-2.p1": [2779.0],
        :
    },
    "time": 1319837460,
    "timeStr": "201110290631"
}
```

Table 1. List of power monitoring REST API

path	method	description
/update	POST	Update data
/latest	GET	Get data for the last minute
/latest,COUNT	GET	Get data for the last N minutes
/summary.s/YYYYmmDDHHMMSS,N	GET	Get data for each second started from YYYYmmDDHHMMSS, for N seconds
/summary.m/YYYYmmDDHHMM,N	GET	Get data for each minute started from YYYYmmDDHHMM, for N minutes
/summary.h/YYYYmmDDHH,N	GET	Get data for each hour started from YYYYmmDDHH, for N hours
/summary.d/YYYYmmDD,N	GET	Get data for each hour started from YYYYmmDD, for N days
/query.s/LOC/YYYYmmDDHHMMSS,N	GET	Get data for locations that name started with LOC
/query.m/LOC/YYYYmmDDHHMM,N	GET	
/unit-config/UNIT_ID	GET	Get configuration
/unit-config/UNIT_ID	PUT	Set configuration

3 Implementation

The detail of the implementation shows as follows.

A power measuring units sends current and voltage values to a data collecting unit every second. This means the data resolution is one second. It has a micro controller (dsPIC30F3013) built-in for signal processing. The size is W 90mm × D 45mm × H 25mm. The production cost is approximately 120 USD, including the cost of 4 current sensors.

A data collecting unit gathers the above data from up to 32 power measuring units, and pushes calculated power usage data to a data store server every 20 seconds. The shape looks like an Ethernet switch, as shown in Figure 1. We use an UTP cable to connect between a data collecting unit and a power measuring unit, and it is used for data transfer and power supply. Figure 3 shows inside of a data collection unit. It consists of an SH3 CPU board (T-SH7706LSR, TAC Inc.), a serial and parallel signal converter, and 33 RJ-45 ports (32 ports for connecting with power measuring units; 1 port for connecting the Internet). Using a Xilinx Spartan-3E FPGA, up to 32 signals from power measuring units are converted to a single serial signal. An embedded Linux system based on buildroot and a power monitoring agent called pmon.py, which is written in Python, are running

on the CPU board. Pmon.py reads data via a serial driver and calculates the active power and reactive power based on voltage, current values, and some adjustment parameters.

A data store server is implemented on the Google App Engine (GAE), which provides scalable and stable data store on the Google's data centers. GAE data store uses Bigtable [2], which is a key-value store database. The data store is replicated across multiple different data centers asynchronously, using Megastore [1]. We have implemented a data store server based on Slim3 [6] which is a full-stack MVC framework optimized for GAE.

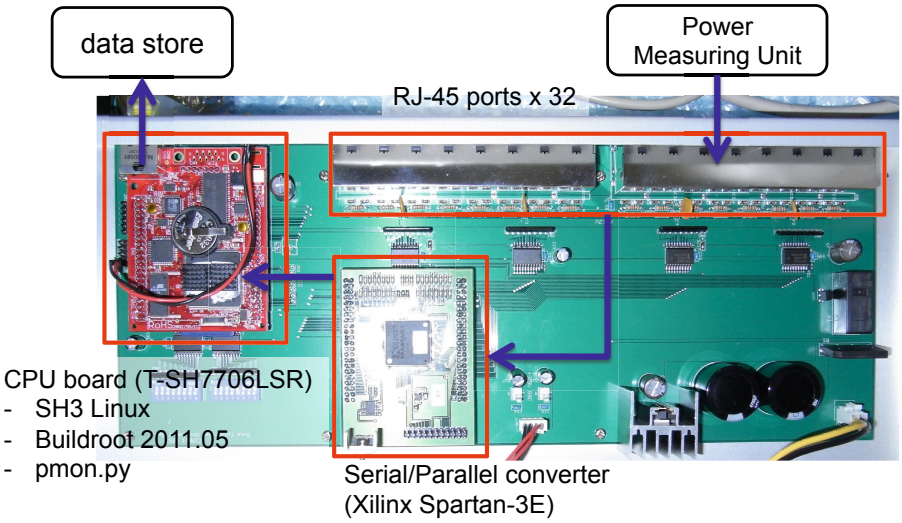


Fig. 3. Data collecting unit [size: W 400mm × D 250mm × H 60mm]

4 Demonstration

We have provided an electricity billing service for our server room (approximately 32m × 15m) users using the proposed system. The development of the system was completed within three months after the Great East Japan Earthquake. We have developed this system including hardware and software from scratch. We have gradually installed a system with 620 measurement points across two buildings at our campus, as shown in Figure 4. This system consists of 177 power measuring units and 10 data collecting units. Clamp-on current sensors are attached each power line of a power distribution board. Data collecting units are installed in free access floor. The billing service manages a mapping table between users and measurement points. It sends a billing report to each user.

Our developed system helps reduce total system cost of a power monitoring system by employing low-cost power measuring units (30 USD per measurement point), and utilizing cloud computing. GAE provides free quota per day. However, the monthly service charge is constantly about 20 USD because some usage of resources, including data store space and the number of read/write requests, are exceeded. In any case, the total cost, including service charge and maintenance fees, is cheaper than on-premise servers.

Using the proposed system, we also have shown that SSS MapReduce implementation [5] achieves lower power consumption compared with Apache Hadoop. We can compare power consumption at second granularity. In addition, we have successfully demonstrated visualization of power usage of network equipments gathering from sensors installed at the US and Japan, in the SC2011 conference.

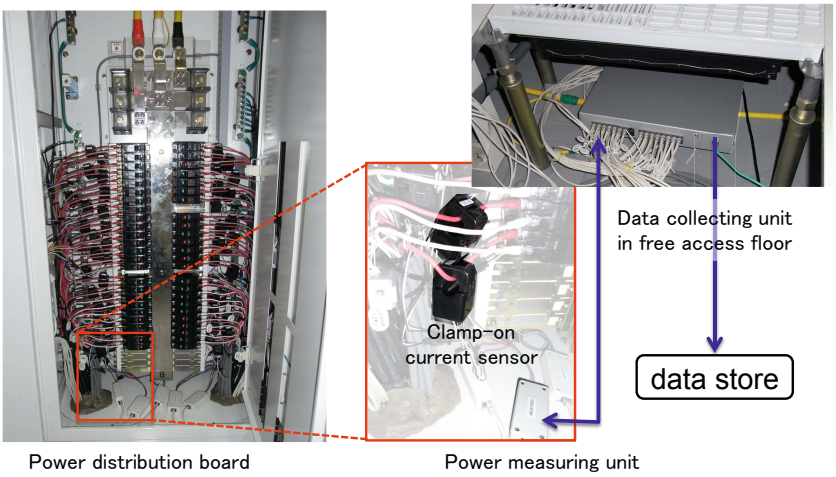


Fig. 4. An installation of the system to a server room at AIST

5 Conclusion

The proposed system helps reduce total cost of a power monitoring system and improve scalability by employing low-cost power measuring units and utilizing cloud computing. A system with 620 measurement points has been successfully installed at our campus, and we have operated it since the third quarter of 2011 to charge electricity bills and evaluate the power efficiency of data processing middleware. We believe the proposed system is feasible to use for other situations such as telecom and internet service providers and large factories.

We have been developing a reactive server consolidation mechanism [4]. We plan to demonstrate the effectiveness of reactive consolidation on real workloads by using the proposed monitoring system. The source code and the further information are available from the project page, `http://aist-power.hpcc.jp/`.

Acknowledgements. We would like to thank the members of the AIST electrical power monitoring system project, namely: Masahiro Murakawa and Yuji Kasai who have mainly developed the power measuring unit. This work was partly funded by the New Energy and Industrial Technology Development Organization (NEDO) Green-IT project.

References

1. Baker, J., Bond, C., Corbett, J., Furman, J., Khorlin, A., Larson, J., Leon, J.M., Li, Y., Lloyd, A., Yushprakh, V.: Megastore: Providing Scalable, Highly Available Storage for Interactive Services. In: Proceedings of the 5th Conference on Innovative Data Systems Research (CIDR), pp. 223–234 (January 2011)
2. Chang, F., Dean, J., Ghemawat, S., Hsieh, W.C., Wallach, D.A., Burrows, M., Chandra, T., Fikes, A., Gruber, R.E.: Bigtable: A Distributed Storage System for Structured Data. In: Proceedings of the 7th Symposium on Operating System Design and Implementation (OSDI), pp. 205–218 (November 2006)
3. Google App Engine, `http://code.google.com/appengine/`
4. Hirofuchi, T., Nakada, H., Itoh, S., Sekiguchi, S.: Reactive consolidation of virtual machines enabled by postcopy live migration. In: Proceedings of the 5th International Workshop on Virtualization Technologies in Distributed Computing (VTDC), pp. 11–18 (June 2011)
5. Ogawa, H., Nakada, H., Takano, R., Kudoh, T.: SSS: An Implementation of Key-value Store based MapReduce Framework. In: Proceedings of the 2nd International Conference on Cloud Computing Technology and Science, pp. 745–761 (2010)
6. Slim3, `https://sites.google.com/site/slim3appengine/`

A JSON Data Formats

We have defined data structures communicating among data collecting units, a data store server, and applications. We use JSON, a language independent data-interchange format. A JSON object is built on two structures: a set of key-value pairs, i.e., hash table, and an ordered list of values, i.e., array. The following shows JSON data structures that we have defined for the AIST power monitoring system. Figure 5, 6, and 7 show update, get, and configure message formats, respectively.

```
{
     "id":   "Data collecting unit ID"
     "time": "Data observed time (elapsed time in seconds since UNIX epoch)"
     "power":  {       # Data for the last 20 seconds per measurement point
        "sensor0.0": [VAL0, VAL1, VAL2, ..., VAL19],
        "sensor0.1": [VAL0, VAL1, VAL2, ..., VAL19],
        "sensor1.0": [VAL0, VAL1, VAL2, ..., VAL19],
        ....
     }
}
```

The each item of `power` shows a pair of a probe ID and an array of the power usage. The probe ID consists of the sensor name and the probe number. The array contains the power usage of the probe for the last 20 seconds since `time`.

Fig. 5. Data update message

```
{
    "time": "Data observed time (elapsed time in seconds since UNIX epoch)"
    "timeStr": "Data observed time (human readable format of YYYYmmDDTHHMM)"
    "power": {
        "LOCATION0": [VAL]
        "LOCATION1": [VAL]
        "LOCATION2": [VAL]
        ...
    }
}
```

The each item of `power` shows a pair of a location string `LOCATION` and the power usage.

Fig. 6. Data get message

```
{
    "id": "Data collecting unit ID"
    "description": "Installation location (free format)"
    "sensors": [
        {
            "id": "Sensor ID"
            "description": "Installation location (free format)"
            "combinations": [CT1, CT2, CT3, CT4]
            "probes": [
                {
                    "scale": VAL   # scale factor
                    "description": "measuring object (free format)"
                }, ...
            ]
            "CTs": [
                {
                    "factors": [Fpa, Fma, Foi, Foq],
                }, ...
            ]
        }, ...
    ]
}
```

The maximum number of `sensors` is 32; the maximum number of `probes` is 4. The `combinations` define probes, which are combinations of single or two current sensors connected with a power measuring unit. For instance, `"[0 1 2 3]"` denotes that there are 4 probes of single-phase 100V; `"[0 0 1 1]"` denotes that there are two probes of three-phase 200V. The `factors` are parameters for phase adjustment, multiplier adjustment, and zero-point adjustment.

Fig. 7. Configuration message format

Performance Evaluation of WDS-Based Mobile ITS Video Control System for Smart APT Traffic Control

Young-Hyuk Kim, Il-Kwon Lim, Jae-Gwang Lee, Jae-Pil Lee,
Hyun Namgung, and Jae-Kwang Lee*

Department of Computer Engineering, Hannam University,
Daejeon, Korea
{yhkim,iklim,leejk,jplee,ghnam,jklee}@netwk.hnu.kr

Abstract. This paper is to represent web-accessible and network-compatible Mobile ITS Video Control System for Smart APT Traffic Control for where Wi-Fi is avaliable. WLAN for IEEE 802.11a has adopted 5GHz frequency in light of the fact that consumer mobile devices work on 2GHz. For test purpose IEEE 802.11p, being a PHY/MAC standard for WAVE, a V2V/V21 Next-gen Network, replaced IEEE 802.11a. Plus, the test practically established WDS in an apartment complex in Daejeon City in confirmation of compatibility of Mobile ITS Video Control System while handling Wi-Fi handover issues.

Keywords: IEEE 802.11a, WDS, m-ITS.

1 Introduction

Modern-day people live in the cutting-edge society of network and vehicle technologies and in the thick of vehicles from time to time congesting traffic. Plus, the society have had a hard time accommodating increasing number of vehicles and handling such tangible and intangible expenditures as traffic jam, pollution, socio-economic factors and loss of economically active population, namely casualty. Conventional solution of road extension may not resolve fast-growing number of vehicle, which is why brand-new corrective actions are in dire need. Cutting-edge network technology may be the resort that modern-day society may take to, in improvement of the conventional traffic system and provision of brand-new traffic service to get rid of traffic issues. In light of these, what is called ITS(Intelligent Transportation System) is on the rise in countries across the world, as backed by government-level support [1-2].

With the concept of Ubiquitous-City(U-City) on the rise as well, pan-national R&D businesses have triggered researchers to be involved, especially for U-Space technology for U-Eco City including urban structure maintenance, maintenance improvement technology, intelligent urban administration and control technology improvement [3]. On the ceiling is furnished with temperature sensor for anti-fire

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_103,

service, with window done with BIPV (Building Integrated Photovoltaic System) for solar energy service. Car-embedded sensors make ITS possible, with traffic control service becoming feasible with serveilance cameras on for what is called Future-oriented Housing.

As far as this paper goes, Smart APT Mobile ITS Video Control System is the thing to concern, in the interests of traffic control and parking administration WLAN for IEEE 802.11a has adopted 5GHz frequency in light of the fact that consumer mobile devices work on 2GHz. For test purpose IEEE 802.11p, being a PHY/MAC standard for WAVE, a V2V/V21 Next-gen Network, replaced IEEE 802.11a. Plus, the test practically established WDS in an apartment complex in Daejeon City in confirmation of compatibility of Mobile ITS Video Control System while handling Wi-Fi handover issues. Chapter 2 of this paper handles ITS-relevant communication technologies and ITS researches, Chapter 3 does test environment and scenario, and Chapter 4 the test result for conclusion and research suggestion in Chapter 5.

2 Related Research

2.1 ITS-Relevant Communication Technology

IEEE 802.11a, WAVE, DSRC and WiBro, but not limited thereto, represents ITS-relevant communication Technology.

IEEE 802.11a

Confirmed in 1999 as a communication standard together with IEEE 802.11b, IEEE 802.11a is a means of wireless communication for high-speed data transfer using high-frequency of wider coverage and thus highly interruptible at where radio wave is incontinuous and overlapping. What retains distance and speed of transmission is OFDM, compatible with 5GHz-frequency ensuring high-speed data transfer at 6~54Mbps and distant communication up to 10~100m, though varied by surrounding environment. IEEE 802.11a is roaming-compatible for handover, broadly categorized as Soft Roaming and Hard Roaming.

Soft Roaming offers seamless connectivity by way of a switch on inter-AP environment, given APs are on line and cabled up, while sharing the same channel.

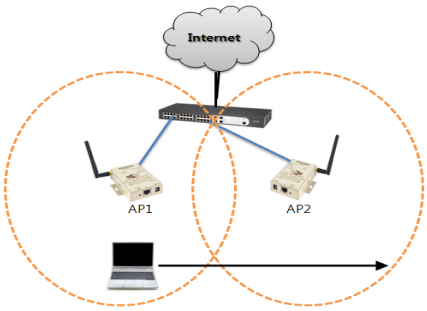


Fig. 1. Soft Roaming

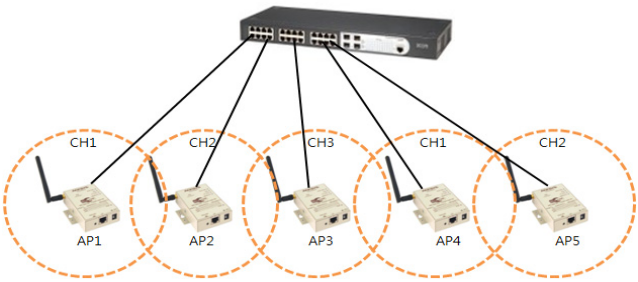


Fig. 2. Hard Roaming

Hard Roaming APs, however, are adopting divergent channels to get rid of interruption and thus features slower speed of roaming. Likewise, Hard Roaming APs should stay on line and cabled up, while on the divergent channels [4].

WAVE(Wireless Access in Vehicular Environment)

TG of IEEE 802.11WGis an American developer of Physical Layer and MAC Layer standards for 5.9GHz DSRC(Dedicated Short Range Communications) WAVE(Wireless Access in Vehicular Environment). WAVE is a means of wireless communication for ground, railroad and maritime transportation, transacting data between high-speed vehicle over 200km/h and roadside communication device(V2I) or arranging inter-vehicle data transaction(V2V), covering distance not longer than 1km. Adopting 5.9GHz-frequency, WAVE adopts adjusted IEEE 802.11a with supplementary protocol. On top of WAVE, scores of other standard-developing institutions are in the course of developing relevant standards. Plus, WAVE has a strategic feature that 802.11p accounts for PHY/MAC while upper-layer security, multi-channel support, system administration and networking service taken by Protocol 1609.1~4, for which the development is underway as supported by IEEE Intelligent Transportation Systems Council SCC32 Committee. On top of IEEE, also developing standards and testing and commercializing WAVE are IntelliDrive and OmniAir Consortium [5].

Table 1. Communication performance comparison

Values	WAVE	DSRC	WiBro	802.11a
Cell Coverage	~1km	~100m	~1km	~100m
Spectrum Band	5.9GHz	5.8GHz	2.4GHz	5GHz
Modulation	QPSK-OFDM	ASK	OFDMA	OFDM
Mobility	200Km/h	160km/h	60Km/h	-
Data Rates	100Mbps	1Mbps	DL 3, UL 1 Mbps	54Mbps
Channels	7	7	-	8

2.2 ITS Service Standardization Trend

VMC(Vehicle Multi-hop Communication)

VMC(Vehicle Multi-hop Communication) is for high-speed vehicle to be compatible with Telematics and ITS services. Electronics and Telecommunications Research Institute of Korea and few others have developed V2V and V2I, supportive technologies for VMC, dating back to Year 2007.

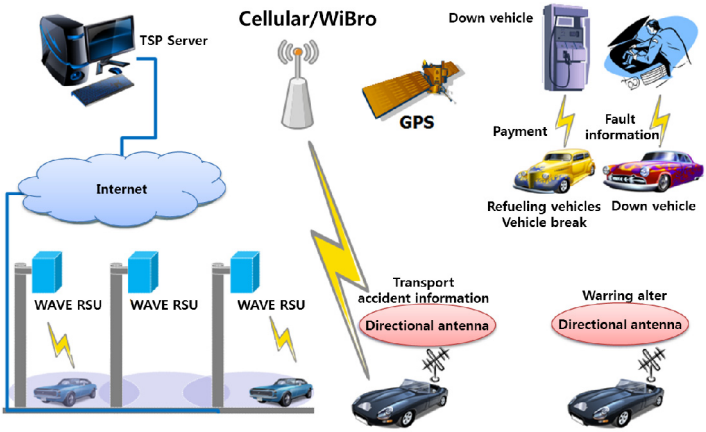


Fig. 3. VMC Tech Scheme

VMC development is underway to make V2V communication, broadcasting and multi-hop routing feasible and for eventual development of communication and safety control service such as accident prevention service, what is called 'herd driving' and group communication. Refer to Fig 3. for tech scheme of VMC. Based upon WAVE, high-tech bi-directional V2V and V2I communications of 10Mbps-caliber are in an attempt for seamless packet (data) transfer.

IntelliDrive

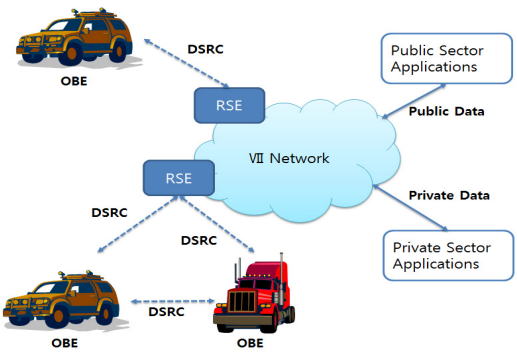


Fig. 4. IntelliDrive Project Scheme

Refer to Fig. 4 for IntelliDrive Project Scheme. Initiated as VII(Vehicle Infrastructure Integration) by US DOT(Department of Transportation), IntelliDrive Project got its name in the early 2009 and is for pan-national V2V and V2I communication system and infrastructure for safety and mobility of traffic as a means of brand-new, state-of-the-art service of traffic. Having backed by Federal DOT since Year 2003 and led by State DoTs and VII Consortia comprising Ford, Nissan Technical Center North America, BMW of North America, GM, Honda R&D Americas, Toyota Motor Engineering North America, Volkswagen of America, Mercedes-Benz Research and Development North America and Chrysler, IntelliDrive is to establish traffic controlling infrastructure to provide on-road drivers with real-time traffic information. Scores of researchers are involved in verification of system services, mobile application, software, hardware and infrastructure by way of Lab Test, Track Test and State-level Test Bed in DOTs for trial run, communication test and infrastructure test.

For IntelliDrive, State DOTs plans to develop V2V- and V2I-compatible WAVE for public and private benefits by way of nationwide roadside stations and other communication infrastructures. Here, WAVE refers to adjusted IEEE 802.11 Wireless Lan to be compatible with the vehicle specification for safety issues[6-7].

3 Test Bedding and Test

3.1 Test Site

Traffic controlling test for Smart APT Mobile ITS Video Control System has been test-bedded in courtesy of an apartment complex located in the city of Daejeon.

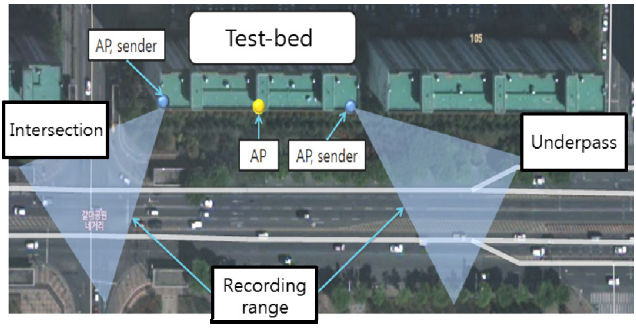


Fig. 5. Test Bedding Environment and Test Scenario

Test site represents congesting traffic and commands a fine view, from an apartment building rooftop, for two different junctions, as what Fig.4 portrays. View is from 40m above ground commanding ‘Underpass’ and ‘Public Park Intersection’. Locale 1, commanding Public Park Intersection, was furnished with AP1, Sender1

and Server, while Locale 2 serving 'Virtual Client' done with, likewise, AP2, Client1 and Client2. Lastly, Locale 3 commanding Underpass is furnished with AP3, Sender2 and Server2.

3.2 Composition of Test Equipment & Test Environment

Refer to Fig. 6 for the composition of test equipment and test environment. Wireless · Wired communication lines were, for test purpose, established by three different APs by way of WDS(Wireless Distribution System) with a rough coverage of 200M. Test environment was set to, for Locale 1, send out wireless video data and, for Locale 3, video data via wire communication. A total of two Clients at locale 2 were linked to switch via wire to communicate with Sender2 and Server2, while accessing, wireless, Sender1 and Server1 via WDS.

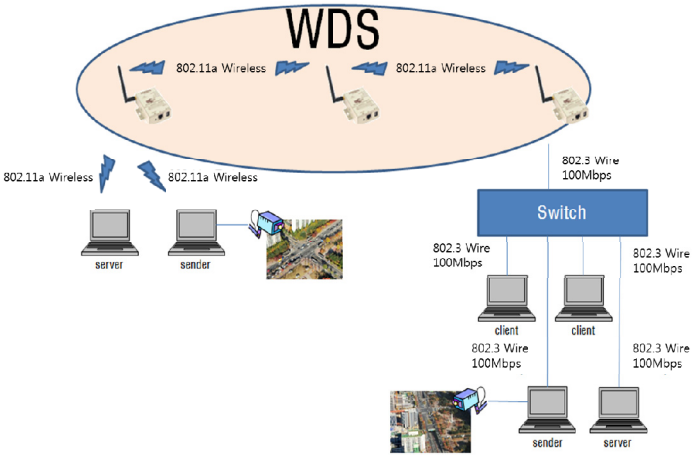


Fig. 6. Composition of Test Equipment & Test Environment

3.3 Test Item and Evaluation Criteria

Table 2. Test Item and Evaluation Criteria

#	Test Item	Unit	Content	Software
1	Data Transfer Rate	Mbps	Data Transfer Rate	Jperf
2	Latency	msec	Latency upon Data Transfer	Wireshark

4 Result

Table 3. Data Transfer Rate

Time Measure (sec)	Data Transfer Rate	Time Measure (sec)	Data Transfer Rate	Time Measure (sec)	Data Transfer Rate
0-1	4.98	20-21	4.06	40-41	6.36
1-2	4.13	21-22	3.60	41-42	7.08
2-3	4.26	22-23	4.19	42-43	4.33
3-4	3.74	23-24	3.60	43-44	4.26
4-5	6.09	24-25	5.77	44-45	4.39
5-6	6.23	25-26	5.64	45-46	6.49
6-7	5.83	26-27	5.51	46-47	6.88
7-8	3.67	27-28	5.57	47-48	5.31
8-9	4.00	28-29	5.57	48-49	6.42
9-10	7.21	29-30	5.77	49-50	6.68
10-11	4.13	30-31	5.05	50-51	6.68
11-12	3.47	31-32	6.95	51-52	5.37
12-13	3.93	32-33	3.80	52-53	6.68
13-14	3.87	33-34	5.96	53-54	4.39
14-15	5.64	34-35	5.44	54-55	4.52
15-16	6.16	35-36	5.57	55-56	6.88
16-17	4.85	36-37	5.18	56-57	6.42
17-18	3.54	37-38	5.57	57-58	5.57
18-19	3.87	38-39	6.55	58-59	6.88
19-20	7.73	39-40	4.98	59-60	7.14

Table 4. Latency

Trial	Sot	Packets Sent	Latency Rate
1	Server1	850	
1	Sender1	150	12%
1	Client (wire)	300	
1	Client (wireless	300	
2	Server1	850	
2	Sender1	150	12%
2	Client (wire)	300	
2	Client (wireless)	300	

5 Conclusion

This paper is to conclude communicability test for Smart APT Mobile ITS Video Control System, the integration of u-City and ITS. Test represents virtual WDS environment via high place APs in an apartment complex for wider coverage,

incorporating 5Ghz-frequency for surveillance cameras, communication devices and inter-server wireless communication for video-controllability.

Test turned out to conclude that Smart APT Mobile ITS Video Control System represent stable data transfer in WDS environment, more seamlessly by way of time-lapse images than video data involving some level of latency. Projected research is thus to confirm the rate of error, as well as designing and testing Smart APT Mobile ITS Video Control System by way of time-lapse imaging.

References

1. Lee, S.Y., Jeong, H.G., Yoon, S.H., Shin, D.K., Lim, K.T.: WAVE Specification and Trend of Next-Generation ITS Communication Technology. In: 2013 Summer Conference on The The Korean Institute of Communications and Information Sciences, pp. 756–757 (2013)
2. Kim, C.S., Jo, U.J.: A Study on the ITS integrated security system. In: 2012 Spring Conference on The Korea Institute of Electronic Communication Sciences, vol. 6(2), pp. 242–245 (2012)
3. Mun, C.Y.: U-Eco City Business Group. Korean Geo-Environmental Society 11(1), 67–72 (2010)
4. IEEE Standard 802.11a-1999: High-speed Physical Layer in the 5 GHz Band (1999)
5. Oh, H.S., Song, Y.S., Cho, H.B.: V2X communication technology and service trends. Journal of Korea Information Science Society, 19–24 (2013)
6. Webinar, I.: Safety Applications for Commercial Vehicles (2012)
7. http://www.its.dot.gov/

Simulation Based Opportunistic Network Coding in Ad Hoc Networks*

Hayoung Oh[1] and Sanghyun Ahn[2,**]

[1] School of Electronic Engineering,
Soongsil University, Korea
[2] School of Computer Science,
University of Seoul, Korea
hyoh@ssu.ac.kr, ahn@uos.ac.kr

Abstract. Network coding is a promising technology that increases the system throughput via reducing the number of transmissions for the packets delivered from the source node to the destination node in the saturated traffic scenario. Nevertheless, some packets can suffer from the metric of end-to-end delay. Since it takes the queuing delay at the intermediate node to wait for other packets to be encoded with (XOR). Therefore, in this paper, we analyze the delay according to the packet arrival rate and propose an enhanced network coding scheme, iXOR (Intelligent XOR). It reduces the average delay even under unsaturated traffic load through the holding-x strategy. Through an analysis and extensive simulations, we show that iXOR is better than the general forwarding scheme (FWD) without XOR and the XOR without the holding-x strategy, x=0, in the aspect of the average delay as well as the delivery ratio.

Keywords: forwarding, XOR, iXOR, DCF, delay analysis.

1 Introduction

Linear network coding and XOR are two major techniques in the network coding. Linear network coding transforms packets with the linear equation at every intermediate node and the destination node only needs to receive enough number of linear equations in the form of coded packets to successfully decode original packets [1][2]. It makes the destination avoid unnecessarily receiving the same packets several times by the retransmition mechanism after transmission failures. On the other hand, XOR reduces the number of transmissions because an intermediate node encodes the packets from both sides into one coded packet and broadcasts it to all the original destinations in the Alice-Bob or X-topology [3]. As a result, recently,

* This work was supported by Seoul Creative Human Development Program (HM120006) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP(Ministry of Science, ICT&Future Planning)) (No. 2010-0027410).

** Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_104,

network coding becomes a promising mechanism due to its effect on improving the network throughput and proof on the theoretical maximum network capacity. It was showed under the wired multicasting network firstly in [4], and the wireless network environment by several papers later in [5][6][7].

IEEE 802.11 recommended DCF (Distributed Coordination Function) as the mandatory function for MAC (Medium Access Control) of the wireless LAN (Local Area network) [8]. In the wireless communication, the transmission failures due to packet collisions can occur when stations transmit at the same time since they share the transmission medium – air. To reduce the collision probability, DCF uses CSMA/CA (Carrier Sense Multiple Access / Collision Avoidance) mechanism in which each station decreases its random back-off timer by one for the idle channel and transmit its packet when the back-off timer reaches zero. However, the network performance can still be decreased due to frequent packet collisions in the saturated traffic scenario. Therefore, most previous network coding schemes have only focused on the throughput improvement by reducing the number of transmitted packets and collisions in the saturated traffic scenario.

Our contribution is the average delay reduction as well as the delivery ratio improvement even in the unsaturated scenario. First, we analyze the delay according to the packet arrival rate and propose the Intelligent XOR (iXOR) using the holding-χ strategy to get more XOR chances even in the unsaturated scenario. Through the event-driven simulations with the NS-3 simulator, we can get the proper holding χ value and evaluate iXOR compared with the general forwarding scheme (FWD) without XOR and the XOR without the holding-χ strategy, $\chi = 0$. The rest of the paper is organized as follows. In Section 2, we show the related work. And we proposed the delay analysis under the dynamic traffic scenarios in DCF without XOR and with XOR in Section 3. Section 4 presents the proposed scheme and Section 5 evaluates the performance. Finally, Section 6 concludes the paper.

2 Related Work

2.1 DCF (Distributed Coordination Function)

IEEE 802.11 recommends DCF as the standard mechanism to reduce the collision probability of stations that share the transmission medium in wireless networks [8]. In wireless networks, packet collision is unavoidable because stations are unaware of the moment when their competitors transmit and several stations transmit simultaneously within the propagation delay. Thus, IEEE 802.11 adopts the CSMA/CA mechanism, in which a station decreases its random back-off timer by one at each slot time of the idle channel and transmits when its back-off timer reaches zero to reduce the packet collision probability.

DCF describes two techniques to employ for packet transmission [22]. The default scheme is the two-way handshaking technique called the basic access mechanism. This mechanism is characterized by the immediate transmission of a positive

acknowledgement (ACK) by the destination station, upon successful reception of the packet transmitted by the sender station. The explicit transmission of an ACK is required since, in the wireless medium, a transmitter can determine if a packet is successfully received by listening for an ACK from the destination station.

In addition to the basic access, an optional four way handshaking technique, known as the request-to-send/clear-to-send (RTS/CTS) mechanism has been standardized. Before transmitting a packet, a station operating in the RTS/CTS mode "reserves" the channel by sending a special RTS short frame. The destination station acknowledges the receipt of the RTS frame by sending back a CTS frame, after which normal packet transmission and ACK response occurs. Since only the RTS frame can suffer from a collision which can be detected by the lack of CTS response, the RTS/CTS mechanism can improve the system performance by reducing the collision duration experience by a message. As an important positive side effect, the RTS/CTS scheme designed in the IEEE 802.11 protocol is suited to combat the so-called problem of Hidden Terminals [23] which occurs when a pair of mobile stations can not hear each other due to some obstacles or fading. This problem has been specifically considered in [24] and in [25] in which the phenomenon of packet capture was also considered.

2.2 Network Coding

Network coding makes the theoretical maximum network capacity achievable practically by reducing the number of packet transmissions. There are two specific mechanisms in network coding, XOR [3] and the linear (random) network coding [1][2]. XOR reduces the number of transmissions in the way that the intermediate node only broadcasts the packet once encoded from the packets sent by several transmitters rather than simply forwards those packets one by one. Especially in [3] [8][9], the authors implemented the XOR-bit level network coding mechanism in a wireless network test-bed and showed that it improves network throughput by 38%. Recently, it was showed that the analog network coding [5] utilizing the signal interference rather than excluding it even reduces transmission time than the traditional bit-level network coding. All these network coding mechanisms can be widely utilized in P2P [16], efficient content distribution [11][12][13], energy efficiency [14][15], opportunistic routing [7], reliability gain [19][10], multi-hop network gain [18][20], relay network gain [17][21][14] and etc.

In contrast, linear coding [1][2] is that intermediate nodes forward every packet linear-transformed by a certain linear equation and destination nodes would decode original packets as long as they receive enough number of linear equations in the form of the number of packets. Linear transformation is a multiplication of a vector so-called coefficient to the bit-pattern of the packet that passes through a station, and it is called linear coding when the coefficient becomes 1 whereas it is called random linear coding when the coefficient is less than 1 and larger than 0. (Random) linear coding can be also applied to packet error recovery, multicast scenario and efficient delivery of urgent messages in VANETs and DTN.

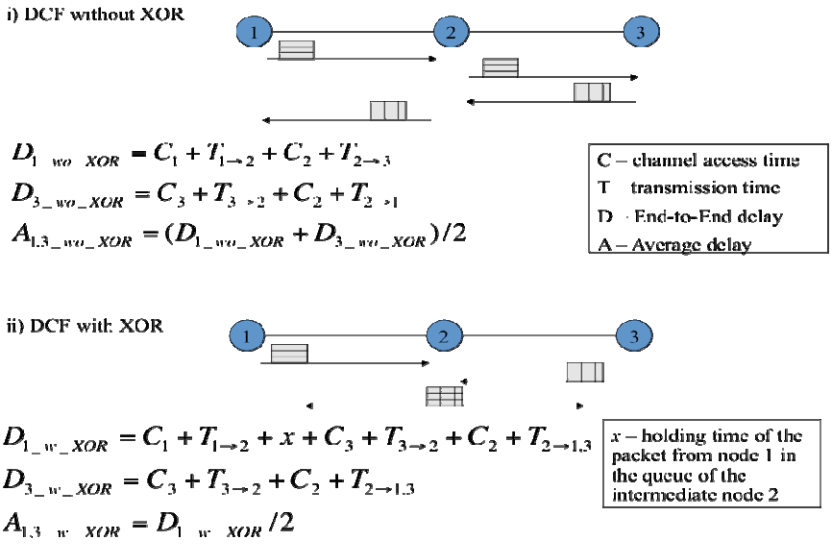


Fig. 1. E2E delay comparison, i) DCF without XOR and ii) DCF with XOR

3 Delay Analysis under the Dynamic Traffic in DCF without XOR and with XOR

Fig. 1 shows the delay comparison between DCF without XOR(the general forwarding scheme (FWD)) and DCF with XOR. First of all, i) the DCF without XOR case shows the end-to-end delay taken by each packet generated from the source nodes 1 and 3 to reach each destination nodes 3 and 1, $D_{1\_wo\_XOR}$ and $D_{3\_wo\_XOR}$, respectively. $D_{1\_wo\_XOR}$ is composed of the time C_1 for the source node 1 to compete for and grabs the channel, the time $T_{1\to 2}$ for the packet from the source node 1 to be transferred to the intermediate node 2, the time C_2 for the intermediate node 2 to compete for and grab the channel, and the time $T_{2\to 3}$ for the packet from the intermediate node 2 to be transferred to the destination node 3. Similarly, $D_{3\_wo\_XOR}$ can be derived like $D_{1\_wo\_XOR}$. And, $A_{1,3\_wo\_XOR}$ means the average delay of the packets from the node 1 and 3.

In contrast, ii) DCF with XOR shows the end-to-end delay of the packets from the source nodes 1 and 3 to the destination nodes 3 and 1 by a broadcast transmission of the coded packet at the intermediate node 2, $D_{1\_w\_XOR}$, $D_{3\_w\_XOR}$, respectively. $D_{1\_w\_XOR}$ is composed of the time C_1 when the source node 1 competes for and grabs the channel, the time $T_{1\to 2}$ when the packet from the source node 1 is transferred to the intermediate node 2, the time χ when the packet from the source node 1 queued in the intermediate node 2 to wait for the packet generated by another source node 3 to be encoded with, the time C_3 when the source node 3 competes for and grabs the channel, the time $T_{3\to 2}$ when the packet from the source node 3 is transferred to the intermediate node 2, the time C_2 when the intermediate node 2 competes for and

grabs the channel, the time $T_{2\rightarrow 1,3}$ when the intermediate node 2 broadcasts and successfully delivers the coded packet to the destination nodes 3 and 1. $D_{3\_w\_XOR}$ can be also interpreted similar to $D_{1\_w\_XOR}$. However, the packet from the node 3 does not wait at the intermediate node 2 because the packet from the node 1 for being coded with is already in the queue of the intermediate node 2. And in fact, $D_{3\_w\_XOR}$ is not needed because it can be a part of $D_{1\_w\_XOR}$. As a result, the average delay $A_{1,3\_w\_XOR}$ is generally shorter than $A_{1,3\_wo\_XOR}$ in the saturated scenario because XOR reduces the number of transmissions. However, $A_{1,3\_w\_XOR}$ can be rather longer than $A_{1,3\_wo\_XOR}$ due to the holding time χ according to the packet arrival rate, specially in the unsaturated scenario.

4 iXOR (Intelligent XOR) Using Holding-χ Strategy in an Ad Hoc Network

To design iXOR (Intelligent XOR) using the holding-χ strategy in an ad hoc network, we define several terms like the followings.

- W_A – the waiting time of an Alice's packet in the queue of the intermediate node for Bob's packet to arrive to be encoded with
- W_{A_i} , W_{A_ii} – two cases of the waiting time of an Alice's packet in the queue of the intermediate node according to the arrival of a Bob's packet.

Fig. 2 shows the meaning of W_{A_i} and W_{A_ii}. Namely, i) **Fig. 2(a)** shows the case that Bob has a packet to transmit at the point of time, t, while an Alice's packet is transmitted to the relay node. As a result, the waiting time of the Alice's packet in the queue of the intermediate node to be encoded with an Bob's packet, W_{A_i} , is the Alice's packet transmission time, T. Otherwise, ii) **Fig. 2(b)** shows the case that Bob has a packet to transmit at the point of time, t, after an Alice's packet is transmitted to the relay node. As a result, the waiting time of the Alice's packet in the queue of the intermediate node to be encoded with the Bob's packet, W_{A_ii}, is the time until the Bob's packet arrival, t+T-T.

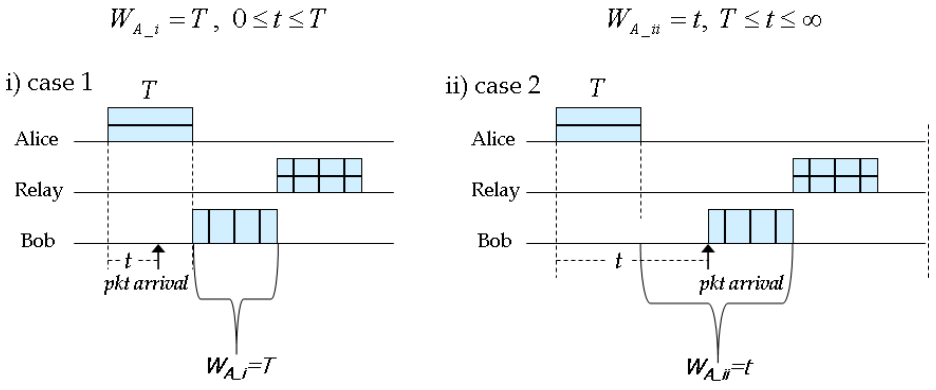


Fig. 2. Two cases of waiting time of an Alice's packet at the intermediate relay node, W_{A_i} and W_{A_ii}

As a result, if we assume that the packet arrival rate λ is exponentially distributed, we can calculate the expected waiting time of the Alice's packet $E[W_A]$ with considering $E[W_{A\_i}]$ and $E[W_{A\_ii}]$ in Eq.(1) and (2).

- $E[W_{A\_i}]$, $E[W_{A\_ii}]$ – the expected waiting time of an Alice's packet in each case

$$E[W_{A\_i}] = q_{A\_i} \cdot W_{A\_i} = \int_0^T \lambda e^{-\lambda t} \cdot T dt = \left(-e^{-\lambda t} \cdot T\right)_0^T = \left(e^{-\lambda t} \cdot T\right)_T^0 = T - e^{-\lambda T} \cdot T$$
$$E[W_{A\_ii}] = q_{A\_ii} \cdot W_{A\_ii} = \int_T^\infty \lambda e^{-\lambda t} \cdot t dt = \left(-\frac{e^{-\lambda t} \cdot (\lambda t + 1)}{\lambda}\right)\Bigg|_T^\infty = \frac{e^{-\lambda T} \cdot (\lambda T + 1)}{\lambda} \tag{1}$$

- $E[W_A]$ – the expected waiting time of an Alice's packet

$$E[W_A] = E[W_{A\_i}] + E[W_{A\_ii}] = \int_0^T \lambda e^{-\lambda t} \cdot T dt + \int_T^\infty \lambda e^{-\lambda t} \cdot t dt = T + \frac{e^{-\lambda T}}{\lambda} \tag{2}$$

q_{A_i} and q_{A_ii} are the probability of case i) and case ii), respectively.

$$q_{A\_i} = \int_0^T \lambda e^{-\lambda t} dt = 1 - e^{-\lambda T}$$
$$q_{A\_ii} = \int_T^\infty \lambda e^{-\lambda t} dt = e^{-\lambda T} \tag{3}$$

T is the channel occupation time (DIFS+backoff+Tx+SIFS+Ack+round-trip propagation delay) and t is the arrival time of an Bob's packet since the arrival of an Alice's packet.

Along the same line, we can also define the waiting time of an Bob's packet in the queue of the intermediate node as follows. Fig. 3 shows the meaning of W_{B_i} and W_{B_ii}.

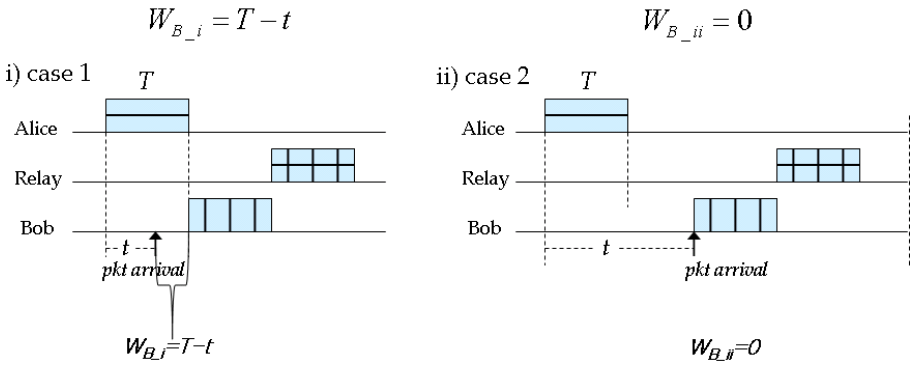


Fig. 3. Two cases of wating time of a Bob's packet in the intermediate node, W_{B_i} and W_{B_ii}

- W_B – the waiting time of Bob's packet in the queue of the intermediate node for an Alice's packet to be delivered to be encoded with
- W_{B_i} , W_{B_ii} – two cases of waiting time of a Bob's packet in the queue of the intermediate node according to the Alice's packet arrival.

As a result, we can also calculate the expected waiting time of a Bob's packet $E[W_B]$ considering $E[W_{B\_i}]$ and $E[W_{B\_ii}]$ in Eq.(4).

- $E[W_{B\_i}]$, $E[W_{B\_ii}]$ – the expected waiting time of a Bob's packet in each case
- $E[W_B]$ – the expectated waiting time of a Bob's packet

$$E[W_B] = E[W_{B\_i}] + E[W_{B\_ii}] = q_{B\_i} \cdot W_{B\_i} + q_{B\_ii} \cdot 0 = E[W_{B\_i}]$$
$$= \int_0^T \lambda e^{-\lambda t}(T - t)dt = \frac{e^{-\lambda t}(\lambda(t-T)+1)}{\lambda}\Bigg|_0^T = T + \frac{e^{-\lambda T} - 1}{\lambda} \quad (4)$$

Finally, we can get the total waiting time of DCF with XOR in Eq.(5).

$$E[W_{XOR}] = E[W_A] + E[W_B] = 2T + \frac{2e^{-\lambda T} - 1}{\lambda} \quad (5)$$

Fig. 4 Shows the comparisons of $E[W_A]$ and $E[W_B]$ for each packet according to the packet arrival rate λ and the packet size T. And Fig. 5 finally shows $E[W_{XOR}]$ of

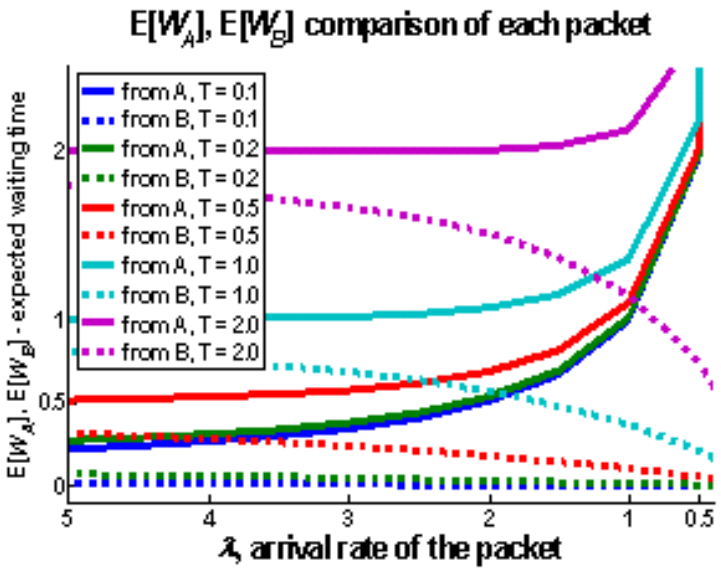


Fig. 4. E[WA] and $E[W_B]$ comparison per packet

the XOR exchange system. It shows that the network coding introduces rather longer delay when the packet arrival rate λ is very low. Therefore, we need to find the proper holding-χ at the intermediate node to reduce the average delay.

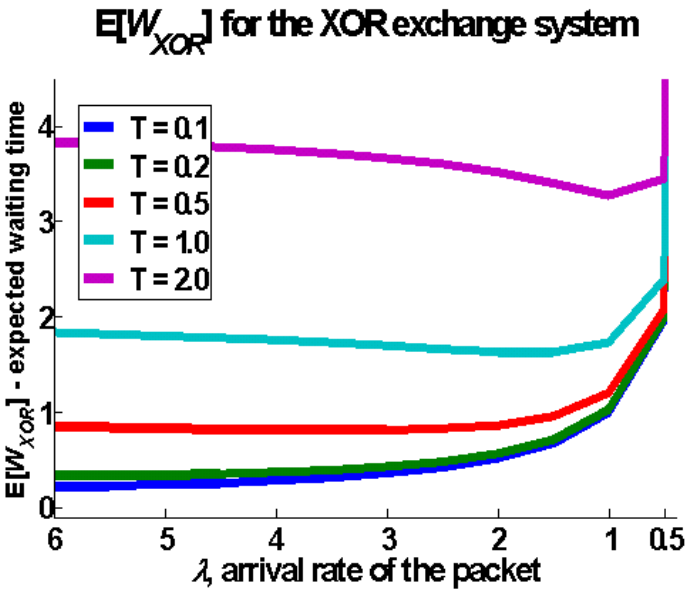


Fig. 5. E[WXOR] for the XOR exchange system

To compare DCF without XOR and with XOR, we explain in detail the case studies of DCF as shown in Fig. 6**.**

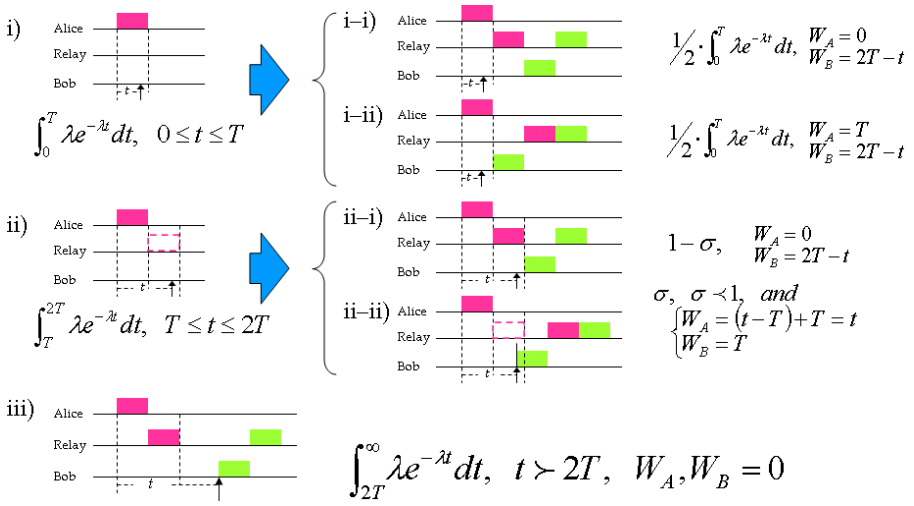


Fig. 6. The case studies of DCF

And we can calculate the total waiting time based on the general DCF in Eq.(6).

$$
\begin{aligned}
E[W_{DCF}] &= E[W_A] + E[W_B] \\
&\approx \left(\int_0^T \lambda e^{-\lambda t} \, 1/2 \, T dt + \int_T^{2T} \lambda e^{-\lambda t} \sigma t dt \right) + \\
&\left(\int_0^T \lambda e^{-\lambda t} (2T - t) dt + \int_T^{2T} \lambda e^{-\lambda t} (1 - \sigma)(2T - t) dt + \int_T^{2T} \lambda e^{-\lambda t} \sigma T dt \right) \\
&\approx \int_0^T \lambda e^{-\lambda t} \, 1/2 \, T dt + \int_0^{2T} \lambda e^{-\lambda t} (2T - t) dt \qquad (\sigma \prec 1) \\
&= -1/2 \cdot e^{-\lambda t} T \Big|_0^T + \frac{e^{-\lambda t}(1 + \lambda(t - 2T))}{\lambda} \Big|_0^{2T} \\
&= T/2 \left(5 - e^{-\lambda T}\right) + 1/\lambda \left(e^{-2\lambda T} - 1\right)
\end{aligned}
\tag{6}
$$

At last, we can get the cross point of the light blue line which gives the motivation of the proposed scheme, iXOR, as shown in Fig. 7. iXOR opportunistically changes the operation from XOR to DCF or reversely according to the packet arrival rate λ. Namely, in iXOR, the relay node holds an Alice's packet until it gets a Bob's packet as long as the arrival rate of the packet is larger than the cross point of the light blue line. On the other hand, the relay node does not hold an Alice's packet and forwards it directly to the destination based on DCF when the arrival rate of the packet is smaller than the cross point of the light blue line to reduce the average delay.

Motivated by the previous part, we focus on the holding time χ to reduce the average delay with iXOR in various scenarios according to the packet arrival rate λ. Fig. 8 shows iXOR (Intelligent XOR) using the holding-χ strategy in ad hoc networks**.** The proposed scheme considers three scenarios according to the packet arrival rate λ. In the first scenario with the low enough arrival rate λ, after a packet from the node 1 has arrived at the intermediate node 2, the intermediate node 2 cannot XOR-encode the packet with another packet from the node 3, because a packet from the node 3 does not arrive at the intermediate node 2 within the holding time χ. Namely, the time point t of the packet arrival from the node 3 is longer than the holding- χ *($t>\chi$ and $\chi \neq 0$)*. In the second scenario with the medium arrival rate, the intermediate node 2 can XOR-encode the two packets from the nodes 1 and 3, because the packet from the node 3 arrives at the intermediate node 2 within the holding-χ *($t<\chi$ and $\chi \neq 0$)*. And, in the third scenario with the high enough arrival rate, the intermediate node 2 can instantly XOR-encode the packets from the nodes 1 and 3 regardless of the holing-χ because a packet from the node 3 immediately arrives at the intermediated node 2 before the holding-χ *($t \leq \chi$ and $x=0$)*. Namely, in the proposed scheme, the intermediate node opportunistically uses the holding-χ strategy according to the packet arrival rate to get more XOR chances. We can figure out the proper χ through simulations.

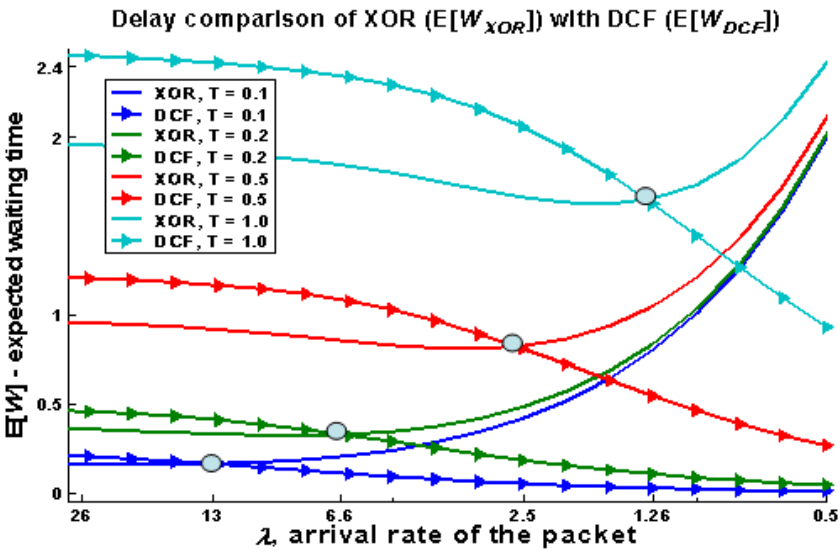


Fig. 7. Delay comparison between DCF with XOR (**E[W$_{XOR}$]**) and without XOR (**E[W$_{DCF}$]**)

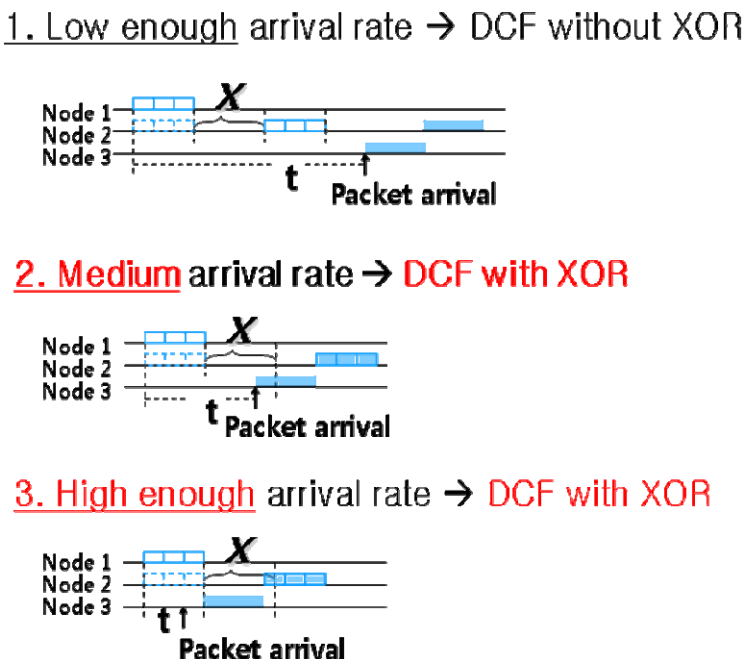


Fig. 8. iXOR (Intelligent XOR) using Holding-*x* Strategy

5 Performance Evaluation

To get the proper holding deadline χ and evaluate the average delay and the delivery ratio of FWD, XOR and iXOR, the event-driven simulations with the NS-3 simulator were performed. The parameters used in the simulations are summarized in Table 1.

Table 1. Simulation Parameters

Meaning	Value
Number of nodes	100
Packet arrival rate	Poisson arrival
Transmission time	12msec
Mean inter-arrival time	48msec
Packet size	1500byte
Data rate	1Mbps
Holding-χ	0.3msec

Fig. 9 shows the average delay of FWD, XOR and iXOR according to the holding time χ, "chi". Because FWD and XOR do not involve the "chi" parameter, their average delays are not affected bys "chi". And when "chi" is zero, iXOR works exactly the same as XOR. While "chi" is increased upto 0.3 msec, the average delay of iXOR outperforms XOR. However, if the "chi" value is larger than 0.3 msec, the holding-χ can be rather an overhead in the unsaturated scenario. Fig. 10 shows the

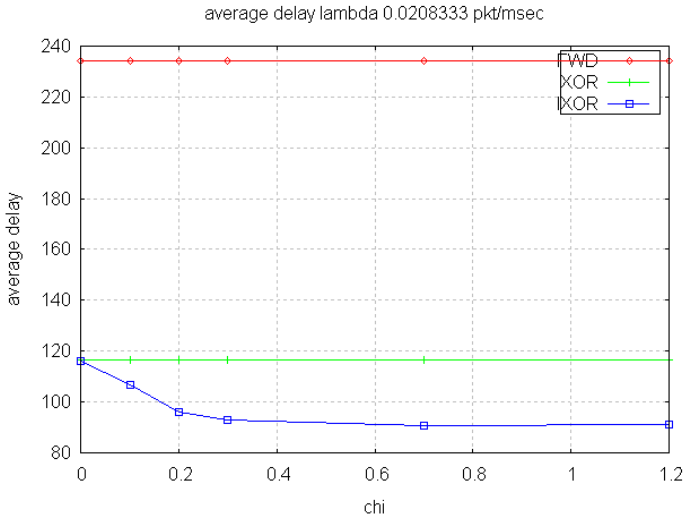


Fig. 9. Average delay vs. chi (=holding-χ)

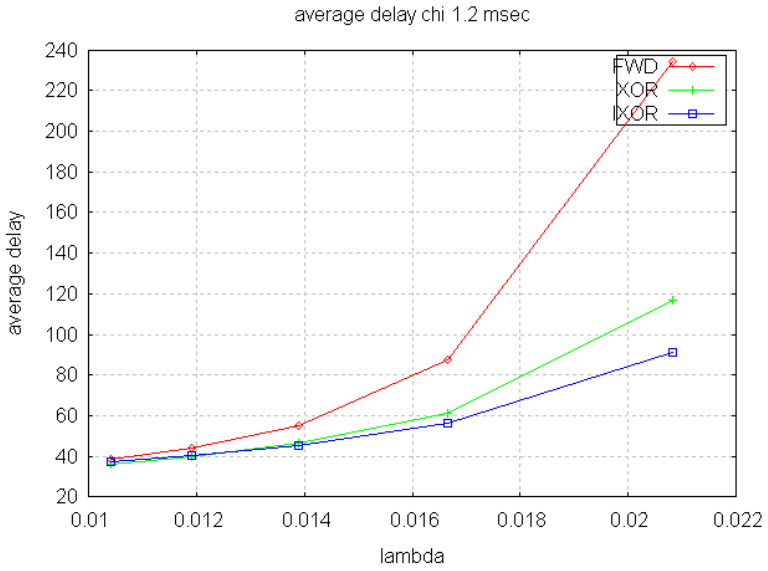


Fig. 10. Average delay vs. lambda (=packet arrival rate λ)

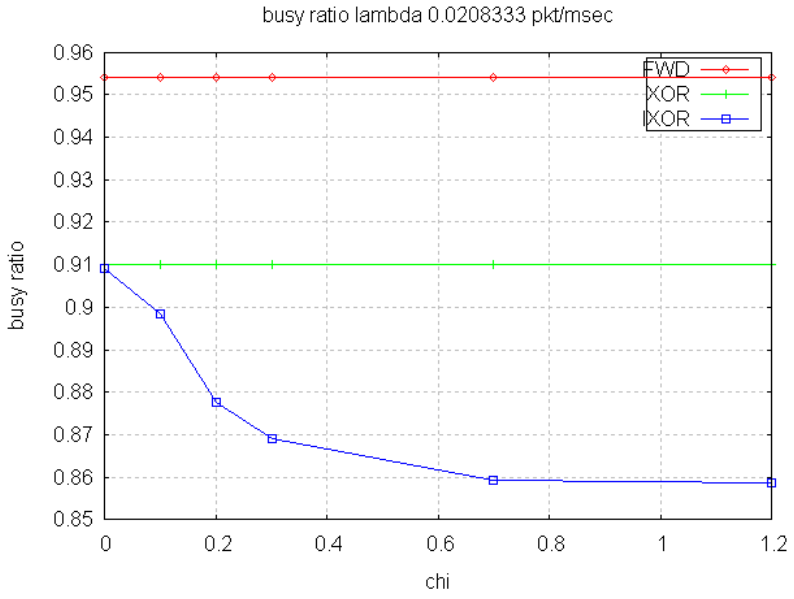


Fig. 11. Busy ratio vs. chi (=holding-χ)

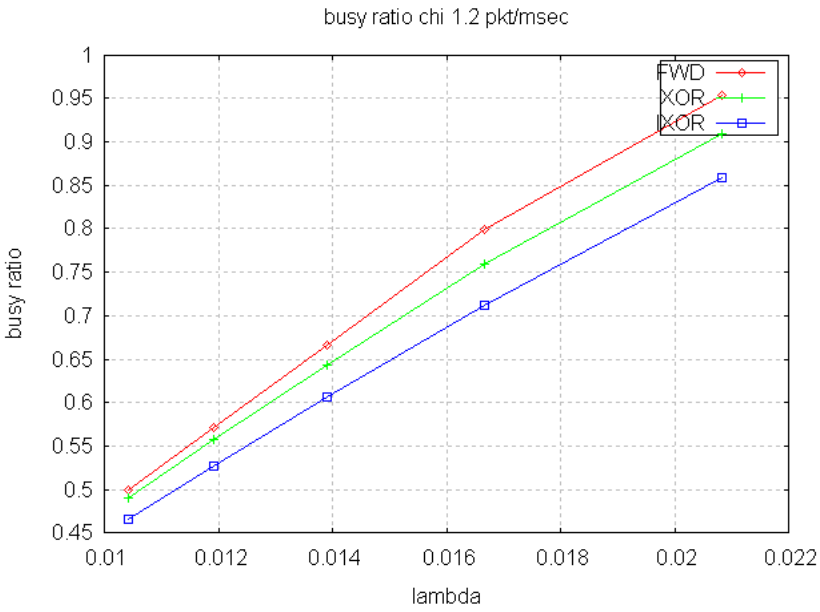


Fig. 12. Busy ratio vs. lambda (=packet arrival rate λ)

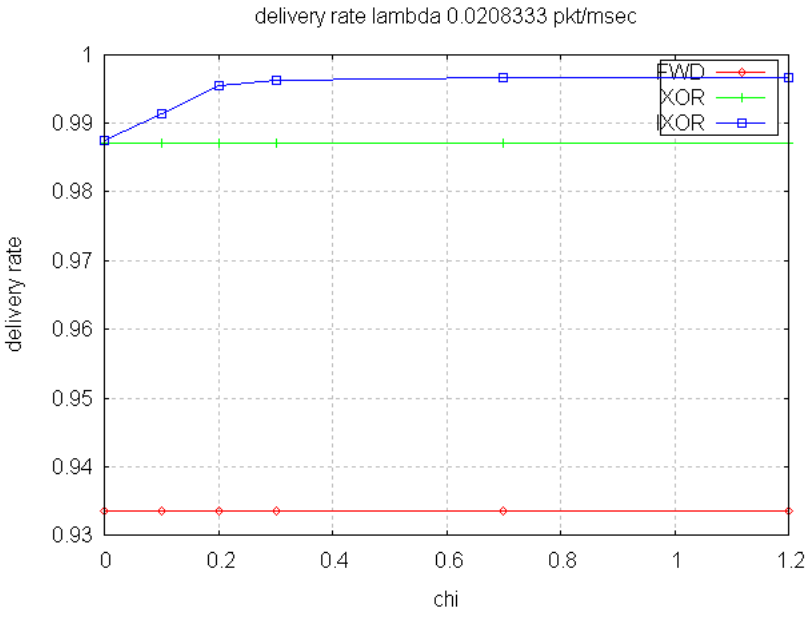


Fig. 13. Delivery ratio vs. chi (=holding-χ)

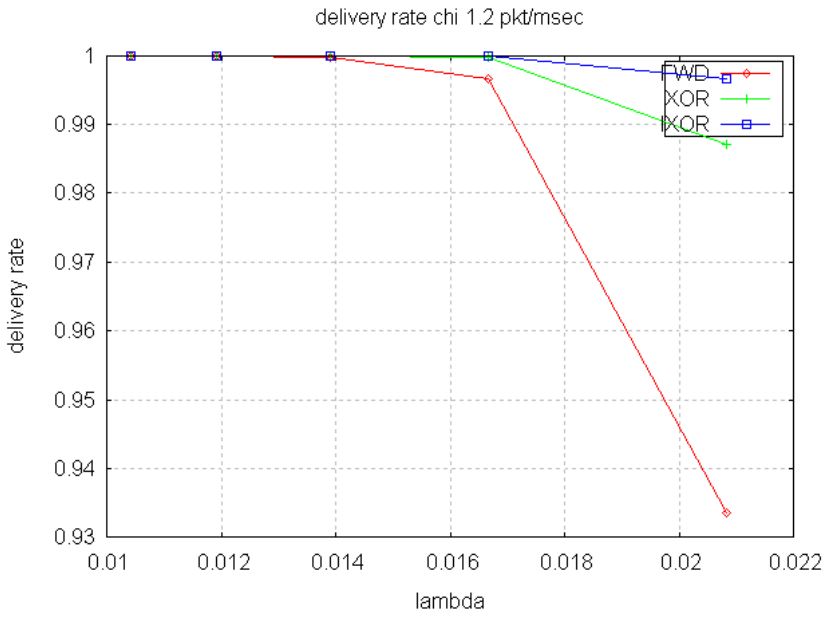


Fig. 14. Delivery ratio vs. lambda (=packet arrival rate λ)

average delay with varying the packet arrival rate λ, "lambda", and iXOR outperforms XOR for all "lambda"s. Fig.s 11 and 12 show the busy ratio of FWD, XOR and iXOR for various "chi"s and "lambda"s. The busy ratio means the wireless channel occupation ratio. iXOR outperforms FWD and XOR because it can reduce the number of transmissions through more coding chances. Similarly, the delivery ratio of iXOR is also better than FWD and XOR as shown in Fig.s 13 and 14.

6 Conclusion

In this paper, we analyzed the E2E delay for the packet arrival rate λ in a network coding capable wireless network and proposed the iXOR using the holding-χ strategy in an ad-hoc network. Through the simulations, we can figure out the proper holding-χ value. And we showed that iXOR outperforms FWD and XOR if the intermediate node opportunistically XOR-encodes packets with the holding-χ Strategy.

References

1. Li, S.-Y.R., Yeung, R.W., Cai, N.: Linear network coding. IEEE Trans. Inf. Theory 49, 371–381 (2003)
2. Fragouli, C., Widmer, J., Le Boudec, J.-Y.: Efficient Broadcasting Using Network Coding. IEEE/ACM Transactions on Networking 16(2) (April 2008)

3. Katti, S., Rahul, H., Hu, W., Katabi, D., et al.: XORs in the Air: Practical Wireless Network Coding. In: ACM SIGCOMM 2006, Pisa, Italy (September 2006)
4. Ahlswede, R., Cai, N., Li, S.-Y.R., Yeung, R.W.: Network Information Flow. IEEE Transactions on Information Theory (2000)
5. Katti, S., Gollakota, S., Katabi, D.: Embracing Wireless Interference: Analog Network Coding. In: ACM SIGCOMM (2007)
6. Katti, S., Katabi, D., Hu, W., Rahul, H., Medard, M.: Practical Network Coding for Wireless Environments. In: Allerton Conference on Communication, Control, and Computing (September 2005)
7. Chachulski, S., Jennings, M., Katti, S., Katabi, D.: Trading Structure for Randomness in Wireless Opportunistic Routing. In: Proc. of ACM SIGCOMM (2007)
8. IEEE, International Standard for Information Technology-Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific Requirements- Part 11: Wireless LAN Medium
9. Chou, P.A., Wu, Y., Jain, K.: Practical network coding. In: Proc. Allerton Conf. Allerton, IL (2003)
10. Le, J., Lui, J.C.S., Chiu, D.-M.: On the Performance Bounds of Practical Wireless Network Coding. IEEE Transactions on Mobile Computing 9(8) (August 2010)
11. Gkantsidis, C., Rodriguez, P.: Network coding for large scale content distribution. In: IEEE INFOCOM, Miami, FL (March 2005)
12. Lee, U., Park, J., Yeh, J., Pau, G., Gerla, M.: CodeTorrent: Content Distribution using Network Coding in VANET. In: MobiShare 2006 (2006)
13. Hou, I.-H., Tsai, Y.-E., Abdelzaher, T., Gupta, I.: AdapCode: Adaptive Network Coding for Code Updates in Wireless Sensor Networks. In: IEEE Infocom 2008 (2008)
14. Cui, T., Chen, L., Ho, T.: Energy Efficient Opportunistic Network Coding for Wireless Networks. In: IEEE Infocom 2008 (2008)
15. Fragouli, C., Widmer, J., Boudec, J.-Y.L.: A network coding approach to energy efficient broadcasting: From theory to practice. In: IEEE INFOCOM, Barcelona, Spain (April 2006)
16. Zhang, X., Li, B.: On the Market Power of Network Coding in P2P Content Distribution Systems. In: IEEE Infocom 2009 (2009)
17. Sagduyu, Y.E., Ephremides, A.: On Joint MAC and Network Coding in Wireless Ad Hoc Networks. IEEE Transactions on Information Theory (2007)
18. Jin, J., Li, B., Kong, T.: Is Random Network Coding Helpful in WiMAX? In: IEEE Infocom 2008 (2008)
19. Ghaderi, M., Towsley, D., Kurose, J.: Reliability Gain of Network Coding in Lossy Wireless Networks. In: IEEE Infocom 2008 (2008)
20. Dimitrios Koutsonikolas, Y., Hu, C., Wang, C.-C.: An Empirical Study of Performance Benefits of Network Coding in Multihop Wireless Networks. In: IEEE Infocom 2009 (2009)
21. Zhang, J., Zhang, Q.: Cooperative Network Coding-Aware Routing for Multi-Rate Wireless Networks. In: IEEE Infocom 2009 (2009)
22. Bianchi, G.: Performance Analysis of The IEEE 802.11 Distributed Coordination Function. IEEE J. Selected Areas in Comm. 18(3), 535–547 (2000)
23. Cali, F., Conti, M., Gregori, E.: IEEE 802.11 Wireless LAN: Capacity Analysis and Protocol Enhancement. In: Proc. IEEE INFOCOM 1998 Conf., pp. 142–149 (1998)

24. Cali, F., Conti, M., Gregori, E.: IEEE 802.11 Protocol: Design and Performance Evaluation of an Adaptive Backoff Mechanism. IEEE J. Selected Areas in Comm. 18(9), 1774–1786 (2000)
25. Chatzimisios, P., Boucouvalas, A.C., Vitsas, V.: Performance Analysis of IEEE 802.11 DCF in Presence of Transmission Errors. In: Proc. IEEE Int'l Conf. Comm., pp. 3854–3858 (2004)

Multipath Routing Method for Supporting QoS and Improving Energy Efficiency in WMSNs

Si-Yeong Bae[1], Sung-Keun Lee[2,*], and Jin-Gwang Koh[1]

[1] Department of Computer Engineering, Sunchon National University, Korea
{bsy233,kjg}@sunchon.ac.kr
[2] Department of Multimedia Engineering, Sunchon National University, Korea
sklee@sunchon.ac.kr

Abstract. QoS support is essential for differentiated data processing according to types and characteristics of traffics occurred from various applications such as environment monitoring, disaster control and data collection as well as efficient use of energy in order to transmit the multimedia data in Wireless Multimedia Sensor Network (WMSNs). This paper proposes a multipath routing method which may extend the life time of the whole network by supporting QoS of WMSNs and improving the energy efficiency. The proposed multipath routing method sets the path considering the distance to sink node, remaining energy of node and link quality. In addition, setting the path according to characteristics of packet in order to support the differentiation of service for guaranteeing QoS is based on the path cost. The proposed method improve the energy efficiency while securing the reliability by enhancing the packet transmission rate, reducing the packet loss rate and delay time preventing the concentration of energy consumption through the multipath and subsequent reorganization of the path.

Keywords: QoS, Energy Efficiency, Multi-path routing, WMSN, Link Quality.

1 Introduction

Wireless Sensor Networks(WSNs) are widely applied to various fields including environment and ecology monitoring, energy control, logistics and inventory control, battle area control and medical monitoring field. The significant roles of WSNs in such application fields are to monitor the target area and transmit the data collected from each sensor node. In general, WSNs consists of many sensor nodes which organize the multi-hop wireless networks by themselves. Each sensor node communicates each other using the low power signals. Recently the researches on Wireless Multimedia Sensor Networks (WMNs) are actively carried out as the modules which are able to collect the multimedia data like CMOS image sensor or microphones have been developed and the cheap hardware have been widely supplied[1][2].

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*,
Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_105,

The node for WSNs is a low power and low priced system which operates with battery. Since the life time of network could be shortened due to fast exhaustion of energy if they use many packets in order to transmit the large scaled multimedia data in sensor nodes which have limited energy, the life time of network should be extended to its maximum by enhancing the efficiency of energy use.

The issue to be discussed in WMNs is to support the Quality of Service(QoS) as well as the efficient energy use and the extension of the whole networks. Especially, the QoS support is essential for differentiated data processing according to types and characteristics of traffics occurred from various applications such as environment monitoring, disaster control and data collection. However, there are several problems to be solved in order to provide a differentiated support to services which require the real time or other reliable transmission in WSNs. That is, the problems are that the wireless channel is not reliable and unpredictable, that since the nodes operate with limited energy, it is difficult to recharge them and that the network topology changes frequently. So, such QoS protocol proposed in existing wireless networks such as IEEE 802.11 could not be applied to WMSNs. The issues of QoS support and efficient energy use in wireless sensor networks have been, in general, handled as separate issues[3][4].

In this paper, the Multipath Routing Method("MPRM") which supports QoS and enhances the energy efficiency by extending the life time of the whole networks is proposed. The selection of multipath in the proposed multipath routing method shall be based on the path cost. The path cost designates the weight value as different value according to priority of packets referring the information for the distance to sink node, remained energy of nodes and link's quality. In addition, the proposed method determines the path according to characteristics of packets based on path cost in order to provide a differentiated service for QoS guarantee. Thus, the proposed method improve the energy efficiency while securing the reliability by enhancing the packet transmission rate and reducing the packet loss rate and the time delay preventing the concentration of energy consumption through the multipath and its subsequent reorganization of path.

This paper consists as follows: Section 2 discusses about related studies and section 3 explains about the proposed Multipath Routing Method. Section 4 verifies the proposed algorithm and analyzes the result and section 5 offers conclusions.

2 Related Work

There are routing algorithms using LQI when selecting path, such as on-demand MinLQI and table-driven MultiHopLQI. MinLQI utilizes minimum LQI value from all the links as a metric value for estimating paths. When establishing a path, it selects the path with maximum metric value. In contrast, MultiHopLQI computes a metric value by summing up all the LQI values of all links from sink node to itself. These routing techniques can improve reliability of data transmission by choosing high-quality link using LQI value. On the other hand, it tends to increase the number of links along the path if there are many links with high quality. Energy consumption also can increase. Furthermore, it does not consider power remainder when selecting

nodes. Hence it is unavoidable to include nodes along the path even though their power remains very low. It results to shorten the network lifetime[5].

3 Multipath Routing Method

3.1 Multipath Routing

This paper determines packet's characteristic depending on a type of application that the packet belongs to, as shown in figure 1 Packet's priority is determined in two ways: by the application originating the packet or by predefined field in the packet. For example, if the packet comes from real-time data service, the packet has high priority, or if the packet is marked as high priority in the field, the packet has high priority. According to the packet's priority, the packet gets transmitted faster or slower.

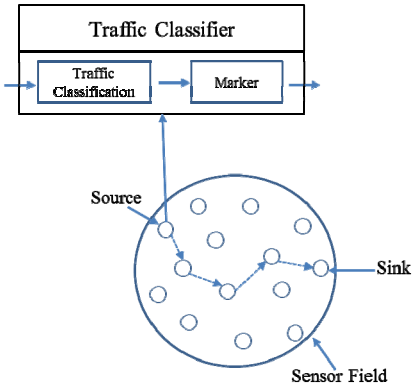


Fig. 1. Traffic control of sensor node

Packet's priority is marked at source node by setting service quality pattern and service quality level depending on traffic pattern and data contents. Packet marking algorithm is used for doing this[6]. Service quality pattern is classified as four categories; energy efficiency, delay sensitive type, reliability, and transmission rate. Incorporating with this, service quality level is classified as three levels; green, yellow, and red. Once the service quality level is decided, it is marked in priority field inside packet prior to being transmitted to sink node. In terms of type of levels, green indicates the most significant level, while red indicates the least significant level. Additionally, priority field is used for storing the packet's priority level. For differentiated service, green packet should be guaranteed to transmit first. Red packet is the least important packet, therefore its loss is acceptable because data can be recovered at sink node by combining existing packets. Importance level of yellow packet lies in between that of green and red.

3.2 Proposal of Multipath Routing Algorithm

Network condition changes very frequently in WSNs. For this reason, all the mobile nodes residing in network pass their routing information to other mobile nodes periodically. When the routing path is changed, its own routing information is also passed to other nodes. In order to maintain routing information for mobile devices, traditional routing mechanisms generate and send much amount of overhead traffic. It is main cause to degrade energy efficiency and shorten network lifetime. Routing table maintained by sensor node does not have to contain information about all the nodes in network. Only information about paths towards sink node is sufficient to operate the network properly. In addition to this, if information about adjacent neighbor sensor node is included, more various functionalities are possible to implement. The proposed routing mechanism adopts energy efficient routing table developing technique, where key means is broadcast messages sent by sink node periodically[7]. Table 1 illustrates the detailed fields of broadcasting message for routing.

Table 1. Fields of broadcasting message for routing

Field	Signification
ID	Broadcasting message Identifier
Flag	To classify broadcasting message into three : first routing construct message, routing table update message and normal broadcasting message
Node Identifier	Sender identifier
Location Information	Sender location information
Hops to Sink	Hop count from current node to sink node
Energy level	Residuary energy level of sender
Path Costs	Path Costs by packet priority

Table 2 shows the detailed fields of Routing Table proposed in this paper. The priority consists of 3 steps of Green, Yellow and Red. Next node means a node to receive the sensing data. Hop count means the number of hop from sink. The remaining energy means a remaining energy of node which is to receive the sensing data. The position information means the position of node which is to receive the sensing data. The path cost means the path costs required to sink node by packet priority.

Table 2. Fields of Routing Table

Node (position)	Priority	Neighbor Node	Hops	Remaining Energy	LQI	Path Costs	Path cost of neighbor node	Next Node

Assume that the networks are dense and there are several paths between two nodes. The path cost required for node *a* to select the next node *b* in order to transmit the packet whose priority is *pri* to sink node could be obtained by following equation 1.

$$C_a^{pri} = \min_{b \in N_a} \left\{ \alpha_{pri} \left(\frac{d_{by} + 1}{d_{ay} + 1} - 0.5 \right) \times \beta_{pri} \left(1 - \frac{LQI_b}{256} \right) \times \gamma_{pri} \left(1 - \frac{e_b^{res}}{e_b^{init}} \right) + C_b^{pri} \right\} \quad (1)$$

where *pri* denotes priority mark, $pri \in \{\text{Green, Yellow, Red}\}$, N_a is a set of neighbor nodes of node *a*. d_{ay} denotes hop distance from node *a* to sink node *y*. d_{by} denotes hop distance between node *b* and sink node *y*. e_b^{init} is an initial energy level of node *b*, and e_b^{res} is remaining energy of node *b*. LQI_b is link quality value for node *b*, ranging from 0 to 255. $\left(\frac{d_{by}+1}{d_{ay}+1} - 0.5 \right)$, $\left(1 - \frac{LQI_b}{256} \right)$ and $\left(1 - \frac{e_b^{res}}{e_b^{init}} \right)$ denote hop distance, link quality and cost depending energy remainder respectively, ranging from 0 to 1.

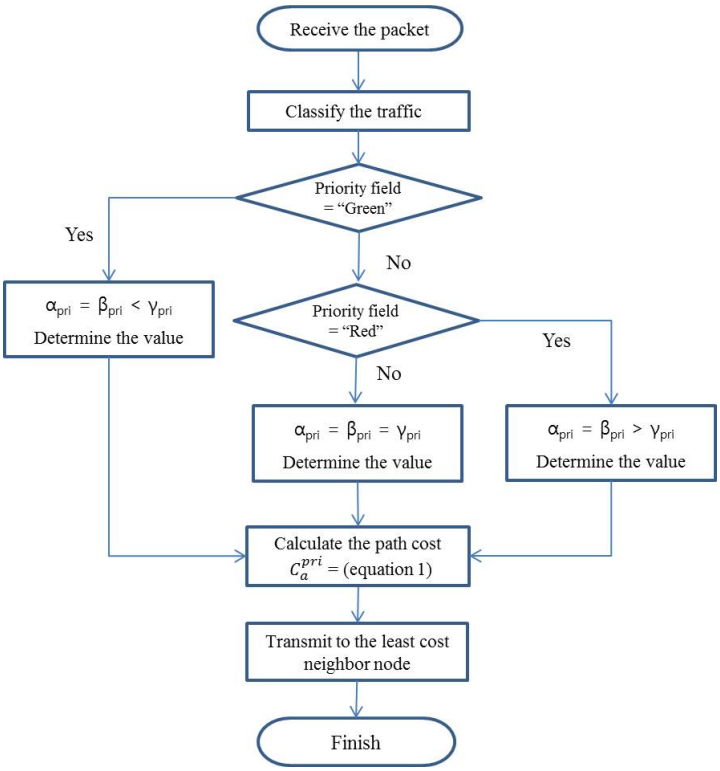


Fig. 2. Flowchart of Multipath Routing Algorithm

α_{pri}, β_{pri} and γ_{pri} are weighted value differently set depending on packet's priority. Since green packet has highest priority, hop counts and link quality should be considered rather than remaining energy when setting the weighted values. If the nodes have the same hop counts, link quality needs to be considered subsequently. Therefore, α_{Green} and β_{Green} is supposed to set to low value, whereas γ_{Green} is set to high value. Yellow packet has middle level of priority. Therefore, hop count, link quality and remaining energy should be considered in balanced manner. Finally, red packet has the lowest priority, hence remaining energy needs to be considered first rather than hop count and link quality. In order to minimize energy consumption, γ_{Red} is set to low weighted value, whereas α_{Red} and β_{Red} are set to be high. Fig. 2 shows the Flowchart of Multipath Routing Algorithm.

4 Verification and Result Analysis

Fig. 3 is the configuration of sensor networks for verification and result analysis. 25 nodes used for verification have a greed form of 5 x 5. Each node transmits the packet to the node adjacent in direction of left, right, up and down. The broadcasting message to update the Routing table is transmitted from the sink. In the event of two adjacent nodes, the Table shall be updated when the standby receiving timer of node is completed.

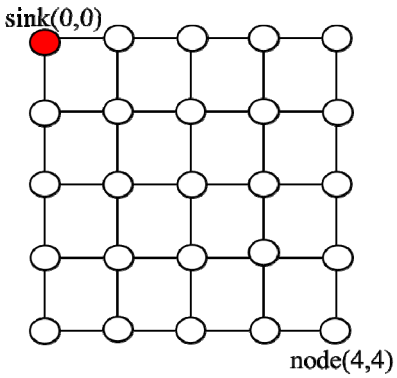


Fig. 3. Configuration of nodes

Fig.4 and 5 show the Routing Table and path selection of nodes (3.4) respectively. The priority was referred the Priority Field. Obtain the least path cost of each neighbor node using formula 1 in order to select the next node and select one neighbor node which has the least path cost by each packet priority.

```
============================================================
                    Path Node pos[3][4]
------------------------------------------------------------
   Hop to Sink : 7
   Next node(Path Cost)of  Green Packet : [2][4] (0.798)
                          Yellow Packet : [3][3] (1.114)
                             Red Packet : [4][4] (1.385)
------------------------------------------------------------
   Next Node(pos infor) : [2][4]
   Hops To Sink : 6
   Path Cost of Green Packet : 0.625
              Yellow Packet : 0.844
                 Red Packet : 1.044
   Remaining Energy : 0.60
   LQI: 108
------------------------------------------------------------
   Next Node(pos infor) : [3][3]
   Hops To Sink : 6
   Path Cost of Green Packet : 0.655
              Yellow Packet : 0.809
                 Red Packet : 1.042
   Remaining Energy : 0.65
   LQI: 80
------------------------------------------------------------
   Next Node(pos infor) : [4][4]
   Hops To Sink : 8
   Path Cost of Green Packet : 0.896
              Yellow Packet : 1.142
                 Red Packet : 1.247
   Remaining Energy : 0.92
   LQI: 80
============================================================
```

Fig. 4. Routing Table of Node (3,4)

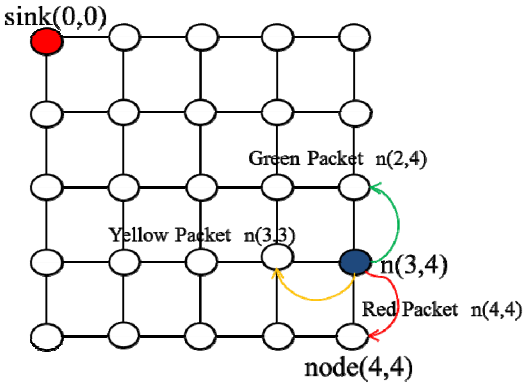


Fig. 5. Path Selection of Node (3,4)

5 Conclusions

The Multipath Routing Method(MPRM) for supporting QoS and improving the energy efficiency in WMSNs is proposed in this paper. The proposed MPRM may support a differentiated service to guarantee QoS by transmitting the high priority data fast considering the characteristics according to classification of applications to which the packets to transmit belong. In addition, the source node sets the multipath referring LQI value with which the number of hops to sink according to the marked Priority, information for remained energy of neighbor node and link quality could be

secured. Since the neighbor node is selected with the least path cost by each packet priority as a result of verification, the differentiated service to guarantee the QoS of packet and an energy efficient routing service could be supported. It is expected that the multipath setting shall be able to improve the energy efficiency of sensor networks by preventing the concentration of energy consumption and the subsequent reorganization of path.

Acknowledgements. 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-2013-H0401-13-2008) supervised by the NIPA(National IT Industry Promotion Agency).

References

1. Akyildiz, I.F., Weilian, S., Sankarasubramaniam, Y., Cayirci, E.: A Survey on Sensor Networks. Comm. IEEE 40(8), 102–114 (2002)
2. Akyildiz, I.F., Melodia, T., Chowdhury, K.R.: A survey on wireless multimedia sensor networks. Computer Net. 51(4), 921–960 (2007)
3. Akyildiz, I., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor networks: A survey. Computer Networks 38(4), 393–422 (2002)
4. Melodia, T., Akyildiz, I.F.: Cross-layer QoS-Aware Communication for Ultra Wide Band Wireless Multimedia Sensor Networks. IEEE J. of Sel. Areas in Comm. 28(5), 653–663 (2010)
5. Lee, W.-J., et al.: Minimum LQI based On-demand Routing Protocol for Sensor Networks. Journal of the Korea Academia-Industrial Cooperation Society 10(5), 3218–3226 (2009)
6. Jang, Y.-J., Kim, J.-H., Lee, S.-K., Koh, J.-G.: Traffic Control Mechanism for QoS provisioning in Wireless Multimedia Sensor Networks. In: ICONI & APIC-IST, pp. 599–603 (2010)
7. Jang, Y.-J., Bae, S.-Y., Lee, S.-K.: An Energy-Efficient Routing Algorithm in Wireless Sensor Networks. In: Kim, T.-H., Adeli, H., Slezak, D., Sandnes, F.E., Song, X., Chung, K.-i., Arnett, K.P. (eds.) FGIT 2011. LNCS, vol. 7105, pp. 183–189. Springer, Heidelberg (2011)

Author Index