
# Computer Science and Information Systems

Published by ComSIS Consortium

**Special Issue on Mobile Collaboration Technologies and Internet Services**

Volume 11, Number 3
August 2014

Computer Science and Information Systems

Vol 11, No 3, August 2014

ComSIS is an international journal published by the ComSIS Consortium



**ComSIS Editorial Office:**
**University of Novi Sad, Faculty of Sciences,**
**Department of Mathematics and Informatics**
Trg Dositeja Obradovića 4, 21000 Novi Sad, Serbia
**Phone:** +381 21 458 888; **Fax:** +381 21 6350 458
www.comsis.org; Email: comsis@uns.ac.rs

**Volume 11, Number 3, 2014**
**Novi Sad**

# Computer Science and Information Systems

Special Issue on Mobile Collaboration Technologies and Internet Services

**ISSN: 1820-0214**

The ComSIS journal is sponsored by:

Ministry of Education, Science and Technological Development of the Republic of Serbia
http://www.mpn.gov.rs/

# Computer Science and Information Systems

## AIMS AND SCOPE

Computer Science and Information Systems (ComSIS) is an international refereed journal, published in Serbia. The objective of ComSIS is to communicate important research and development results in the areas of computer science, software engineering, and information systems.

We publish original papers of lasting value covering both theoretical foundations of computer science and commercial, industrial, or educational aspects that provide new insights into design and implementation of software and information systems. ComSIS also welcomes survey papers that contribute to the understanding of emerging and important fields of computer science. In addition to wide-scope regular issues, ComSIS also includes special issues covering specific topics in all areas of computer science and information systems.

ComSIS publishes invited and regular papers in English. Papers that pass a strict reviewing procedure are accepted for publishing. ComSIS is published semiannually.

### Indexing Information

ComSIS is covered or selected for coverage in the following:

- Science Citation Index (also known as SciSearch) and Journal Citation Reports / Science Edition by Thomson Reuters, with 2013 two-year impact factor 0.575
- Computer Science Bibliography, University of Trier (DBLP),
- EMBASE (Elsevier),
- Scopus (Elsevier),
- Summon (Serials Solutions),
- EBSCO bibliographic databases,
- IET bibliographic database Inspec,
- FIZ Karlsruhe bibliographic database io-port,
- Index of Information Systems Journals (Deakin University, Australia),
- Directory of Open Access Journals (DOAJ),
- Google Scholar,
- Journal Bibliometric Report of the Center for Evaluation in Education and Science (CEON/CEES) in cooperation with the National Library of Serbia, for the Serbian Ministry of Education and Science,
- Serbian Citation Index (SCIndeks),
- doiSerbia.

### Information for Contributors

The Editors will be pleased to receive contributions from all parts of the world. An electronic version (MS Word or LaTeX), or three hard-copies of the manuscript written in English, intended for publication and prepared as described in "Manuscript Requirements" (which may be downloaded from http://www.comsis.org), along with a cover letter containing the corresponding author's details should be sent to official journal e-mail.

**Criteria for Acceptance**

Criteria for acceptance will be appropriateness to the field of Journal, as described in the Aims and Scope, taking into account the merit of the content and presentation. The number of pages of submitted articles is limited to 20 (using the appropriate Word or LaTeX template).

Manuscripts will be refereed in the manner customary with scientific journals before being accepted for publication.

**Copyright and Use Agreement**

# Computer Science and Information Systems

Volume 11, Number 3, Special Issue, August 2014


## CONTENTS

# EDITORIAL

Various collaboration technologies and Internet services have successfully and continually improved enterprise work efficiency, and have influenced and changed the quality of our lives over the past decade. Mobile collaboration technologies and Internet services, however, still lack robust functionality and content representation support. Meanwhile, the explosive growth of data traffic for user services threatens the current mobile systems. Especially, mobility models, architectures, and application services have posed various challenges to those in academia, industry, and governmental research labs. In particular, the key challenges for improving efficiency, scalability, and reliability are the development of mobile collaboration technologies and Internet services and the measurement of precise performance of mobile collaboration technologies and Internet services. These challenges allow us to design and develop new models, architectures, and services for future mobile systems. New research results in theory, simulation, and experimental approaches on mobile collaboration technologies and Internet services are welcome.

This special issue covers the following main topics:

- Security and data management
- Cloud architecture
- Hospital wireless sensor network
- Internet of things service interface
- Perspective security techniques for public Internet services
- Mobile devices and wireless communication technologies
- Protocols and algorithms for cooperative wireless relay networks
- Security and dependable applications
- Secure machine-to-machine communication
- Network architecture of big data
- Trust and privacy in wireless sensor networks
- Innovative cooperative communication technologies
- Social networking services
- Innovative intelligent algorithms and network security.

These subjects, as well as some others, are the focus of this special issue of "Mobile Collaboration Technologies and Internet Services". The special issue is organized as follows:

The first paper by Yi-Li Huang, Fang-Yie Leu, Jian-Hong Chen and William Cheng-Chung Chu, propose a true random-number encryption method employing block cipher and pseudo random number generator. They present a more secure one, called the True Random Number Encryption Method (TRNEM for short), which employs the current time, true random numbers and system security codes as the parameters of the encryption process to increase the security level of a system. The same plaintext file encrypted by the TRNEM at different time points generates

different ciphertext files. So the ciphertext files are difficult to be cracked. They also analyze the security of the Data Encryption Standard (DES), Advanced Encryption Standard (AES) and TRNEM, and explain why the TRNEM can effectively defend some specific attacks, and why it is safer than the DES and AES.

The second paper by Chin-Ling Chen, Woei-Jiunn Tsaur, Yu-Yi Chen and Yao-Chung Chang, presents a secure mobile digital rights management (DRM) system based on cloud architecture. They show that information security can be achieved efficiently via cloud server architecture and a cryptography mechanism. The proposed scheme focuses on using a mobile device to access the cloud service. The DRM mechanisms can protect digital content; once the mobile users pass the authentication they can access the cloud services, with authenticated users able to easily use mobile devices to read digital content.

The third paper by Mohammadreza Sahebi Shahamabadi, Borhanuddin M Ali, Nor Kamariah Bt. Noordin, Mohd Fadlee B. A. Rasid, Pooria Varahram and Antonio J. Jara, describes a Network Mobility - Hospital Wireless Sensor Network (NEMO-HWSN) solution to support IPv6 Low-power Personal Area Networks (6LoWPAN) network mobility in hospital wireless sensor network. They survey IPv6 mobility protocols and later propose a suitable solution for a hospital architecture based on 6LoWPAN technology. Moreover, they discuss an important metric like signaling overload to optimize the power consumption and how it can be optimized through the mobility management. This metric is more effective on the mobile router as a coordinator in NEMO since a mobile router normally constitutes a bottleneck in such a system. Finally, they present their initial results on a reduction of the mobility signaling cost and the tunneling traffic on the mobile personal area network.

The fourth paper by Hae-Min Moon and Sung Bum Pan, suggests long distance face recognition for enhanced performance of Internet of things service interface. While the existing face recognition algorithm uses single distance image as training images, the proposed algorithm uses face images at distance extracted from 1 to 5m as training images. In the proposed Linear Discriminant Analysis (LDA)-based long distance face recognition algorithm, the bilinear interpolation is used to normalize the size of the face image and a Euclidean distance measure is used for the similarity measure. As a result, the proposed face recognition algorithm is improved in its performance by 6.1% at short distance and 31.0% at long distance, so it is expected to be applicable for Ubiquitous Sensor Network (USN)'s robot and surveillance security systems.

The fifth paper by Tianhan Gao, Nan Guo, Kangbin Yim and Qianyi Wang, proposes a Privacy-Preserving Security (PPS) scheme for multi-operator wireless mesh networks with enhanced user experience. By hybrid utilization of the tri-lateral variable pseudonym approach and different kinds of tickets under identity-based proxy signature and proxy blind signature (PBS), anonymity, untraceability, as well as sophisticated unlinkability are satisfied during Mesh Client (MC)'s roaming. User accountability is also achieved through PBS-based e-cash system that is incorporated into their mutual authentication protocols together with key agreement features. Their analysis shows that PPS is able to implement desired security objectives and high efficiency.

The sixth paper by Hae-Duck Joshua Jeong, Sang-Kug Ye, Jiyoung Lim, Ilsun You and Woo-Seok Hyun, describes a computer remote control system using speech

recognition technologies of mobile devices and wireless communication technologies for the blind and physically disabled population as assistive technology. These people experience difficulty and inconvenience using computers through a keyboard and/or mouse. The purpose of this system is to provide a way that the blind and physically disabled population can easily control many functions of a computer via speech. The configuration of the system consists of a mobile device such as a smartphone, a Personal Computer (PC) server, and a Google server that are connected to each other. Users can command a mobile device to do something via speech; such as directly controlling computers, writing emails and documents, calculating numbers, checking the weather forecast, or managing a schedule. These commands are then immediately executed. The proposed system also provides blind people with a function via text to speech of the Google server if they want to receive contents of a document stored in a computer.

The seventh paper by Young-Long Chen, Mu-Yen Chen, Fu-Kai Cheung and Yung-Chi Chang, proposes a hybrid architecture based on power-efficient gathering in sensor information system (PEGASIS) and low-energy adaptive clustering hierarchy (LEACH). This architecture can achieve an average distribution of energy loads, and reduced energy consumption in transmission. To further extend the system lifetime, they combine the intersection-based coverage algorithm (IBCA) with LEACH architecture and the hybrid architecture to prolong the system lifetime that introducing sensor nodes to enter sleep mode when inactive. This step can save more energy consumption. Simulation results show that the performance of their proposed LEACH architecture with IBCA and the hybrid architecture with IBCA perform better than LEACH architecture with PBCA in terms of energy efficiency, surviving nodes and sensing areas.

The eighth paper by Taerim Lee, Hyejoo Lee, Kyung-Hyune Rhee and Sang Uk Shin, introduces a Distributed Text Processing System based on Hadoop, called DTPS, and explains about the distinctions between DTPS and other related researches to emphasize the necessity of it. In addition, this paper describes various experimental results in order to find the best implementation strategy in using Hadoop MapReduce for the distributed indexing and to analyze the worth for practical use of DTPS by comparative evaluation of its performance with similar tools. To be short, the ultimate purpose of this study is the development of useful search engine specially aimed at big data indexing as a major part for the future e-Discovery cloud service.

In the ninth paper by Yongman Han, Jongcheon Choi, Seong-Je Cho, Haeyoung Yoo, Jinwoon Woo, Yunmook Nah and Minkyu Park, to detect block and remove pirated software (illegal programs) on Online Service Provider (OSP) and Peer-To-Peer (P2P) networks, they study a new filtering approach using software birthmark, which is unique characteristics of program and can be used to identify each program. Software birthmark typically includes constant values, library information, sequence of function calls, and call graphs, etc. They target Microsoft Windows applications and utilize the numbers and names of Dynamic-Link Libraries (DLLs) and Application Programming Interfaces (APIs) stored in a Windows executable file. Using that information and each cryptographic hash value of the API sequence of programs, they construct software birthmark database. Whenever a program is uploaded or downloaded on OSP and P2P networks, they can identify the program by comparing software birthmark of the program with birthmarks in the database. It is

possible to grasp to some extent whether software is an illegally copied one. The experiments show that the proposed software birthmark can effectively identify Windows applications. That is, their proposed technique can be employed to efficiently detect and block pirated programs on OSP and P2P networks.

The tenth paper by Inshil Doh, Jiyoung Lim, Shi Li and Kijoon Chae, proposes key establishment mechanisms for secure communication among entities in the cellular Machine-to-Machine (M2M) network. Considering the characteristics of cellular M2M networks, traditional security solutions are not proper to be applied to cellular M2M networks because the M2M network itself is vulnerable to various attacks. Their mechanism includes pairwise keys for the M2M communication and the group communication among the M2M Equipments (M2MEs). Their key agreement proposal can provide security and reliability for the cellular M2M communication.

The eleventh paper by Hsing-Chung Chen, Cheng-Ying Yang, Hui-Kai Su, Ching-Chuan Wei and Chao-Ching Lee, describes a new e-mail delivery mechanism using secure multicast key protocol with ID-based factorial number structure (ID-based FNS) in an encryption multicast system. In the e-mail delivery mechanism, the message of e-mail is required to encrypt first before sending out in order to safeguard the secrecy of the message on a public channel, such as wire-lined public switching communication links and wireless communication systems. Without loss generality, the public-key system is usually adopted in the multicast environment for the convenience at the easy key management need. As a manner of fact, the mechanism is having analyzed the space occupation; their scheme outperforms the existed methods in the way of magnitude order of reconstructing secure command key. In addition, the extraction of direct secure command key associated each intended receiver is fast operated to succeed the subsequent e-mail message recovery.

The twelfth paper by Hyoung Woo Park, Il Yeon Yeo, Jongsuk Ruth Lee and Haengjin Jang, presents the important paradigm shifts of network architecture caused by big data traffic. They show the new network architecture which resulted from their experience of the European Organization for Nuclear Research (CERN) Large Hadron Collider (LHC) data service. They also illustrate the effect of the throughput improvements of the proposed network architecture using Network Simulator (NS)-2. An interesting feature of their new approach for network architecture is a kind of recycling-friendly architecture because the proposed architecture requires a plentiful number of legacy network cables and legacy low-end network devices instead of buying expensive and cutting-edge network devices. According to their investigation, the future network architecture of the big data center will be a dual matrix architecture in which the big data part will be located at the front and the center of the architecture in order to reduce the number of interactions between the big data traffic and the legacy traffic.

The thirteenth paper by Guowei Wu, Xiaojie Chen, Lin Yao, Youngjun Lee and Kangbin Yim, proposes wormhole attack detection based on transmission range that exploits the local neighborhood information check without using extra hardware or clock synchronizations. Extensive simulations are conducted under different mobility models. Simulation results indicate that the proposed method can detect wormhole attacks effectively and efficiently in wireless senor networks.

The fourteenth paper by Chien-Erh Weng, Jyh-Horng Wen, Hsing-Chung Chen and Lie Yang, describes the backoff procedure characteristics of the Markov chain

model with direct/cooperative transmission strategies with Request to Send / Clear to Send (RTS/CTS) mechanism. There is no guarantee that the earlier strategy adopting cooperative transmission presents better performance. With the population of multimedia applications in Wireless Local Area Networks (WLANs), they extend the Markov chain model to support the Quality of Service (QoS) requirements. Differentiating the contention window size is better than differentiating the arbitration interframe space in terms of throughput and delay. Nevertheless, differentiating the arbitration interframe space is a fast way to access the channel. This can be explained by the fact that the different contention window durations can differentiate the probability of collision and provide priority, whereas the arbitration interframe space only provides the priority by differentiating the duration that the station accesses channel.

The fifteenth paper by Guangzhi Zhang, Yunchuan Sun, Mengling Xu and Rongfang Bie, proposes a new approach to cluster the Weibo data by analyzing the users' reposting behavior data besides the text contents. To verify the proposed approach, a data set of users' real behaviors from the actual Social Networking Service (SNS) platform is utilized. Experimental results show that the proposed method works better than previous works which depend on the text analysis only.

Finally, in the last paper by Neng-Yih Shih and Hsing-Chung Chen, they study decision approach for selecting candidates in soft-handover procedure in 4th generation mobile communication via grey relational analysis of the series similarity and approximation. The multi-generating and second grey relational analysis procedure is proposed in this paper. It could be applied to the kind of application for selecting candidates in soft-handover procedure; during to the properties of the multi-generating data series are similar to the velocity and acceleration series. With several simulations are validated, the approach could be used to deal with the candidates selecting in soft-handover, and output the best results of feasibility and effectiveness for user equipment in 4th generation mobile communications.

We strongly believe that the papers presented in this special issue make significant contributions to the work and studies conducted by academic researchers, industry professionals, students, and everyone in the areas of mobile collaboration technologies and Internet services.

We would like to thank all the authors for their valuable contributions. Our special thanks go to prof. Mirjana Ivanović, Editor in Chief of the Computer Science and Information Systems (ComSIS) Journal, for inviting us to prepare this special issue and for his productive comments and great support throughout the entire publication process.

Editors of the special issue
Hae-Duck Joshua Jeong (Korean Bible University, Seoul, South Korea)
Fatos Xhafa (Technical University of Catalonia, Barcelona, Spain)
Makoto Takizawa (Hosei University, Tokyo, Japan)

Computer Science and Information Systems 11(3):905–924 DOI: 10.2298/CSIS130921056H

# A True Random-Number Encryption Method Employing Block Cipher and PRNG

Yi-Li Huang, Fang-Yie Leu, Jian-Hong Chen, William Cheng-Chung Chu

Department of Computer Science, Tunghai University,
No. 1727, Section 4, Taiwan Boulevard, Taichung City, Taiwan
{yifung, leufy, g01350027, cchu}@thu.edu.tw

**Abstract.** In January 1999, distributed.net collaborated with the Electronic Frontier Foundation to break a DES (i.e., Data Encryption Standard) key, spending 22 hours and 15 minutes, and implying that the DES is no longer a secure encryption method. In this paper, we propose a more secure one, called the True Random Number Encryption Method (TRNEM for short), which employs current time, true random numbers and system security codes as parameters of the encryption process to increase the security level of a system. The same plaintext file encrypted by the TRNEM at different time points generates different ciphertext files. So these files are difficult to be cracked. We also analyze the security of the DES, AES (i.e., Advanced Encryption Standard) and TRNEM, and explain why the TRNEM can effectively defend some specific attacks, and why it is safer than the DES and AES.

**Keywords:** DES, AES, true random number, SSC, block cipher, wrapped ciphertext file

## 1. Introduction

Due to the popularity of computer systems and network services, the Internet-access security and information security have been a part of the focuses of computer research since when accessing the Internet, users may anytime anywhere face different kinds of attacks [1]. Thus, protecting important data stored in a computer or a cloud system and messages delivered in a network system is a challenge. Data Encryption Standard (DES) [2] and Advanced Encryption Standard (AES) [3,4] were then developed. However the DES has been cracked and the AES may someday be solved, e.g., by differential attack [5] or linear attack [6]. On the other hand, security data is often encrypted by random numbers which play a critical role in information security services, e.g., when employing an one-way hash function [7] to generate message digests, encrypting messages [8], and signing an electronic document with a digital signature [9,10]. Unfortunately, true random numbers are difficult to obtain since it is hard for us to design themin a deterministic way. However, human activities and the information having been collected in a website as well as their description own the characteristics similar to those of a true random number since before reading them, we do not know what has been collected and how they are described. These data often continuously and randomly vary at different time. In fact, we can randomly select a short fragment of the data as true random numbers from a randomly chosen website and use the segment to

encrypt plaintext. In this study, we develop a data protection mechanism, named True Random Numbers Encryption Method (TRNEM for short), which encrypting plaintext by employing true random numbers is a secure encryption approach which is difficult to be cracked by using brute force attacks and ciphertext analyses.

The rest of this paper is organized as follows. Section 2 briefly describes the related studies of this paper, including AES, and DES, and their vulnerabilities. Section 3 introduces the encryption/decryption process of the TRNEM. The security and performance of the TRNEM and the comparison between the TRNEM and the AES are presented in Section 4. Section 5 concludes this paper and outlines our future studies.

## 2. Common Block Cipher

Currently, the most common block cipher modes are the DES and AES.

### 2.1. Data Encryption Standard (DES)

DES [2] is a symmetric block cipher algorithm in which the encryption and decryption details are almost the same. The length of a key is 56 bits (the key is typically expressed as a 64-bit number, but the first eight bits are used for parity check). The DES encrypts a 64-bit plaintext block into a 64-bit ciphertext block. Its key generation process can be mainly divided into two steps, the initial permutation and the inverse permutation.

In the initial permutation step, the 64-bit input block is permuted to generate two outputs L0 and R0, each of which is 32 bits long. After 16 times of iteration, L0 and R0, respectively, become L16 and R16, which are then input to the inverse permutation process to recover these bitsto their original sequence. The result is the corresponding ciphertext block. DES [11] is unsafe because a brute force attack may succeed. Currently, one of its threats is the linear cryptanalysis [12] which collected 243 known plaintexts. The cracking time complexity ranges between $2^{39}$ and $2^{43}$ [13]. But the complexity can be reduced to 1/4 [14] with the help of a chosen-plaintext attack.

Three effective DES attacks, include differential cryptanalysis [15], linear cryptanalysis [12] and Davies' attack [16], which can break the 16 rounds of DES with the time complexity lower than that of a brute-force method.

### 2.2. Advanced Encryption Standard (AES)

AES [17] algorithm was developed based on bit permutation and substitution. It re-arranges the sequence of the original data, and substitutes a data unit by another. As an iterative and symmetric-key block cipher technique with 128, 192, or 256 bits as its key length, AES encrypts a data block with 10 rounds of duplication and transformation. Each round comprises the SubBytes, ShiftRows, MixColumns and AddRoundKey steps, except the final round in which the MixColumns is substituted by an AddRoundKey. Generally, in the AddRoundKey step, each byte of the data is bitwise-xored with a round key.

In the SubBytes step, each byte is substituted by another one following the content of a predefined lookup table. The ShiftRows rotates a row of a state where a state is an AES calculation on a 4×4 column-major order matrix of bytes. The initial value of this matrix is a plaintext block. In the MixColumns step, a column-wise linear transformation is performed by multiplying a constant matrix and the state matrix to produce a new state matrix.

In 2009, the side-channel attack [18,19] successfully cracked an easy version of the AES. But the National Security Agency (NSA) reviewed all the AES finalists, and claimed that all of them were secure enough for U.S. Government non-classified data. But the weak version that has been successfully cracked and the number of encryption loop of this version are almost the same as those of original version. Cryptographers are worrying about the security of the AES. If the penetrating capabilities of some well-known attack are improved, this block encryption system may someday be cracked again.

### 2.3 Block Cipher Mode of Operation

An operation mode is mainly used to encrypt and authenticate delivered messages. An operational model defines the process of encrypting a data block, often based on a given initialization vector (IV for short) as an additional parameter to further enhance the security of the encrypted data.

If different IVs are given, the same plaintext will generate different ciphertext, even though the plaintext is encrypted by using the same key. The purpose is to avoid regenerating the same ciphertext.

The Cipher Block Chaining (CBC), the Propagating Cipher Block Chaining (PCBC), Cipher feedback (CFB), Output feedback (OFB) and Counter (CTR) are block cipher standards having been recognized by the National Institute of Standards and Technology (NIST). With the CBC mode, as shown in the following two statements, a plaintext block $P_i$ is XORed with the ciphertext $C_{i-1}$ generated in the previous encryption round. The XORed result and the encryption key K are then input to the Block-Cipher-Encryption function to produce the ciphertext $C_i$ where $C_0$ is the *IV* of the CBC mode.

$$C_1 = E_K(P_1 \oplus IV)$$

$$C_i = E_K(P_i \oplus C_{i-1}), 2 \le i \le n$$

With the PCBC mode, as illustrated in the following two statements, a plaintext block $P_i$is XORed with $P_{i-1} \oplus C_{i-1}$. The XORed result and the encryption key K are then input to the Block-Cipher-Encryption function to produce the ciphertext $C_i$ where $P_0 \oplus C_0$ is the *IV* of the PCBC mode.

$$C_1 = E_K(P_1 \oplus IV)$$

$$C_i = E_K(P_i \oplus P_{i-1} \oplus C_{i-1}), 2 \le i \le n$$

The following two statements show the encryption process of the CFB mode. The ciphertext generated in the previous round and the encryption K are input to the Block-Cipher-Encryption function. The result is then XORed with plaintext $P_i$ to yield the ciphertext $C_i$ where $C_0$ is the *IV* of the CFB mode.

$$C_1 = E_K(IV) \oplus P_1$$

$$C_i = E_K(C_{i-1}) \oplus P_i, 2 \le i \le n$$

In the OFB mode, $O_{i-1}$ and the encryption key $K$ are input to the Block-Cipher-Encryption function to produce $O_i$. $O_i$is then XORed with plaintext $P_i$ to produce the ciphertext $C_i$where $O_0$ is the *IV*.

$$C_0 = P_i \oplus E_K(IV)$$

$$C_i = P_i \oplus E_K(O_{i-1}), 2 \le i \le n$$

Similar to that of the OFB mode, the CTR mode ciphers a plaintext block with a stream-cipher method. It generates the next key-stream block by using a counter which is often a function of time with a very long repeating cycle. During encryption, the encryption key K and the counter are input to the Block-Cipher-Encryption function. The result is then XORed with plaintext $P_i$ to produce the ciphertext $C_i$. After an encryption round, the counter value is increased by one. The new value is used to encrypt the next plaintext block.

Although these modes provide a security system with data integrity and confidentiality, they are still vulnerable to known plaintext-ciphertext cryptanalysis attacks.

## 3. The Proposed Method

In this section, we first define the parameters and codes used by the TRNEM.

### 3.1. The Parameters

The parameters are as follows.

*File name*: which is the name of the file being encrypted. Its length is 16 characters. If originally the length is longer than 16, we keep the first 16 characters and truncate the remaining ones. However, if the length is shorter than 16, we extend it by duplicating the file name n times until the length is equal to or longer than 16, $n$>1, and then extract the first 16 characters.

*Filename_ext*: which isthe filename extension of the file. Its length is also 16 characters. If originally it is longer than 16, we extral the first 16 and truncate the remaining ones. If the length is shorter than 16, we extend it with the same method as that used to extend its file name. However, if the length is zero, we put 16 *s as the filename_ext.

*SSC*: which stands for system security code.*SSC* has 16 members where *SSC*($i$) is the $i^{th}$

system security code, $1 \leqq i \leqq 16$, and the length of $SSC(i)$ is 128 bits.

$\triangle h$：which is a variable of 11 bits long for indicating the length of a pseudo random number sequence (PRNS), i.e., $1 \leqq \triangle h \leqq 2047$.

$K\triangle h$：which is an encryption key of 128 bits long. It is generated by the concatenation of 12 $\triangle hs$, but discarding the last 4 bits.

*KCT*: which is a current-time encryption key defined as a bit sequence obtained by concatenating the following items, including $\triangle h$, and current values of the system clock which contains nanosecond, second, minute, hour, and nanosecond of the clock, i.e., $KCT=\triangle h$||nanosecond||second||minute|| hour||nanosecond ||$\triangle h$ , where "||" denotes concatenation. $\triangle h$ consists of 4 digits, nanosecond is 9 digits long, each of the remaining items is 2 digits in length and each digit is 4 bits long, i.e., $|KCT| = 128$ bits (= 4+9+2+2+2+9+4=32 digits).

*WI* (Web-Index): We randomly select an URL as the *WI* from those dynamically crawled webpages (named crawled files), $1 \leqq WI \leqq 1023$.

*Sd* (Start-distance): which isthe start point of the encrypting segment extracted from the $WI^{th}$ crawled file. The start point is the $Sd^{th}$ character of the file, $1 \leqq Sd \leqq 1023$.

*TRNS*: which stands for True Random Number Sequence (TRNS). It is the segment extracted from the $Sd^{th}$ character of the $WI^{th}$ web's content.

$\triangle L$: which is the length of TRNS, $1024 \leqq \triangle L \leqq 2047$.

RIG$y$(X): which is the value of the y right-most bits of the key $X$, i.e., if $X=x[1]\ x[2]\ldots x[|X|]$, RIGy($X$)= $x[|X|\text{-}(y\text{-}1)]\sim x[|X|]$, where $x[i]$ is the $i^{th}$ bit of $X$, $i$=1,2,…$|X|$, and $y$=8, 128 or 256, e.g., when $y$=256, RIG256($X$)=$x[|X|\text{-}255]\ \sim x[|X|]$, and when $y$=8, RIG8($X$)=$x[|X|\text{-}7]\sim x[|X|]$. If $X$ is a character string, we treat it as a long bit string by sequentially substituting these characters by their ASCII codes, e.g., if $X$=abc, 616263 will be the corresponding bit string of 24 bits long.

LEF$y$($X$): which is the value of the $y$ left-most bits of $X$, LEF$y$($X$)=$x[1]\sim x[y]$. For example, when $y$=128, LEF128($X$)=$x[1]\sim x[128]$, and when $y$=8, LEF8($X$)= $x[1]\sim x[8]$.

### 3.2. The Equations used to Generate Encryption Keys

The equations employed in this study are as follows.

$$\triangle h = [(\sum_{i=1}^{4} \mathrm{RIG20}(SSC(i)) + \mathrm{RIG20}(\textit{file name})) * (\sum_{i=5}^{8} \mathrm{RIG20}(SSC(i)) + \mathrm{RIG20}(\textit{filename\_ext})) + (\sum_{i=9}^{12} \mathrm{RIG20}(SSC(i)) + \mathrm{LEF20}(\textit{file name})) * (\sum_{i=13}^{16} \mathrm{RIG20}(SSC(i)) + \mathrm{LEF20}(\textit{filename\_ext}))] \bmod 2047 + 1 \quad (1)$$

which randomly varies each time when it is invoked. It is the first parameter adopted by the TRNEM.

$$DA = \text{HMAC}(SSC(1) \oplus KCT \| SSC(2) \oplus KCT \| SSC(3) +_2 K \triangle h \| SSC(4) +_2 K \triangle h, SSC(5) \oplus KCT) \quad (2)$$

which randomly varies each time when it is invoked. It is the first dynamic key employed by the TRNEM.

$$DB = \text{HMAC}(SSC(6) \oplus DA \| SSC(7) \oplus DA \| SSC(8) +_2 KCT \| SSC(9) +_2 KCT, DA +_2 K \triangle h) \quad (3)$$

which randomly varieseach time when it is generated. It is the second dynamic key employed by TRNEM.Eqs. (2) and (3), that respectively generate dynamic keys *DA* and *DB*, together are called Equation-group 1.

$$CDA = [((SSC(10) \oplus DA) +_2 SSC(11)) +_2 (K \triangle h \oplus SSC(12))] \oplus (SSC(13) +_2 K \triangle h) \quad (4)$$

$$CDB = [((SSC(14) \oplus DB) +_2 SSC(15)) +_2 (DA \oplus K \triangle h)] \oplus (SSC(16) +_2 DA) \quad (5)$$

Eqs. (4) and (5), that respectively produce the encrypted dynamic keys *CDA* and *CDB*, together are called Equation-group 2.

$$\triangle L = [\text{LEF12}(SSC(2)) * \text{RIG12}(DA) + \text{LEF12}(SSC(3)) * \text{RIG12}(DB) + (\text{LEF12}(K \triangle h) + \text{LEF12}(DA) + \text{LEF12}(DB)) * \text{LEF12}(SSC(4))] \bmod 1024 + 1024 \quad (6)$$

$$WI = [\text{LEF12}(SSC(5)) * \text{LEF12}(DA) + \text{LEF12}(SSC(6)) * \text{LEF12}(DB) + (\text{LEF12}(K \triangle h) + \text{RIG12}(DA) + \text{RIG12}(DB)) * \text{LEF12}(SSC(7))] \bmod 1023 + 1 \quad (7)$$

$$Sd = [\text{LEF12}(SSC(8)) * \text{LEF12}(DA) + \text{LEF12}(SSC(9)) * \text{LEF12}(DB) + (\text{RIG12}(K \triangle h)^2 + \text{RIG12}(DA)^2 + \text{RIG12}(DB)^2) * \text{LEF12}(SSC(10))] \bmod 1023 + 1 \quad (8)$$

$$Pk_1 = \text{HMAC}(SSC(11) +_2 DA \| SSC(12) +_2 DA \| SSC(13) \oplus DB \| SSC(14) \oplus DB, (SSC(15) +_2 DB) \oplus DA) \quad (9)$$

which as a pseudorandom key is the first pointing key employed by the TRNEM to generate the *PRNS*1, *PRNS*2 and *CTRNS*. Eqs. (6) ~ (9), that respectively generate $\triangle L$, *WI*, *Sd* and $Pk_1$, together are called Equation-group 3.

$E(k, str)$ : An encryption function defined as:

$$E(k, str) = k \oplus s_1 \| k \oplus s_2 \| k \oplus s_3 \| \ldots \| k \oplus s_n, \quad (10)$$

where $str = s_1 s_2 s_3 \ldots s_n$ is a string.

$$TRNS(j) = \text{HMAC}(\text{E}(SSC(j), TRNS) \| E(SSC(17\text{-}j), TRNS), SSC(j+7) +_2 DB), 1 \leqq j \leqq 4 \quad (11)$$

$$\triangle t = (\text{RIG12}(DA)^3 + \text{RIG12}(DB)^3 + \text{LEF12}(DA)^3 + \text{LEF12}(DB)^3 + \text{RIG12}(TRNS(1))^3 + \text{LEF12}(TRNS(1))^3) \bmod 1023 + 1 \quad (12)$$

which as a pesudorandom parameter is the length of $PRNS2$. $\triangle t$ together with $\triangle h$ are adopted to protect the CTRNS and ciphertext in the wrapped ciphertext file.

$$Pk_2 = \text{HMAC}(TRNS(2) \oplus DA \,||\, TRNS(3) \oplus DB \,||\, TRNS(4) +_2 DA, TRNS(1) \oplus DB) \quad (13)$$

which as a pseudorandom key is the second pointing key employed by the TRNEM to generate the ciphertext.

Eqs. (10) ~ (13), that respectively produce $E(k, str)$, $TRNS(1) \sim TRNS(4)$, $\triangle t$ and $Pk_2$, together are called Equation-group 4.

### 3.3. The TRNEM Encryption Process

Fig. 1 illustratively summarizes the encryption flow of the TRNEM. The details are as follows.

**Step 1**: Generating $\triangle h$ and $K\triangle h$. The TRNEM's encryption process invokes the non-invertible $\triangle h$ generation equation defined above to read the file name of the file being encrypted. The file name, filename extension and *SSCs* are the parameters used to produce $\triangle h$ and $K\triangle h$.

**Step 2**: Generating dynamic keys *DA* and *DB*. The TRNEM derives *KCT* from $\triangle h$ and current time (*CT*), and invokes Equation-group 1 which uses $K\triangle h$, *KCT* and *SSCs* as its parameters to produce dynamic keys *DA* and *DB*.

**Step 3**: Encrypting dynamic keys. The TRNEM invokes Equation-group 2 which consisting of two invertible equations defined above employs the generated *DA*, *DB*, *SSCs* and $K\triangle h$ as the parameters to produce *CDA* and *CDB* so that the TRNEM can securely store *CDA* and *CDB* into the wrapped ciphertext file and decrypt *DA* and *DB* from *CDA* and *CDB* carried in the received wrapped ciphertext file.

**Step 4**: Generating $\triangle L$, *WI*, *Sd* and $Pk_1$. The TRNEM invokes Equation-group 3, consisting of four non-invertible equations defined above, to respectively produce $\triangle L$, *WI*, *Sd* and $Pk_1$ by employing the generated *DA*, *DB*, *SSCs* and $K\triangle h$ as input parameters.

**Step 5**: Generating $TRNS(1) \sim TRNS(4)$, $\triangle t$, and $Pk_2$. The TRNEM randomly reads data of $\triangle L$ bytes from the chosen webpage indexed by *WI* and the first character is the $Sd^{th}$ byte of the webpage. These data are our *TRNS*. The TRNEM invokes Equation-group 4 which consisting of some non-invertible equations defined above in turn invokes the generation equations of the *DA*, *DB*, *SSCs* and *TRNS* to produce $TRNS(1) \sim TRNS(4)$, $\triangle t$ and $Pk_2$.

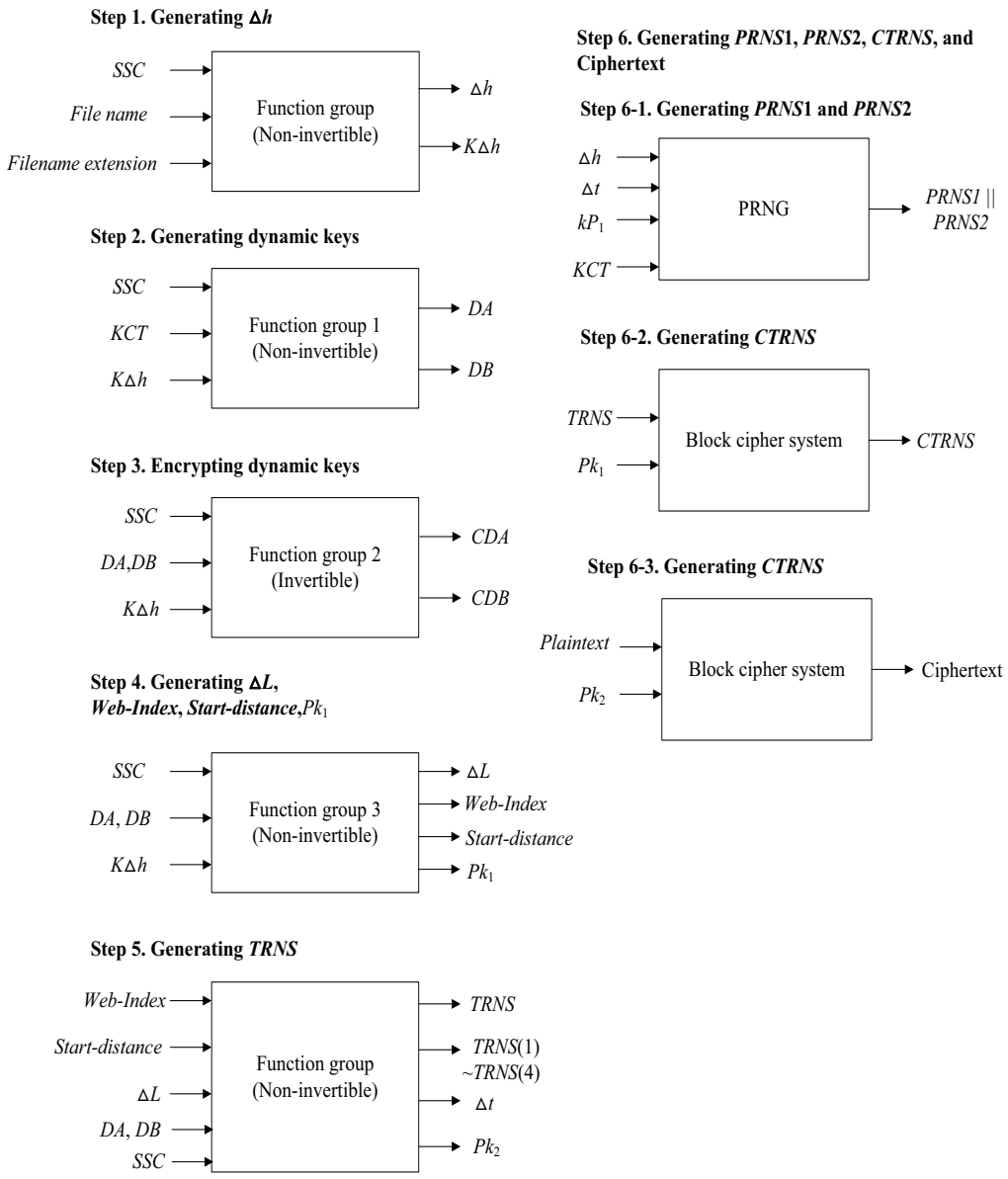



**Fig. 1.** The encryption flow of the TRNEM

**Step 6**: Generating *PRNS*1, *PRNS*2, *CTRNS* and ciphertext.

**Step 6-1**: Generating *PRNS*1 and *PRNS*2. The TRNEM grabs the time parameters from system clock to produce a new *KCT*. After that, *KCT*, $\triangle h$, $\triangle t$ and $Pk_1$are input to the pseudo random number generator (PRNG for short) to produce *PRNS*1 and *PRNS*2.

**Step 6-2**: Generating *CTRNS*. The *CTRNS* is produced by the adopted block cipher system (e.g., AES) with the TRNS as the plaintext and key $Pk_1$ as an input parameter.

**Step 6-3**: Encrypting plaintext (generating ciphertext). A plaintext to be encrypted and key $Pk_2$ are input to the adopted block cipher system to produce the corresponding ciphertext.

**Step 7**: Generating a wrapped ciphertext file. The TRNEM concatenates *PRNS*1, *CDA*, *CDB*, *CTRNS*, the ciphertext generated in Step 6 and *PRNS*2 to produce a wrapped ciphertext file, the format of which is shown in Fig. 2.

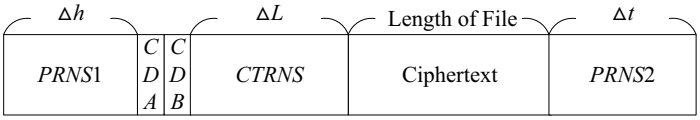



**Fig. 2.** The format of the wrapped ciphertext file generated by the TRNEM

### 3.4. The TRNEM Decryption Process

Fig. 3 illustrates the decryption process of the TRNEM. The details are described below.

**Step 1**: Calculating $\triangle h$ and removing *PRNS*1 from the received wrapped ciphertext file. To decrypt the ciphertext, a user needs to invoke the $\triangle h$ generation equation, which in turn reads the file name and filename extension of the designated file to produce $\triangle h$, with which to delete *PRNS*1 from the wrapped ciphertext file. It further generates $K\triangle h$.

**Step 2**: Retrieving and calculating *DA* and *DB*. Reads *CDA* and *CDB* from the wrapped ciphertext file and decrypts them by using the following two decryption equations, i.e., Eqs. (14) and (15), to obtain the dynamic keys *DA* and *DB*.

$$DA = [CDA \oplus (SSC(13) +_2 K\triangle h)] -_2 (K\triangle h \oplus SSC(12)) -_2 SSC(11) \oplus SSC(10) \quad (14)$$

where $-_2$ is the inverse operation of $+_2$ [20].

$$DB = [CDB \oplus (SSC(16) +_2 DA)] -_2 (DA \oplus K\triangle h)) -_2 SSC(15) \oplus SSC(14) \quad (15)$$

**Step 3**: Calculating $\triangle L$ and retrieving *CTRNS*

(1) Invoking Eq. (6) which employs *SSCs*, $K\triangle h$, *DA* and *DB* as its parameters to calculate $\triangle L$.

(2) Retrieving *CTRNS* from the wrapped ciphertext file based on the calculated $\triangle L$ since *CTRNS* is $\triangle L$ in length.

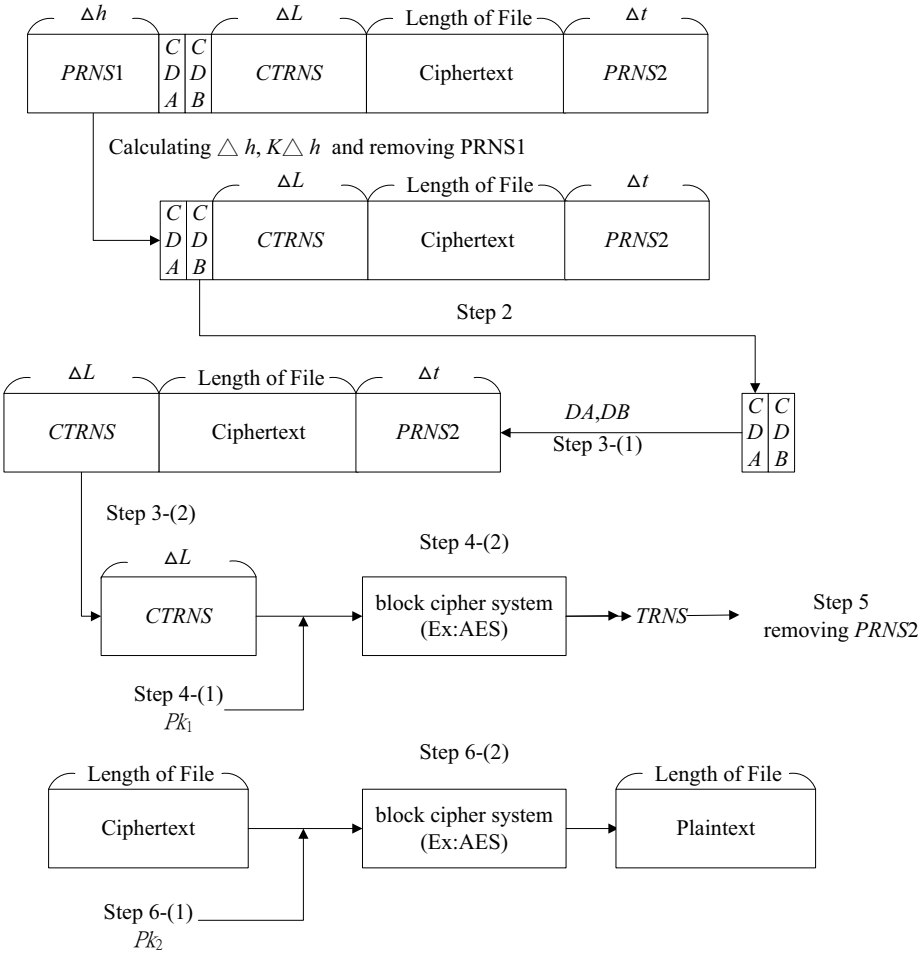



**Fig. 3.** The decryption flow of the TRNEM

**Step 4**: Retrieving *TRNS*. Retrieve *TRNS* by inputting *CTRNS* and $Pk_1$ to the adopted block cipher system.

(1) Producing $Pk_1$ by invoking Eq. (9) which employs*DA*, *DB* and *SSCs* as its parameters.
(2) Invoking the adopted block cipher system to decrypt the *CTRNS* retrieved from the wrapped ciphertext file with $Pk_1$so as to produce *TRNS*.

**Step 5**: Retrieving $\triangle t$ and removing *PRNS*2.

(1) Producing *TRNS*(1) ~ *TRNS*(4) by invoking Eqs. (10) and (11) which employ *SSCs* and *TRNS* as their parameters.
(2) Producing $\triangle t$ by invoking Eq. (12) which utilizes *DA*, *DB* and *TRNS*(1) as its parameters.
(3) Removing *PRNS*2 from the wrapped ciphertext file.

**Step 6**: Generating $Pk_2$ and decrypting the ciphertext.
(1) Invoking Eq. (13) which uses *TRNS*(1)~*TRNS*(4), *DA* and *DB* as its parameters to produce $Pk_2$.
(2) Decrypting the plaintext from the ciphertext by inputting $Pk_2$ and the ciphertext to the adopted block cipher system so as to revert the plaintext.

### 3.5. The Features and Advantanges of the TRNEM

The TRNEM has five features, including

(1) employing filename, filename extension, system security codes and a non-invertible equation to generate $\triangle h$, making $\triangle h$ be one with high security;
(2) utilizing current time to produce dynamic keys *DA* and *DB* which are different when they are generated at different time points since current time continuously varies;
(3) using the *DA* and *DB* to fetch the true random numbers, with which to encrypt the plaintext so as to enhance the security of the ciphertyext;
(4) employing scalable parameters $\triangle h$, $\triangle L$ and $\triangle t$, with which to construct a wrapped ciphertext file. The purpose is enhancing the security of the ciphertext file;
(5) the *CDA*, *CDB* and CTRNS are embedded in the wrapped ciphertext file to effevtively protect the ciphertext.

Beside the mentioned security features, based on the Interent as its data pool, the TRNEM creates a true random number sequence to encrypt plaintext so that the ciphertext has a very high degree of security. Furthermore, the ciphertext is embedded in the position located between *PRNS*1 and *PRNS*2. Hackers cannot directly obtain the (plaintext, ciphertext) pairs from the wrapped ciphertext file, thus highly enhancing the security of the TRNEM.

## 4. Security and Performance Analysis

In this section, we analyze the security levels of different TRNEM parameters and generated data, including $\triangle h$, dynamic keys *DA* and *DB*, the *TRNS*, a wrapped ciphertext file, and $Pk_2$. We also evaluate the security and performance of the TRNEM.

### 4.1. Security of $\Delta h$

There are three major reasons to say that $\triangle h$ possess high security. According to Eq. (1), without the 16 system security codes *SSC*(1) – *SSC*(16), hackers cannot correctly calculate $\triangle h$, even though they have caught the file name and filename extension. Also, the value generated for each term contained in Eq. (1) is larger than $2^{20}$ which is very larger than the upper limit of $\triangle h$, i.e., 2047. After that, the value generated before the modulus operation is reduced to $\triangle h$ through the non-invertible modulus equation,

implying that $\triangle h$ has high randomness and security. Furthermore, the $\triangle h$ is an internal variable of the TRNEM. Hackers cannot derive it from the ciphertext and solve it.

Authough, hackers can try a lot of file names and filename extensions to respectively substitute for the original file name and filename extension contained in the encryption expression, attempting to analyze the possible $\triangle h$. However, the length of the wrapped ciphertext file is $\triangle h$+32+$\triangle L$+|the plaintext| + $\triangle t$ bytes, in which 32 is the length of *DA*+*DB*, and $\triangle L$ and $\triangle t$ randomly change at each encryption, even the file name, filename extension and plaintext remain unchanged. $\triangle h$ is well protected due to the dynamic values of $\triangle L$ and $\triangle t$. Now, we dare to say that $\triangle h$ and the protected system are very safe.

### 4.2. Security of Dynamic Keys DA and DB

There are two methods to obtain *DA*. First, hackers may directly generate *DA* by employing Eq. (2). However, *SSC*(1) ~ *SSC*(5), $K\triangle h$ and *KCT* are unknown to hackers and *KCT* continuously changes at each encryption. That means hackers cannot directly generate *DA* by employing Eq. (2). Second, hackers may crack *CDA* to obtain *DA*. However, according to Eq. (4), they need *SSC*(10)~*SSC*(13) and $K\triangle h$. But these parameters are unknown to hackers. Furthermore, *CDA* is embedded in the wrapped ciphertext file. Hackers need $\triangle h$ to correctly fetch it. But $\triangle h$ is unknown to hackers. Thus, *DA* is secure. Similarly, to derive *DB* from *CDB*, hackers need *SSC*(14) ~ *SSC*(16), $K\triangle h$ and *DA* which are unknown to hackers, i.e., *DB* is secure.

### 4.3. Security of the TRNS

The TRNEM collects a webpage based on a randomly chosen *WI*, and accesses the content of the webpage from the position indicated by the Sd to the position pointed to by $Sd + \triangle L$. In other words, $|TRNS| = \triangle L$. But the characteristics and contents of different pages vary with time, and the page contents may be changed frequently. Under this circumstance, extracting web contents from a randomly chosen webpage can make a number sequence, i.e., the *TRNS*, truly random.

Hackers may obtain *TRNS* by decrypting *CTRNS* embedded in the wrapped ciphertext file. However, to fetch the *CTRNS*, parameters $\triangle h$, $\triangle L$, length of the plaintext and $\triangle t$ are required. But, hackers cannot obtain them from the wrapped ciphertext file. That is, hackers cannot correctly fetch *CTRNS* from this wrapped ciphertext file. Furthermore, if CTRNS is known by hackers, they still cannot decrypt *CTRNS* to obtain *TRNS* since $PK_1$ is unknown to them. So, *TRNS* is secure.

### 4.4. Security of a Wrapped Ciphertext File

This system adopts a wrapping ciphertext approach, in which the ciphertext as shown in Fig. 2 is wrapped by *PRNS*1 of length $\triangle h$, *CTRNS* of length $\triangle L$, and *PRNS*2 of length

$\triangle t$. Parameters $\triangle L$ and $\triangle t$ are different at each encryption even though the plaintext is the same. Hackers cannot obtain $\triangle h$, $\triangle L$ and $\triangle t$ to unwrap the ciphertext, i.e., hackers cannot collect (plaintext, ciphertext) pairs when plaintext is known. They need to crack $\triangle h$ before solving other parameters, meaning that the ciphertext file is securely protected by the TRNEM.

### 4.5. Security of the $Pk_2$

$Pk_2$ as a pseudorandom key is the second pointing key employed by the TRNEM to generate the ciphertext. The security of the ciphertext strongly depends on the security of $Pk_2$ and the block cipher system. In the following, we would like to identify the security level of $Pk_2$. Theorem 1 proves that its security level is the same as that when it is solved by using a blind guess method.

**Theorem 1.** If the key length of the TRNEM is n bits, then the probability of solving the correct value of $Pk_2$ from the wrapped ciphertext file is $1/2^n$.

*Proof.* $Pk_2$ as an internal pseudorandom key used by the TRNEM does not appear in the wrapped ciphertext file. Hackers cannot directly break it. Two methodscan be used to break $Pk_2$, excluding the blind guess approach. The first is that,according to Eq. (13), i.e., $Pk_2$ = HMAC($TRNS(2)\oplus DA \parallel TRNS(3)\oplus DB \parallel TRNS(4) +_2 DA$, $TRNS(1)\oplus DB$), only the one who knows$DA$, $DB$, $TRNS(1) \sim TRNS(3)$ can correctly generate $Pk_2$. However, the dynamic keys $DA$ and $DB$ are secure and the true random number sequences, i.e., $TRNS(1) \sim TRNS(3)$, derived from$TRNS$ and $DB$ are secure, too, based on the abovementioned description. The dynamic keys $DA$ and $DB$, which are functions of $KCT$, vary randomly each time when it is generated, implying that the generated messages, $TRNS$ and hence, $TRNS(1) \sim TRNS(3)$ change randomly each time when they are produced so that $Pk_2$is secure.

The second method is breaking the block cipher system to obtain plaintext from the ciphertext. However, the ciphertext embedded in the position located between $PRNS1$ and $PRNS2$ is secure, according to that described in section 4.4. That is, hackers cannot break $Pk_2$ from the ciphertext. In the worst case, if the ciphertext is known by the hackers, they need to break the block cipher system. But this is not an easy work since hackers require a massive amount of (plaintext, ciphertext) pairs given the same parent key. Hence they still cannot break the block cipher system to obtain $Pk_2$.

There are no useful method to obtain $Pk_2$ other than the blind guess approach. Therefore, the probability of solving the correct value of$Pk_2$ from the wrapped ciphertext is $1/2^n$. Q.E.D.

### 4.6. Security of the TRNEM

Due to involving current time and TRNS, the encryption results generated on the same plaintext at different time points vary, implying that TRNEM can effectively prevent those linear cryptanalysis attacks [21,22]. In the TRNEM, the mechanism that wraps a

ciphertext file can effectively defend the known plaintext attacks because hackers cannot correctly collect different (plaintext, ciphertext) pairs.

In fact, the TRNEM integrates time variables, i.e., the current time and TRNS. So the wrapped-ciphertext-file mechanism can effectively prevent the protected system from brute force attacks.

### 4.7. Generation Times of Parameters

The TRNEM generates ciphertext by using a block cipher system (e.g., AES or DES). To generate a wrapped ciphertext file, we produce several parameters introduced above. Table 1 lists the times required to produce these parameters. We also used these parameters to produce the *PRNS*1, *PRNS*2, *CDA*, *CDB* and *CTRNS*. Table 2 lists the times required to produce them and the wrapped ciphertext file.

**Table 1.** The times required to produce different required parameters

| Parameter | Parameter generation time (ms) |
|---|---|
| Eq. (1):$\triangle h$ | 0.01948 |
| Eq. (2):*DA* | 0.30646 |
| Eq. (3):*DB* | 0.27196 |
| Eq. (4):*CDA* | 0.23230 |
| Eq. (5):*CDB* | 0.23624 |
| Eq. (6):$\triangle L$ | 0.00963 |
| Eq. (7):*WI* | 0.00958 |
| Eq. (8):*Sd* | 0.01020 |
| Eq. (9):$Pk_1$ | 0.42009 |
| Eq. (11):*TRNS*(1) | 15.06665 |
| Eq. (11):*TRNS*(2) | 14.96699 |
| Eq. (11):*TRNS*(3) | 15.41611 |
| Eq. (11):*TRNS*(4) | 15.52798 |
| Eq. (12):$\triangle t$ | 0.00820 |
| Eq. (13):$Pk_2$ | 0.28322 |
| Total | 62.78509 |

**Table 2.** The time required to produce the wrapped ciphertext file

| Item | Generation time (ms) |
|---|---|
| *PRNS*1\|\|*PRNS*2(length of$\triangle h$+$\triangle t$) | 20 |
| *CDA* | 0.2323 |
| *CDB* | 0.23624 |
| *CTRNS*(length of$\triangle L$) | 13 |
| Ciphertext | The same as the time required by AES or DES |

No matter what size of the file to be encrypted is, the times the TRNEM spent to generate *PRNS*1, *PRNS*2, *CDA*, *CDB* and *CTRNS* are themselves the same. Compared with other block cipher techniques, it only takes a very short extra time to encrypt a file. But the security level on the contrary dramatically increases.

Since the lengths of *PRNS*1||*PRNS*2 (i.e., $\triangle h+\triangle t$) and *CTRNS* (i.e., $\triangle L$) are not fixed, we individually chose the max lengths of them to calculate their generation times. Ciphertext is encrypted by block ciphering. Its generation time is the same as those of the adopted block cipher system, e.g., AES and DES. The extra time required by the TRNEM is 95.78529 ms.

The difference between the decryption process and the encryption process of the TRNEM is that when decrypting the ciphertext file, *DA* and *DB* are acquired by invoking Eqs. (4) and (5), which further invoke the invertible equations to generate *CDA* and *CDB* where *CDA* and *CDB* are retrieved from the wrapped ciphertext file. *TRNS* is decrypted by inputting *CTRNS* and $\mathrm{Pk}_1$ to the adopted block cipher system where *CTRNS* is also retrieved from the wrapped file. Since the formulas used to generate other parameters for decryption are the same as those when encrypting the plaintext file, the times required to produce $\triangle h$, $\triangle L$, $Pk_1$, *TRNS*(1), *TRNS*(2), *TRNS*(3), *TRNS*(4), $\triangle t$ and $Pk_2$ are then individually the same as those when encrypting the file. Table 3 lists the times required to produce parameters for decrypting *CTRNS*, and Table 4 shows the ciphertext decryption time.

**Table 3.** The times required to produce different parameters for decrypting the wrapped ciphertext file

| Parameter | Parameter generation time (ms) |
|---|---|
| Eq. (1):$\triangle h$ | 0.01948 |
| Eq. (4):*CDA* | Reads *CDA* from the wrapped file |
| Eq. (5):*CDB* | Reads *CDB* from the wrapped file |
| Eq. (14):*DA* | 0.19569 |
| Eq. (15):*DB* | 0.19926 |
| Eq. (6):$\triangle L$ | 0.00963 |
| Eq. (9):$Pk_1$ | 0.42009 |
| Eq. (11):*TRNS*(1) | 15.06665 |
| Eq. (11):*TRNS*(2) | 14.96699 |
| Eq. (11):*TRNS*(3) | 15.41611 |
| Eq. (11):*TRNS*(4) | 15.52798 |
| Eq. (12):$\triangle t$ | 0.0082 |
| Eq. (13):$Pk_2$ | 0.28322 |
| Total | 62.1133 |

**Table 4.** The times required to decrypt the *CTRNS* and the wrapped ciphertext file

| Item | Generation time (ms) |
|---|---|
| *TRNS* | 13 |
| plaintext | The same as the time required by AES or DES |

Table 5 lists the computational efforts in terms of different numbers of operations employed by the encryption/ decryption processes of the TRNEM.

**Table 5.** All computational efforts in terms of different numbers of operations employed by the encryption/ decryption processes of the TRNEM

| TRNEM | Encryption | Decryption |
|---|---|---|
| Eq. (1):$\triangle h$ | 18+s + 2*s + 1 mod | 18+s + 2*s + 1 mod |
| Eq. (2):*DA* | 3$\oplus$s (128 bits) + 2+$_2$s (128 bits) + 1HMAC | does not generate |
| Eq. (3):*DB* | 2$\oplus$s (128 bits) + 3+$_2$s(128 bits) + 1HMAC | does not generate |
| Eq. (4):*CDA* | 3$\oplus$s (128 bits) + 3+$_2$s(128 bits) | does not generate |
| Eq. (5):*CDB* | 3$\oplus$s (128 bits) + 3+$_2$s(128 bits) | does not generate |
| Eq. (6):$\triangle L$ | 3*s + 4+s + 1 mod | 3*s + 4+s + 1 mod |
| Eq. (7):*WI* | 3*s + 4+s + 1 mod | does not generate |
| Eq. (8):*Sd* | 6*s + 4+s + 1 mod | does not generate |
| Eq. (9): $Pk_1$ | 3$\oplus$s (128 bits) + 3+$_{2s}$ (128 bits) + 1HMAC | 3 $\oplus$ s (128 bits) + 3+$_2$s(128 bits) + 1HMAC |
| Eq. (11): *TRNS*(1) | 2$E$(*k*,*str*) + 1+$_{2s}$ (128bits) + 1HMAC | 2$E$(*k*,*str*) + 1+$_2$s (128bits) + 1HMAC |
| Eq. (11): *TRNS*(2) | 2$E$(*k*,*str*) + 1+$_{2s}$ (128bits) + 1HMAC | 2$E$(*k*,*str*) + 1+$_2$s (128bits) + 1HMAC |
| Eq. (11): *TRNS*(3) | 2$E$(*k*,*str*) + 1+$_{2s}$ (128bits) + 1HMAC | 2$E$(*k*,*str*) + 1+$_2$s (128bits) + 1HMAC |
| Eq. (11): *TRNS*(4) | 2$E$(*k*,*str*) + 1+$_{2s}$ (128bits) + 1HMAC | 2$E$(*k*,*str*) + 1+$_2$s (128bits) + 1HMAC |
| Eq. (12):$\triangle t$ | 18*s + 5+$_2$s + 1mod | 18*s + 5+$_2$s + 1mod |
| Eq. (13): $Pk_2$ | 3$\oplus$s (128 bits) + 1+$_{2s}$ (128 bits) + 1HMAC | 3 $\oplus$ s (128 bits) + 1+$_2$s (128 bits) + 1HMAC |
| Eq. (14): *DA* | does not generate | 3 $\oplus$ s (128 bits) + 1+$_2$s (128 bits) + 2-$_2$s (128bit) |
| Eq. (15): *DB* | does not generate | 3 $\oplus$ s (128 bits) + 1+$_2$s (128 bits) + 2-$_2$s (128bit) |
| *CTRNS*/*TRNS* | The same as the time of AES or DES | The same as the time of AES or DES |
| Ciphertext/ plaintext | The same as the time of AES or DES | The same as the time of AES or DES |

### 4.8. Performance Analysis

Table 6 summarizes the computational efforts required by the DES, AES, and TRNEM to encrypt and decrypt a data file.

**Table 6.** The summary of the computational efforts required by the DES, AES and TRNEM to encrypt and decrypt a data file.

| Scheme | Encryption | Decryption |
|---|---|---|
| DES (64-bit block) [23,24] | 16 $\oplus$s (32 bits) + 16 $\oplus$s (48 bits) + 1 IP (64 bits) + 1 IP-1 (64 bits ) + 128 S-Box (6 bits) + 16 Expansions (48 bits) + 16 Permutations (32 bits) | The number of operations is the same as that of the encryption process. |
| AES (128-bit block, 128-bit key) [25] | (AddRoundKey)<br>176 $\oplus$s (8 bits)<br>(SubBytes)<br>160 Substitutions (8 bit) [26]<br>(ShiftRows)<br>30 ShiftRows (128 bit)<br><br>(MixColumns)<br>36 Rijndael columns mixing [26] (128 bits) | The number of operations is the same as the sum of the numbers of those operations employed by the encryption process for the three stages, including AddRoundKey, SubBytes, and ShiftRows<br>(MixColumns)<br>36 Rijndael columns mixing [27] (128 bits).<br>(Generally, the operations of a decryption process are often more complex than those of the corresponding encryption process.) |
| TRNEM | 30+s + 32*s + 17$\oplus$s (128 bits) + 24+$_2$s (128 bits) + 8 $E(k,str)$ + 8 HMAC + 5 mod + 2*(176 $\oplus$ s (8 bits) +160 Substitutions (8 bit)+ 30 ShiftRows +36 Rijndael columns mixing ) in which the last term 2*(176 $\oplus$s …) is the time required to produce *CTRNS* from *TRNS* and generate ciphertext from plaintext | 22+s + 23*s + 12$\oplus$s (128 bits) + 15+$_2$s (128 bits) + 4-$_2$s(128bit) + 8 $E(k,str)$ + 6 HMAC + 3 mod +2*(176 $\oplus$ s (8 bits) +160 Substitutions (8 bit)+ 30 ShiftRows +36 Rijndael columns mixing ) in which the last term 2*(176 $\oplus$ s …) is the time required to produce *TRNS* from *CTRNS* and generate plaintext from ciphertext |

The following analyses show that TRNEM is more secure than the AES. First, the plaintext is encrypted by the pseudorandom key $PK_2$ when the TRNEM employs the adopted block cipher system. If the block cipher system is the AES, then the TRNEM is still more secure than it since, by Theorem 1, $PK_2$ varies at each encryption, whereas the parent key adopted by the AES is fixed for encrypying a file.Second, the ciphertext of the TRNEM is embedded in a wrapped ciphertext file. It is not easy for hackers to correctly fetch the ciphertext and analyze it. But AES does not have this protection mechanism.

Third, the AES suffers brute force attacks, e.g., the known plaintext/ciphertext attack [28], chosen plaintext attack, such as differential cryptanalysis attack [30], and linear cryptanalysis attack [21,22] since the AES is an combinatorial-logic style encryption method [29]. However, in the TRNEM, when a plaintext block is encrypted at different time points, different current time key*KCTs*and hence different other keys, including *DA*, *DB*, $Pk_1$, *TRNS*(1)~*TRNS*(4) and $PK_2$, are produced, thus resulting in diffenent wrapped ciphertext files. The value of*KCT* randomly changes and has no regular rule. Hence, the following keys generated, including *DA*, *DB*, $Pk_1$, *TRNS*(1) ~ *TRNS*(4) and $PK_2$, also randomly vary. Therefore, they can effectively defend the abovementioned attacks. In summary,*KCT*and *TRNS*are the two keys making the TRNEM more secure than the AES.

As shown in Fig. 2, due to concatenating *PRNS*1, *CDA*, *CDB*, *CTRNS* and *PRNS*2, and the lengths of them are, respectively, $\triangle h$, |*CDA*|, |*CDB*|, $\triangle L$ and $\triangle t$. Therefore, the data transmission efficiency of the TRNEM is

$$\frac{|ciphertext|}{\triangle h + |CDA| + |CDB| + \triangle L + \triangle t + |ciphertext|}$$

## 5. Conclusions and Future Work

This system utilizes a wrapping ciphertext approach, which prevents hackers from identifying the correct position of ciphertext. So the hackers cannot easily crack the protected ciphertext. Additionally, the TRNEM encrypts plaintext by using *TRNS*, which is highly random by randomly choosing a webpage and randomly accessing its content $\triangle h$ in length. Moreover, even though given the same plaintext, the TRNEM generates different ciphertext at different time points. This can effectively prevent hackers from issuing known plaintext/ciphertext attacks. So we dare to say that the TRNEM is very secure.

However, a portable encryption/decryption system, like DES and AES, does not create system parameters in it. To develop an algorithm, with which the system security codes in the TRNEM can be generated by the input password or parent key, is necessary and important. Furthermore, to enhance the performance of the TRNEM, the block cipher system adopted by the TRNEM does not need to be DEA or AES. To develop a secure and efficient encryption/decryption method, we plan to utilize the keys generated by the TRNEM, e.g.,$K\triangle h$, *DA*, *DB*, $PK_1$, $PK_2$ and *SSCs*, as the parameters to establish a new block cipher system, which is then substituted for the AES or DES to perform the block ciphering for the TRNEM. These constitute our further studies.

**Acknowledgments.** The work was partially supported by TungHai University under the project GREENs and the National Science Council, Taiwan under Grants NSC 102-2221-E-029-003-MY3, NSC 101-2221-E-029-003-MY3 and NSC 100-2221-E-029-018.

## References


1. Wiki, Computer insecurity, http://en.wikipedia.org/wiki/Computer_insecurity
2. Category, G.M.:The PPP DES Encryption Protocol. RFC 2419,September 1998, Version 2 (DESE-bis)
3. Daemen, J., Rijmen, V.: The Design of Rijndael: AES The Advanced Encryption Standard. New York, USA, Springer-Verlag.(2002)
4. Prodanović, R., Simić, D.: Holistic Approach to Wep Protocol in Securing Wireless Network Infrastructure. Computer Science and Information Systems, vol. 3, issue 2, 97-113.(2006)
5. Bahrak, B., Aref, M.R.: Impossible Differential Attack on Seven-round AES-128. Published in IET Information Security, vol. 2, Issue 2, 28 – 32. (2008)
6. Wiki, Linear cryptanalysis, http://en.wikipedia.org/wiki/Linear_cryptanalysis
7. Li, P., Sui, Y., Yang, H., Li, P.: The Parallel Computation in One-Way Hash Function Designing. International Conference on Computer, Mechatronics, Control and Electronic Engineering, Conference, vol. 1, 189 - 192. (2010)
8. Wang, M., Zhu, G., Zhang, X.: General Survey on Massive Data Encryption.International Conference on Computing Technology and Information Management, vol. 1, 150- 155.(2012)
9. Kaur, R., Kaur, A.: Digital signature. International Conference on Computing Sciences, 295-301. (2012)
10. Živković, Z.V., Stanojević, M.J.: Simulation Analysis of Protected B2B e-commerce Processes. Computer Science and Information Systems, vol. 3, issue 1, 77-91. (2006)
11. Wiki, DES, http//en.wikipedia.org/wiki/Data_Encryption_Standard
12. Matsui, M.: The First Experimental Cryptanalysis of the Data Encryption Standard. In Advances in Cryptology CRYPTO'94, Lecture Notes in Computer Science 839, Springer Verlag, 1-11. (1994)
13. Junod, P.: On the complexity of Matsui's attack. Selected Areas in Cryptography, Lecture Notes in Computer Science 2259, 199-211. (2001)
14. Knudsen, L.R., Mathiassen, J.E.: A Choice Plaintext Linear Attack. DES Fast Software Encryption, 62-272. (2000)
15. Biham, E., Shamir, A.: Differential Cryptanalysis of the Data Encryption Standard - Advances in Cryptology.The Annual International Cryptology Conference, CRYPTO '92, 487-496. (1992)
16. Biham, E., Biryukov, A.: An Improvement of Davies' attack on DES. Journal of Cryptology, vol. 10, no. 3, 195-206. (1997)
17. National Institute of Standards and Technology, Advanced Encryption Standard, NIST FIPS PUB 197.(2001)
18. Bernstei, D.J.: Cache-timing Attacks on AES. Citeseer. 2005.04. http://cr.yp.to/antiforgery/cachetiming-20050414.pdf
19. Tromer, E., Osvik, D.A., Shamir, A.: Efficient Cache Attacks on AES, and Countermeasures. Journal of Cryptology, vol. 23, Issue 1, 37-71.(2010)
20. Huang, Y.L., Leu, F.Y., Wei, K.C.: A Secure Communication over Wireless Environments by using a Data Connection Core. Mathematical and Computer Modeling, vol. 58, issues 5–6, 1459–1474. (2013)
21. Matsui, M.: Linear cryptanalysis method for DES cipher, in Advances in cryptography - Eurocrypt 1993. Springer-Verlog, Berlin, 386-397.(1993)
22. Biham, E.: On Matsui's Linear Cryptanalaysis. Springer-Verlag 1998, 341-344.(1998)
23. Katz, J., Lindell, Y.: Introduction to Modern Cryptography, Chapman & Hall/CRC Press.(2008)
24. Bellare, M., Rogaway, P.: Introduction to Modern Cryptography, Chapter 3, May 11, 2005. http://digidownload.libero.it/persiahp/crittografia/2005_Introduction_to_Modern_Cryptography.pdf
25. Federal Information Processing Standards Publication 197: Announcing the Advanced Encryption Standard (AES).(2001)

26. Cui L., Cao, Y.: A New S-Box Structure Named Affine-Power-Affine. International Journal of Innovative Computing, Information and Control, vol. 3, no. 3, 751-759. (2007)
27. Daemen, J., Rijmen, V.: AES Proposal: Rijndael. The First Advanced Encryption Standard Candidate Conference, NIST, 1999.
28. Wiki, Known-plaintext attack, http://en.wikipedia.org/wiki/Known-plaintext_attack
29. Qaosar,M., Ahmad, S.: A Combinational Logic Approach by using HDL to Implement DES Algorithm. Canadian Journal on Electrical and Electronics Engineering, vol. 3, no. 7, 384-389. (2012)
30. Biham, E., Shamir, A.: Differential Cryptanalysis of the Full 16-round DES. In E. F. Brickell, editor, Advanced in Cryptology-Crypto'92, vol. 740 of *Lectures* Notes in Computer Science, 487-496. (1992)

**Yi-Li Huang** received his master degrees from National Central University of Physics, Taiwan, in 1983. His research interests include security of network and wireless communication, solar active-tracking system, pseudo random number generator design and file protection theory. He is currently a senior instructor of Tunghai University, Taiwan, and director of information security laboratory of the University.

**Fang-Yie Leu** received his B.S., M.S. and Ph.D. degrees from National Taiwan University of Science and Technology, Taiwan, in 1983, 1986 and 1991, respectively, and another M.S. degree from Knowledge Systems Institute, USA, in 1990. His research interests include wireless communication, network security, Grid applications and Chinese natural language processing. He is currently a full professor of Tunghai University, Taiwan, the director of database and network security laboratory of the University, the chair of MCNCS and CWECS workshops, and the editorial board member of several international journals. He is also a member of IEEE Computer Society.

**Jiang-Hong Chen** graduated from Computer Science Department, Tunghai University, Taiwan, in 2012. He is now a master student of this department. His research interests include wireless communication and network security.

**William C. Chu**, the Director of Software Engineering and Technologies Center of Tunghai University, a professor of the Department of Computer Science, he had served as the Dean of Engineering College at Tunghai University, Taiwan. From 2008 to 2011, Dean of Research and Development office at Tunghai University from 2004 to 2007, Taiwan. In 1992, he was also a visiting scholar at Stanford University. His current research interests include software engineering, embedded systems, and E-learning. Dr. Chu received his MS and PhD degrees from Northwestern University in Evanston Illinois, in 1987 and 1989, respectively, both in computer science.

*Received: September 21, 2013; Accepted: January 21, 2014.*

Computer Science and Information Systems 11(3):925–941 DOI:10.2298/CSIS130919057C

# A Secure Mobile DRM System Based on Cloud Architecture

Chin-Ling Chen[1], Woei-Jiunn Tsaur[2], Yu-Yi Chen[3] and Yao-Chung Chang[1]

[1]Department of Computer Science and Information Engineering
Chaoyang University of Technology,
Taichung, 41349, Taiwan
{clc@mail.cyut.edu.tw; cyc200@gmail.com}
[2]Department of Information Management
Da-Yeh University,
Changhua, 51591, Taiwan
wjtsaur@mail.dyu.edu.tw
[3]Department of Management Information systems
National Chung Hsing University,
Taichung, 402, Taiwan
chenyuyi@nchu.edu.tw

**Abstract.** Public cloud architecture offers a public access software service. Users can login to access the cloud resources via various devices. The main advantage of the SaaS (Software as a Service) cloud service is that it supports different software and devices, in order to open web browsers, to authenticate the users through the standard format. E-books are protected by digital rights management (DRM), and users can use mobile devices to read them. However, the users' identity need to be authenticated or the communication between the user and the cloud server will be at risk. The processes by which users submit their proof of identity to the cloud needs to be protected. In this paper, information security can be achieved efficiently via cloud server architecture and a cryptography mechanism. The proposed scheme focuses on using a mobile device to access the cloud service. The DRM mechanisms can protect digital content; once the mobile users pass the authentication they can access the cloud services, with authenticated users able to easily use mobile devices to read digital content.

**Keywords:** Cloud, DRM, Authentication, Mobile Devices, Security

## 1. Introduction

First, we introduce cloud architecture, the DRM concept of cloud architecture, and the analysis of DRM implementation using mobile devices.

### 1.1 Cloud Architecture

As long as information is stored in a cloud, users can access the cloud service through the Internet and mobile devices[1,2] anytime and anywhere. The user does not need to

know what kind of cloud architecture is present (such as cluster computing, grid computing, distribution computing, etc). The user need only send the request to the cloud and it will perform the most efficient operations.

The early goals of cloud architecture were to combine many computers of distributed computations via the Internet. The running program was divided into many threads and distributed into many computers for execution, with the result being presented immediately. Cloud architecture was gradually developed into service-oriented applications, with users being able to use the cloud properties: permanently available, fast computing, etc, with simple steps such that users could access the services provided by the cloud [3].

In the early stages, users communicated with different devices provided by the cloud architecture, and the cloud structure communication services needed to be robust. Current cloud structure has adopted a hierarchical structure. The top of the user services request message is forwarded and handled by the internal framework. Users do not directly communicate with the internal structure of clouds, and this ensures internal safety; this is called object-to-object architecture [3], and is distinct from the early host-to-host architecture. The present cloud structure can be divided into the following three modes [4,5]; the structure is shown in Figure 1.

- Public Cloud
- Private Cloud
- Hybrid Cloud

When the user's request message passes through the interface of a public cloud, malicious packets are filtered out by the firewall. The authenticated user's request will be forwarded to the API server. The API server need not be in the same geographical region. For example, when a Google Docs file is stored in a US database, the document can be opened by other people to co-edit it, and other users may edit the same document in different countries; however, all of them use the service through the same API.

On the other hand, a private cloud is an internal self-management systems database which develops and maintains the normal operation of the API. The network is connected through a local network; this incurs greater cost for small and medium enterprises. Thus, the Hybrid Cloud was developed. The Hybrid Cloud structure acts as a proxy server in most enterprises. Its main goal is to identify staff identification. In this way, it allows enterprises to control their own staff permissions, while the database is maintained by the provider. Costs are there by reduced.

### 1.2 Cloud Services Model

If cloud services are provided by a single industry, the cloud may not be able to satisfy all of a user's requirements. Thus, a user may use cloud network services provided by different industries. Cloud size can be divided into the following modes: Domestic clouds and Transborder clouds [6].

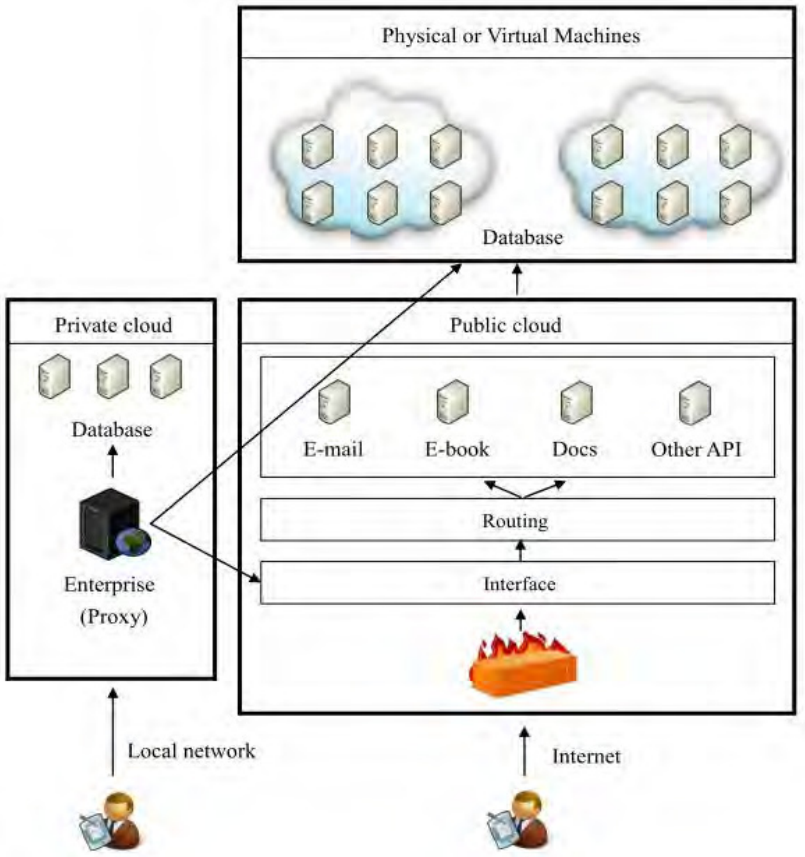



**Fig. 1.** Cloud architecture

(1) Domestic clouds: The entire cloud is physically located within one jurisdiction. The provider provides devices or data exclusively for specific enterprises. The provider need not provide additional service via third party provider to a specific provider, with the resultant advantages of uniform size and high data security.
(2) Transborder clouds: Devices can transmit data to a server (such as Google). Although the Google servers may be located in different countries, users can determine which server stores the data, even if they cannot find some data. Google Docs is a similar concept: someone can open a file and other people can edit the same document in different locations in different countries.

In February 2000, Amazon.com suffered from DDOS attacks which caused serious damage [7]. A new technology was developed to defend against such attacks. Now, packets will be filtered, and it will be determined whether they are normal or not by the firewall before users communicate with the cloud. The private key of the cloud system is not stored in the user's equipment. The user must use a secure encryption method (such as Public Key Infrastructure (PKI) or Secure Socket Layer (SSL) to transmit messages to the cloud, and then the cloud's stored user identity verification table will identify the user.

In 2007, vendors pushed the OpenID [8] verification specification 2.0 and attributes of a standard 1.0. OpenID, aiming to provide different cloud providers with a means to authenticate users' identities. The users only need to register once with OpenID, and they can then log into the authentication pages. However, OpenID alliance should ensure the users' safety and be able to determine if a cloud is illegal or not, otherwise users' privacy will be easily revealed by a malicious attacker masquerading as a cloud service.

Cloud services have been a hot topic in recent years. Despite the lack of a concrete definition of a cloud, there seems to be a common consensus as to what constitutes a cloud [6]. The National Institute of Standards and Technology (NIST) [9] has proposed the following five basic characteristics of current cloud architecture:

(1) On-demand self-service: A consumer can unilaterally provide computing capabilities, such as server time and network storage, as needed automatically, without requiring human interaction with each service's provider.

(2) Broad network access: Capabilities are available over the network and accesses through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, laptops and PDAs).

(3) Resource pooling: The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control over or knowledge regarding the exact location of the provided resources, but may be able to specify location at a higher level of abstraction (e.g., country, state or data center). Examples of resources include: storage, processing, memory, network bandwidth, and virtual machines.

(4) Rapid elasticity: Capabilities can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out, and rapidly released to quickly scale in. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be purchased in any quantity at any time.

(5) Measured Service: Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled and reported, providing transparency for both the provider and consumer of the utilized service.

As long as a service is connected to a network and uses the network to achieve a certain goal, it can be called a cloud service (for example: E-mail, E-books, Google Docs, Google TV, Cloud Printer, etc). That is, users rely on the application data stored on a remote server, with no additional devices installed in the personal applications. Users' data and information can be stored anywhere in the cloud.

A prerequisite of connecting to the cloud is network connectivity. Users can send a message, use the service, receive messages etc; all rely on a network connection to communicate with the cloud. The user's information must be protected so there must be communication through a mechanism to protect the user's identity and information. Different suppliers have different protection mechanisms, for example, Google protection mechanism is used for SSL.

The remainder of this paper is organized as follows: Section 2 reviews the DRM related work. In section 3, we introduces the proposed protocol. Section 4, we analyze the security of the proposed scheme, and we provide conclusions in Section 5.

## 2. The DRM Related Work

E-books are currently the main product of cloud services. At present, the main providers are companies like Google, Apple, and Microsoft, although the E-book format has yet to be standardized. However, the main specifications of E-books are DRM, DRM-Free, and Adobe PDF format.

Take, for example, DRM-Free with permission [10]; on April 2, 2007, Apple announced that half of the DRM-protected music on iTunes would be sold via DRM-Free. The price would be lower for higher music quality. In this way, DRM-protected MP3 digital products it needed to pay for the license; there was the limitation that only Apple-related products could share this benefit. On the other hand, Google's DRM-Free forbids users to copy or print the digital content [11].
The primary business objectives of DRM are:

- Providers must specify the user's rights
- Digital content cannot be tampered
- The print and copy permissions of the digital content need to be authorized
- Digital content's Copyright notice

The Provider sells DRM-protected digital content that can be used to control the consumer's rights [12, 13, 14, 15]. However, the DRM cloud provider authorizes the users to access the digital content via a one-time sale.

Microsoft for digital content protection [16] must install the RMS software at the user end, and is limited to Windows OSes. The encryption method is the RSA [17] key component. When a user requests to authorize the use of a right, it allows the designated user and is authorized to grant the permission. However, the SP2 version added an offline authorization function, and authors use the RMS application to create file permissions; this specifies the authorization conditions. This is a special license which can be granted by the offline state RMS-protected content permissions.

Our proposed architecture allows consumers to download E-books and enjoy the benefits of the trial period (e.g., DRM-Free). When users buy the products, the users' permission will be changed via License (such as DRM). However, our architecture is such that, via the Internet, it is possible at any time to record a user's E-book page number. The advantages of our approach are that it allows users to read E-books on different devices, and it can be easily modified to record the number of pages. It can also prevent purchased E-book users illegally forwarding documents to other users.

### 2.1 Discussion of Using a Mobile Device to Implement DRM

With the rapid development of smart mobile devices, it is now possible to easily access network resources. Even though it is well-known that mobile devices are undermined by several recent threats [18], these mobile devices (such as PDAs or Tablet PCs) and cloud services can be combined to form an easy to use communication platform.

Users can access the cloud services through different mobile devices, however, the hardware of such mobile devices is limited in the following ways [7, 19]:

- bandwidth limitations
- connection stability
- low computational ability
- limited battery capacity
- small storage capacity

From the mobile user's viewpoint, the user must provide his/her identification before using the cloud service. This is different from using a smart card, since not every device can read smart cards. Moreover, different operating systems have different peripheral limitations (such as iPad). Although users can browse the web, the device does not provide a general standard interface (such as USB) to provide the smart card reading function or other more secure mechanisms (such as a biometric identification mechanism).

On the other hand, the mobile device's computing power is limited. In order to send a protected message from these mobile devices it is necessary to consider other appropriate security mechanisms. Google or Apple, and other providers of these cloud services, do not provide a clear definition for the services model of the cloud. In this paper, we present a mobile device-based DRM system to achieve the following objectives:

(1) Provide a process for clearer communication enabling a unified authentication.
(2) Reduce the computation of communication for mobile devices.
(3) The suppliers can use their encryption method to protect the security of E-books.
(4) The E-book providers for DRM purpose of sale and limitation are not the same.

In order to achieve the required level of security, we have integrated the mobile devices and the cloud services model to allow users to access an E-book resource under secure authentication.

## 3. Proposed Authentication Protocol

Because Linux is outstanding for parallel computing and executing efficiency [4], the proposed cloud service for mobile DRM systems is based on Linux. Linux is open source, so users can develop various APIs to meet their requirements.

The user's message will first pass through the firewall to confirm whether or not the packets are normal. For authenticating users, the cloud server aims to produce the session key between user and cloud. The cloud confirms the identity of the user's mobile device. The user need not worry about the messages sent to the cloud end or the internal processing. Our proposed architecture is shown in Figure 2:

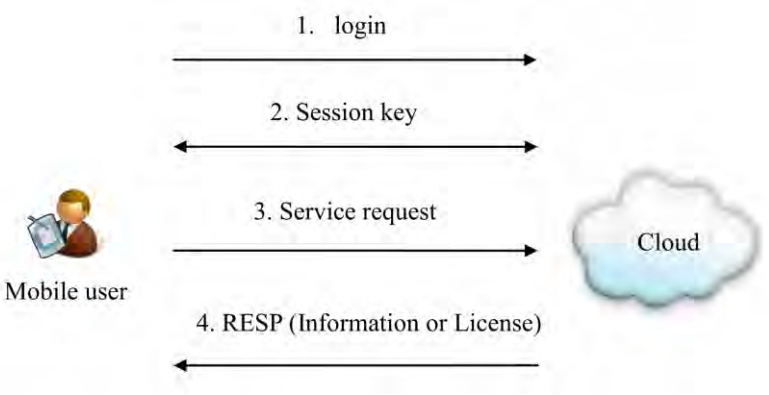



**Fig. 2.** Our proposed architecture

Step 1: User logs into cloud for authentication via mobile device.
Step 2: Cloud confirms the user's identification and generates the session key.
Step 3: The user's message is protected by the session key and a service request is made to the cloud.
Step 4: Cloud responds to user's request (such as the E-book pages or E-book usage rights).

Our scheme is to record the page number of the user's last review. We limit the user's communication time with the cloud to negotiate the session key by changing the license permission. The advantages are that users can read E-books on different devices, and we can prevent access to the E-books.

The following notation is used in this paper:

| | |
|---|---|
| $\oplus$ | exclusive -or operation |
| $\|\|$ : | concatenation operation |
| $ID$ : | user identification |
| $PW$ : | user password |
| $IMEI$ : | identity of the mobile device, International Mobile Equipment Identification |
| $N_u, N_s$ : | nonces |
| $SK$ : | session key between user and cloud |
| $E-book_{req}$ : | mobile user's first request of the E-book |
| $M_{req}$ : | E-book page number request after last view |
| $RESP$ : | response message of the cloud to user's request |
| $E_{SK}(m)$ : | use the symmetrical key $SK$ to encrypt a message $m$ |
| $D_{SK}(m)$ : | use the symmetrical key $SK$ to decrypt a message $m$ |
| $A \stackrel{?}{=} B$ : | determine whether or not A and B are equal |
| $h(.)$ : | one way hash function |

### 3.1 Registration Phase

The user proposes an identification *ID* and password *pw* to the cloud through a secure channel. The cloud stores the user's authentication information in the verification table.

### 3.2 Authentication Phase

The user authenticates with the cloud, and generates a session key. Figure 3 shows our proposed authentication protocol.

**Step 1**: User enters *ID* and *pw*, and generates a nonce $N_u$ and computes $C_1$ and $C_2$ as follows:

$$C_1 = h(h(pw) \oplus IMEI) \oplus N_u \tag{1}$$

$$C_2 = h(ID \| h(pw) \| h(N_u \oplus IMEI)) \tag{2}$$

Afterward, the user sends (*ID*, *IMEI*, $C_1$ and $C_2$) to the cloud.

**Step 2**: The cloud first checks *ID* and uses the *ID* to identify the corresponding *pw* on the verification table. Then the cloud computes $N'_u$ and performs the authentication as Eq. (4)

$$N'_u = C_1 \oplus h(h(pw') \oplus IMEI) \tag{3}$$

$$h(ID \| h(pw') \| h(N'_u \oplus IMEI)) \stackrel{?}{=} C_2 \tag{4}$$

If Eq. (4) holds, then the cloud completes the user's authentication. The cloud generates $N_s$ and computes the communication session key for the next communication as follows:

$$SK = h(N_u \| N_s) \oplus IMEI \tag{5}$$

Afterward, the cloud sends $h(h(pw) \| IMEI \| N_u) \oplus N_s$ and $h(IMEI \| N_u \| N_s)$ to the user.

**Step 3**: The user computes $N'_s$ and checks $N'_s$

$$N'_s = h(h(pw) \| IMEI \| N_u) \oplus h(h(pw) \| IMEI \| N_u) \oplus N_s \tag{6}$$

$$h(IMEI \| N_u \| N_s) \stackrel{?}{=} h(IMEI \| N_u \| N'_s) \tag{7}$$

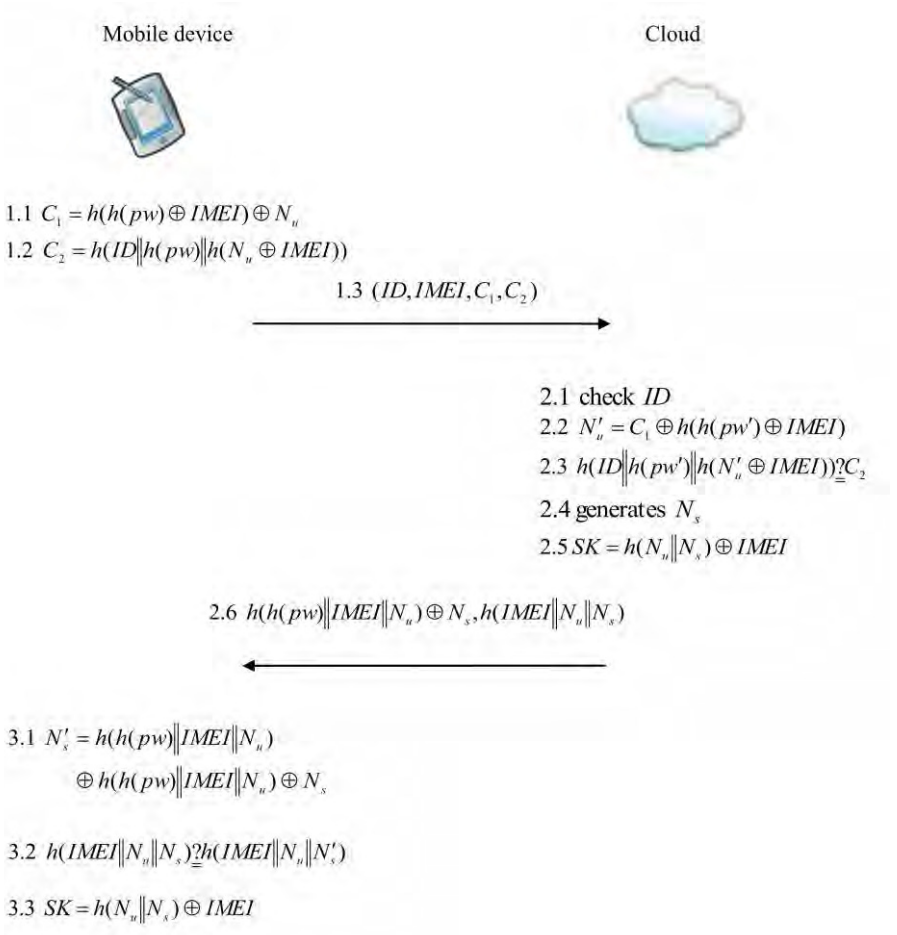



**Fig. 3.** The overview of our proposed authentication phase

If Eq. (7) holds, the cloud completes the mutual authentication with the user, and then the user can communicate the service message with the cloud. The user also generates a session key *SK* for the next communication.

$$SK = h(N_u \| N_s) \oplus IMEI \tag{8}$$

### 3.3 Service Response Phase

The user presents the service request by using the previous generated session key, and the cloud responds to the user's request. Figure 4 shows our proposed service response process.

**Step 1**: The user chooses the cloud service API license or asks to respond to the request; the cloud authenticates the user identity, generating a symmetric encryption message as follows:

$$C_3 = E_{SK}(E-book_{req}) \tag{9}$$

$$\text{or} \quad C_3' = E_{SK}(M_{req}) \tag{10}$$

A new nonce $N_{u+1}$ is generated and an authentication message is computed as follows:

$$C_4 = h(h(pw) \oplus IMEI) \oplus N_{u+1} \tag{11}$$

$$C_5 = h(ID \| h(pw) \| h(N_{u+1} \oplus IMEI)) \tag{12}$$

Afterward, the user sends $(ID, C_3, C_4, C_5)$ to the cloud.

**Step 2**: The cloud first checks *ID*, and uses the corresponding session key $SK$ to decrypt the service request.

$$E-book_{req} = D_{SK}(C_3) \tag{13}$$

$$\text{or} \quad M_{req} = D_{SK}(C_3') \tag{14}$$

The cloud computes $N'_{u+1}$

$$N'_{u+1} = C_4 \oplus h(h(pw') \oplus IMEI) \tag{15}$$

$$h(ID \| h(pw') \| h(N'_{u+1} \oplus IMEI)) \stackrel{?}{=} C_5 \tag{16}$$

If Eq. (16) holds, then the cloud generates the next nonce $N_{s+1}$, and calculates the new session key $SK_{new}$ as follows:

$$SK_{new} = h(N_{u+1} \| N_{s+1}) \oplus IMEI \tag{17}$$

User computes $C_6$ as follows:

$$C_6 = E_{SK_{new}}(RESP) \tag{18}$$

Afterward, the cloud sends $C_6'$, $h(h(pw) \| IMEI \| N_{u+1}) \oplus N_{s+1}$ and $h(IMEI \| N_{u+1} \| N_{s+1})$ to the user.

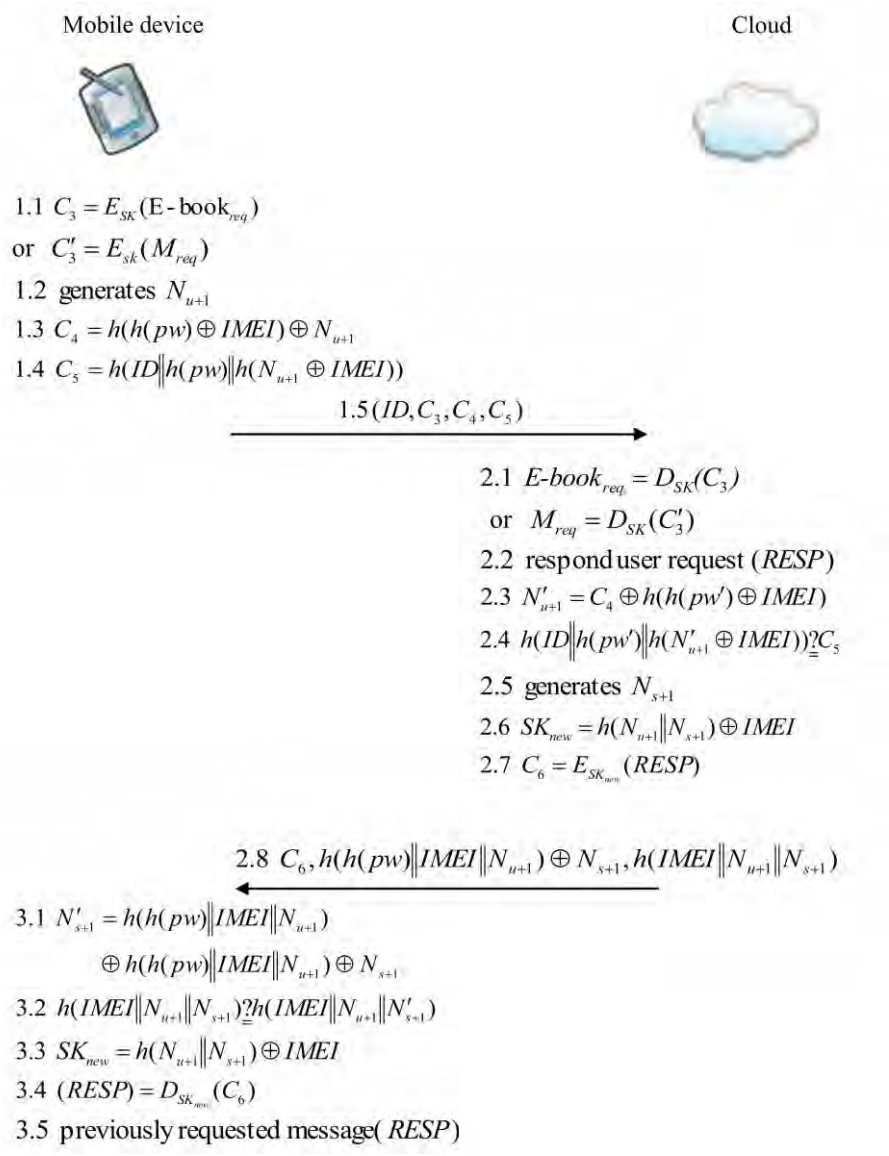



**Fig. 4.** The overview of our proposed service response phase

**Step 3**: The user computes $N'_{s+1}$ as follows:

$$N'_{s+1} = h(h(pw) \| IMEI \| N_{u+1}) \oplus h(h(pw) \| IMEI \| N_{u+1}) \oplus N_{s+1} \tag{19}$$

And authenticates the $N'_{s+1}$

$$h(IMEI \| N_{u+1} \| N_{s+1}) \stackrel{?}{=} h(IMEI \| N_{u+1} \| N'_{s+1}) \tag{20}$$

If Eq. (20) holds, then the user uses the previously generated $N_{u+1}$ to compute the $SK_{new}$ as follows:

$$SK_{new} = h(N_{u+1} \| N_{s+1}) \oplus IMEI \tag{21}$$

and decrypts $C_6$ to obtain the response message $RESP$

$$RESP = D_{SK_{new}}(C_6) \tag{22}$$

Thus, the user can access the previously requested message $RESP$.

In our proposed protocol, the cloud service user can continue to maintain a secure communication with the cloud.

## 4. Security Analysis

The following analysis is to show how our proposed scheme can prevent various attacks.

### 4.1 DOS Attack Prevention

As with the Amazon cloud infrastructure sites [5], the user's communication messages are the first through the firewall filters on the server. The user can synchronize with cookies to reduce abnormal malicious attacks, and users also must be limited to connect with the clouds. If users use the browser to perform malicious attacks, the server automatically locks the user's behavior. For example, if the same IP requests 1000 messages in one second, the user is regarded as a malicious attacker, and the server will block the IP services.

### 4.2 Password Guessing Attack Prevention

As the users do not store any user data in their mobile device, an attacker cannot achieve offline password guessing attacks via the mobile device. The cloud protects the users' accounts on the cloud end. If an attacker or a legitimate user enters consecutive incorrect passwords, the server will block the account, and the user will be requested to change the password and to send the registration information via email. Thus, there is no way to use online password guessing attacks since the attacker or the user does not know the previous password set.

### 4.3 Insider Attack Prevention

High value assets of the cloud system [5] request the user to change their password regularly, and the private key of the server will also be regularly changed. While the cloud stores the password, it does not directly store user passwords, and it is protected

by a one way hash function in order for users to store their passwords. For example, a user's password $pw$ and Linux's private key $x$ are protected by an MD5 hash function $h(h(pw) \oplus x))$. So, even if an insider attacker (root) steals the verification table, the attacker cannot use brute-force attacks to guess the user's password and identify the server's private key.

### 4.4 Reply Attack Prevention

Because the nonces $N_u$ and $N_s$ are not the same, even if the attacker were to intercept the messages ( $C_1 = h(h(pw) \oplus IMEI) \oplus N_u$ and $h(h(pw) \| IMEI \| N_u) \oplus N_s$ ), in order to make a forged message $C_4 = h(h(pw) \oplus IMEI) \oplus N_{u+1}$ and $h(h(pw) \| IMEI \| N_{u+1}) \oplus N_{s+1})$ , an attacker cannot use the intercepted messages to communicate with a user on the cloud during the authentication phase.

### 4.5 Impersonation Attack Prevention

Since each communication is recorded for a user's *ID* and *IMEI*, the user's password $pw$ is protected by a one way hash function ( $C_1 = h(h(pw) \oplus IMEI) \oplus N_u$ ), so the attacker cannot successfully fake being the user during the communication process. Neither can an attacker fake being the server. Moreover, the user's password is difficult to work out. Only the legal cloud can compute the correct $N'_{u+1} = C_4 \oplus h(h(pw') \oplus IMEI)$ , so the attacker cannot fake being the cloud.

### 4.6 Man-in-the-Middle Attack Prevention

Each message is protected by two unknown nonces $N_X$ and $h(pw)$ , so even if an attacker intercepts the messages $C_1 = h(h(pw) \oplus IMEI) \oplus N_u$ and $h(h(pw) \| IMEI \| N_u) \oplus N_s$ , the attacker cannot pass the authentication by the following equations: $h(ID \| h(pw') \| h(N'_{u+1} \oplus IMEI)) \stackrel{?}{=} C_2$ and $h(IMEI \| N_u \| N_s) \stackrel{?}{=} h(IMEI \| N_u \| N'_s)$ . Thus, the Man-in-the-Middle attack will be prevented.

### 4.7 Parallel Sessions Attack Prevention

The user transmits the communication messages $(C_1, C_2, C_4, C_5)$ to the cloud, and the cloud responds with the messages $h(h(pw) \| IMEI \| N_u) \oplus N_s$ and $h(IMEI \| N_u \| N_s)$. Both

of the communication messages of the hash value are different; thus, the proposed scheme prevents parallel session attacks.

### 4.8 Session Key Error or Tampering

Our protocol aims at reducing the computation cost on the mobile device. Once the session key is checked, if an error occurs or the key is tampered with during the authentication, the user just needs to be authenticated again and log into the cloud to access the cloud services.

### 4.9 Comparison

From Figure 5, it can be seen that we combine the charging mechanisms and replace usage rights with licenses in order to change the method of E-book usage rights via purchase. The proposed scheme enables the cloud to easily record a user's reading information, and the last viewed page immediately, despite interface and device limitations. Users can read E-books free from the various devices and paid software (such as office series) limitations anytime and anywhere. We use symmetric encryption for the device to reduce the computation and communication cost, which is different from other suppliers' encryption mechanisms.

| | DRM model | Interface model | Limited device or software | Protected mechanism | Install related API |
|---|---|---|---|---|---|
| Apple | B | Cloud | Limited to a single product brand equipment | N/A | iTunes |
| Google | A, B | Cloud | free | SSL | adobe reader |
| Microsoft | A, C | Client-Server | Office Series | RSA | RMS |
| Our scheme | A, B, C | Cloud | free | Symmetric encryption | adobe reader |

A：DRM, B：DRM-Free, C：Exchange License

**Fig. 5.** The comparisons of the related works

## 5. Conclusions

The proposed cloud scheme not only provides more convenient E-book services, but allows users to apply to other cloud services, with the digital content stored in the cloud. Users can access E-books using different devices, anytime and anywhere. The digital content is protected by DRM, which is flexible via changing the license usage mechanism such that the cloud can record the user's information.

Our proposed protocol allows users to use different mobile devices to access the cloud services. The mobile devices do not need to store the user's privacy and cloud's related messages. In the communication process, we use low complexity functions (such

as hash function, exclusive-OR and lightweight operations [20, 21]) to reduce the computing cost of the mobile device, and we also address mutual authentication issues. This study realizes the following goals:
(1) Propose a cross-vendor authentication of the cloud.
(2) Resist known attacks.
(3) Provide a low computing cost for mobile user.
(4) Provide a user friendly use for the digital content.
(5) Provide a device-independent management for DRM.

Considering the distributed nature of protected DRM contents and also that the proposed protocol allows to use different mobile devices, some possible future work could be to extend the work in a way to be also applicable to interconnected federated cloud, such as proposed in [22].

## References

1. Albano, P., Bruno, A., Carpentieri, B., Castiglione, A., Castiglione, A., Palmieri, F., Pizzolante, R. and You, I.: A Secure Distributed Video Surveillance System Based on Portable Devices, Lecture Notes in Computer Science, Vol. 7465, pp 403-415, (2012).
2. Pizzolante, R., Carpentieri, B. and Castiglione, A.: Text Compression and Encryption through Smart Devices for Mobile Communication, Proceeding of 2013 Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS 2013), July 3-rd to July 5-th, 2013, Asia University, Taichung, Taiwan, pp. 672 - 677.
3. Ohlman, B., Eriksson, A., Rembarz, R.: What Networking of information Can Do for Cloud Computing. the 18th IEEE International Workshops on Enabling Technologies : Infrastructures for Collaborative Enterprises, 78-83, (2009).
4. Buyya, R., Yeo, C.S., Venugopal, S., Broberg, J., Brandic, I.: Cloud Computing and Emerging IT Platfoms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility. Future Generation Computer Systems, Vol. 25, No. 6, 599-616, (2009).
5. Subashini, S., Kavitha, V.: A Survey on Security Issues in Service Delivery Models of Cloud Computing. Journal of Network and Computer Applications, Vol. 34, No. 1, 1-11, (2011).
6. Svantesson, D., Clarke, R.: Privacy and Consumer Risks in Cloud Computing. Computer Law & Security Review, Vol. 26, 391-397, (2010).
7. Chen, C.L.: A Secure and Traceable E-DRM System Based on Mobile Device. Expert Systems With Applications, Vol. 35, No. 3, 878-886, (2008)
8. OpenID. http://openid.net/government/, Access available 13/8/2013.
9. Mell, P., Grance, T.: The NIST Definition of Cloud Computing (Draft). http://csrc.nist.gov/publications/drafts/800-145/Draft-SP-800-145_cloud-definition.pdf, Access available 4/8/2011, (2011).
10. Apple DRM-Free. http://www.apple.com/pr/library/2007/04/02itunes.html, Access available 13/8/2013.
11. Google DRM-Free. http://books.google.com/support/partner/bin/answer.py?hl=en&answer=170424, Access available 13/8/2013.
12. Google DRM. http://books.google.com/help/ebooks/content.html, Access available 13/8/2013.
13. Google adopts Adobe ebook DRM. http://blogs.adobe.com/digitalpublishing/2010/12/google-ebooks.html, Access available 13/8/2013.

14. Chen, Y.Y., Wang, Y.J. and Chen, J.C.: A Fair-use DRM System Based on Web Service. Eighth International Conference on Intelligent Systems Design and Applications, Vol. 3, No. 11, 11-16, (2008).
15. Lee, W. B., Wu, W. J., Chang C. Y.: A Portable DRM Scheme Using Smart Cards. Journal of Organizational Computing and Electronic Commerce, Vol. 17, No. 3, 247-258, (2007).
16. Windows Rights Management Services http://technet.microsoft.com/zh-tw/library/cc706990(WS.10).aspx, Access available 13/8/2013.
17. Rivest, R., Shamir, A., Adleman, L.: A Method for Obtaining Digital Signatures and Public Key Cryptosystems. Communications of the ACM, Vol. 21, No. 2, 120-126, (1978).
18. Castiglione, A., De Prisco, R. and De Santis, A.: Do You Trust Your Phone?, Lecture Notes in Computer Science, Vol. 5692, pp 50-61, (2009).
19. Chen, C.L.: All-In-One Mobile DRM System Design. International Journal of Innovative Computing, Vol. 6, No. 3A, 897-911, (2010).
20. Chen, C. L., Tsai, Y. T.: Aniello Castiglione and Francesco Palmieri, Using Bivariate Polynomial to Design a Dynamic Key Management Scheme for Wireless Sensor Networks. Computer Science and Information Systems, Vol. 10, No. 2, 589-609, (2013).
21. Chen, C. L., Tsai, W. C.: Using a Stored-value Card to Provide an Added-value Service of Payment Protocol in VANET. 2013 Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS 2013), July 3-rd to July 5-th, 2013, Asia University, Taichung, Taiwan.
22. Esposito, C., Ficco, M., Palmieri, F. and Castiglione A.: Interconnecting Federated Clouds by Using Publish-Subscribe Service, Cluster Computing, In press, DOI 10.1007/s10586-013-0261-z, (2013).

**Chin-Ling Chen**, PhD, is a member of the Chinese Association for Information Security. From 1979 to 2005, he was a senior engineer at the Chunghwa Telecom Co., Ltd. He is currently a professor of the Department of Computer Science and Information Engineering at Chaoyang University of Technology, Taiwan. His research interests include cryptography, network security and electronic commerce. Dr. Chen had published over 50 SCI/SSCI articles on the above research fields in international journals.

**Woei-Jiunn Tsaur**, PhD, worked as a project manager and technology consultant from 1994 to 2003 in R&D Division of Syscom Computer Engineering Co., a research center of software development in Taiwan. Since 1999, he has been with the Department of Information Management at Da-Yeh University, Taiwan, where he is currently a full professor. His research interests include network security, security topics in operating systems, applied cryptography, information security management and computer networks. He has directed many research projects in the areas of network security and cloud computing security. Dr. Tsaur is also a member of the IEEE and the Chinese Cryptology and Information Security Association.

**Yu-Yi Chen**, PhD, is presently an associate professor of the Department of Management Information systems, National Chung Hsing University, Taiwan. His research interests include computer cryptography, network security, and e-commerce.

**Yao-Chang Chung** was born in 1987. He received the B.S degree in Department of Computer Science and Information Engineering from St. John's University, Taipei Taiwan in 2010. He received his Master degree at the Department of Computer Science and Information Engineering, Chaoyang University of Technology in 2012. His research interests include information security and cloud security.

*Received: September 19, 2013; Accepted: January 6, 2014.*

Computer Science and Information Systems 11(3):943–960 DOI: 10.2298/CSIS130917058S

# A NEMO-HWSN Solution to Support 6LoWPAN Network Mobility in Hospital Wireless Sensor Network

Mohammadreza Sahebi Shahamabadi[1], Borhanuddin M Ali[1], Nor Kamariah Noordin[1], Mohd Fadlee b. A. Rasid[1], Pooria Varahram[1], and Antonio J. Jara[2]

[1] Faculty of Engineering, Universiti Putra Malaysia,
43400 UPM Serdang, Malaysia
m.saheb.sh@gmail.com, {borhan, nknordin, fadlee, varahram}@upm.edu.my
[2] University of Applied Science Western Switzerland (HES-SO),
Switzerland
jara@ieee.org

**Abstract.** IPv6 Low-power Personal Area Networks (6LoWPANs) have recently found renewed interest because of the emergence of Internet of Things (IoT). Mobility support in 6LoWPANs for large-scale IP-based sensor technology in future IoT is still in its infancy. The hospital wireless network is one important 6LoWPAN application of the IoT, it keeps continuous monitoring of vital signs of moveing patients. Proper mobility management is needed to maintain connectivity between patient nodes and the hospital network. In this paper, first we survey IPv6 mobility protocols and propose a solution for a hospital architecture based on 6LoWPAN technology. Moreover, we discuss an important metric like signaling overload to optimize the power consumption and how it can be optimized through the mobility management. This metric is more effective on the mobile router as a coordinator in network mobility since a mobile router normally constitutes a bottleneck in such a system. Finally, we present our initial results on a reduction of the mobility signaling cost and the tunneling traffic on the mobile PAN.

**Keywords:** 6LoWPAN, NEMO, Handoff, Mobility, Wireless Sensor Networks, Healthcare.

## 1. Introduction

Over the past two decades, communication networks have experienced tremendous growth and expansion all over the world. The explosive growth of many types of mobile devices such as smart phones, variations of tablet computers, and laptops, has fueled the demand for more bandwidth with varying Quality of Service (QoS), with pervasive connectivity and at affordable costs [1]. These mobile devices are generally very powerful in themselves with ever more innovative user interfaces, better information security and privacy, capability for higher end-to-end data transfer rate, streaming or interactive communications, and many other features [2]. Mobile wireless network generally encompasses Wireless Sensor Networks (WSNs), ad-hoc and mesh networks and infrastructure based cellular networks. These groups of networks can service a wide

array of application areas such as the ubiquitous broadband access [3], mobile peer-to-peer, WiFi hot-spots, vehicular networks, sensor networks, and many more.

WSNs can be used for a wide range of applications, from environmental monitoring, home and industrial automation, military, to education, transport, healthcare and many more. It has been developed over IEEE 802.15.4 which is a layer_2 standard defined for Personal Area Network (PAN). WSN is designed for infrastructure-less type of networks which does not require an established network to be set up unlike the case with cellular based networks. WSN is also designed to connect to the Internet, this is done via a suitable node called the gateway [4]. However, IEEE 802.15.4 is defined to be of limited capabilities by way of smaller frame sizes, low memory capacity and data rate, respectively. It was primarily designed for short range communications with efficient power management. Eventually it creates a Low-power Personal Area Networks (LoWPANs) that supports a large number of nodes with energy saving capability [5]. The Internet Engineering Task Force (IETF) defines IPv6 Low-power Personal Area Network (6LoWPAN) which is an IPv6-based LoWPAN on the basis of IEEE 802.15.4 for communications with the Internet. With its vast address space, 6LoWPAN allows global connectivity between a large number of IPv6 intelligent devices over large areas. The protocol also enables the nodes to be self-organized i.e. can do self-detection, self-healing, and self-configuring, without human intervention [4].

For the success of IoT in general, and for healthcare in particular, mobility support is essential [6]. Mobility support is required to maintain fault tolerance of the network and full access to information regardless of their locations. In healthcare, some of the main applications for 6LoWPAN are for real-time monitoring of vital signs some examples being ECG (electrocardiogram), heart rate, SPo2, blood pressure, weight and breathing rate of patients. Moreover, it is important that these monitoring could be performed while the patients move around within the hospital [7]. In addition, because of the criticalness of healthcare provisioning mobility protocol needs to be reliable under any conditions, that is, it has to reduce packet loss, end-to-end delay, and network failures. Therefore, among the aims of a portable monitoring system are: firstly to control and monitor the patients in any location, and secondly to store the information as the Knowledge Based System (KBS) in order to study and survey symptoms and predict illness [8].

The design features of 6LoWPAN node like packet size restrictions, energy and power restrictions and delays in the reception of messages, have constrained host-based mobility protocols such as MIPv6 [9], HMIPv6 [10], FMIPv6 [9]. The Mobile Node (MN) which a mobile patient would carry, is involved in most of the mobility management signaling, and this weighs on the MN in the way of power consumption [11]. Hence, Proxy MIPv6 (PMIPv6) [12] is more appropriate in this respect to support 6LoWPAN mobility rather than the host-based solutions, but it has two shortcomings: that it cannot support multi-hop and that it requires 64 bit network prefix to be assigned to each MN [13].

Mobility solutions can give different kinds of efficiency and performance depending on the applications. In order to have a real-time access to the patients' body sensors to control body parameters, the use of Hospital Wireless Sensor Networks (HWSNs) is the best choice. Hence, this paper [14] grantees a reliable continuous and real-time remotely monitoring solution of hospitalized patients in a hospital infirmary based on an HWSN with intra-handover mechanism support. Thus, HWSN based on 6LoWPAN (HWSN6)

has been defined for hospitals as smart building, equipped with MNs, Border Routers (BRs) and gateways. Although this mobility solution has been tuned for hospital applications and therefore made more compatible with it, but the energy constraint of mobile patient nodes which comprises a set of sensor nodes as a PAN has not been considered [4]. It also did not consider network mobility especially on the aspect of energy consumption in Mobile Router (MR) this will constrain the PAN lifetime.

From this brief discourse, it is anticipated that 6LoWPAN will become more popular in the near future. This is primarily because it has a wide address space that is well suited to individually address all objects that are connected to the Internet. Nevertheless, power consumption is a serious issue in 6LoWPAN, hence, mechanisms need to be sought in order to optimize this resource. One example of a busy device is MR; it is a very complex device that manages significant mobility functions [1].

In this paper, we propose a new mobility solution for mobile networks such as mobile patient nodes that comprise of a set of sensor nodes that consitutes a single unit called mobile patient node in HWSN6 scenario [15]. In this scheme, the MR that acts as a coordinator manages the mobility and PAN functions. This mobility solution decreases the amount of message on MR, and prolongs the lifetime of a patient PAN via MR.

This paper is organized as follows: a review of the related works is presented in section 2. A discussion on system architecture is given in section 3. Section 4 presents the HWSN6 mobility scenario. In section 5, our mobility mechanism scheme is evaluated. Finally, simulation results and conclusion are discussed in section 6 and 7 respectively.

## 2. Related Works

From sensor networks point of view, movement occurs in 6LoWPAN nodes when an MN or a mobile PAN tries to leave its current link and connect to a new point of attachment. 6LoWPAN device/s should do self-configuration and self-detection and automatically introduce themselves in any movement to keep the connectivity. This process usually starts by binding message exchange through Neighbor Discovery (ND), and then establishing a bi-directional tunnel that connects the Home Agent (HA) and the MN. Mobility is categorized into two groups: micro-mobility or macro-mobility and involves two processes roaming and handover. Roaming is moving from the previous 6LoWPAN area to a new PAN and handover is the changing of current point of attachment and data flows to another point of attachment. Micro-mobility or ntra-PAN mobility occurs when an MN leaves its current position and moves to another point of attachment within the same 6LoWPAN network. On the other hand, macro-mobility or inter-PAN is the mobility between network domains where there would be a network address change [11]. Figure 1 displays the possible node mobility movement for supporting IPv6 in WSN 6LoWPAN. When the whole PAN changes its point of attachment similar to NEMO (NEtwork Mobility), this is called WPAN mobility [16].

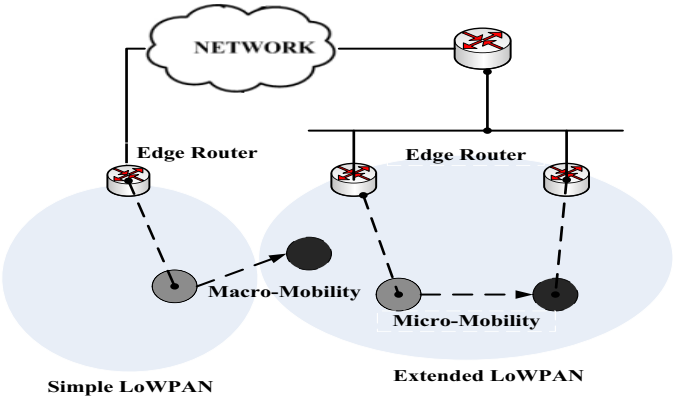



**Fig. 1.** 6LoWPAN Micro-Mobility and Macro-Mobility

The chart in figure 2 depicts the various mobility protocols and their hierarchies in MIPv6 when an MN changes its point of attachment in the network, it should update its current Care-of Address (CoA) by itself and informs the HA of its CoA using the Binding Update message (BU) [17]. An enhancement to the MIPv6, Hierarchical Mobile IPv6 (HMIPv6) was introduced, whereby it separates global mobility from local mobility [10]. Then, for the optimization of MIPv6, Fast handover for Mobile IPv6 (FMIPv6) was introduced. It reduces handoff delays by performing CoA configuration even before an MN leaves its current network [18]. In [19], they presented an authentication protocol for HMIPv6 roaming service to establish secure communications, when an MN is roaming into a foreign network. In the host-based mobility management protocols, an MN is involved in the processing of mobility and signaling to configure an IP address on a new link management [10]. FMIPv6, HMIPv6 and MIPv6 are of type host-based mobility protocol, but they are not suitable for 6LoWPAN due to its constraints [7].

From figure 2, network-based mobility is more appropriate in low-power sensor nodes because it relieves the MN from participating in any mobility operation, thereby extending its network lifetime [6]. In this respect, the Proxy Mobile IPv6 (PMIPv6) is more suitable as a mobility solution for IPv6 devices as it undertakes the responsibility of performing the handover process from the MN with a single hop. Even through this helps to conserve energy in IPv6 devices but single hop communication is not appropriate for 6LoWPAN devices because this may impose high transmission power to the energy constraint devices in order to reach distant PMIPv6 gateway [12]. Sensor Proxy Mobile IPv6 (SPMIPv6) is an optimization of PMIPv6 which is more suitable for energy constraint devices. It reduces signaling and mobility costs compared with MIPv6 and PMIPv6 [20]. LoWMob has been subsequently introduced for mobile 6LoWPAN nodes based on network-side and intra-mobility. The communication between MNs and gateways with the participant of the 6LoWPAN static nodes is made to be multi-hop rather than a single hop as in the previous protocols. The signaling overhead is reduced through supporting packet format at the adaptation layer [16]. A distributed version of LoWMob referred to as DLoWMob optimizes the mobility process. This is done by way of the following procedures: (i) supporting points to distribute the gateways traffic and

to enhance the multi-hop routing path between source and destination nodes, (ii) considering security aspects, (iii) equipping SNs with antennas in order to get the Angle of Arrival (AoA) measurements, and (iv) equipping SNs with a radio-triggered component to manage the sleep state by sending wake up radio signal [16]. Another protocol called Inter-MARIO has been proposed to perform handover based on the pre-configuration mechanism of 6LoWPAN mobility. This solution runs pre-configuration via the partner nodes to save the information on the PAN coordinators in the neighborhood PANs and reduces the mobility handover delay [21].

The philosophy behind NEMO protocol is that it runs Mobile IP and full IPv6 stacks only at MR/edge router, and does not run Mobile IP for attached nodes. This mobility solution fits the 6LoWPAN model perfectly as LoWPAN nodes are not adjustable for dealing with MIPv6 [20]. Lightweight NEMO protocol compresses the packet header to reduce the signaling overhead between MRs and gateways, this is done by using a compressed mobility header to support the 6LoWPAN mobility [11]**.** Inter-PAN mobility solution proposes an adaptation layer packet format for 6LoWPAN mobility signaling to reduce handover time. It provides extra information about the frequencies of the surrounding PANs at the border nodes [22]. To support mobility in 6LoWPAN sensor nodes, Sensor NEMO (SNEMO) has been introduced, it presents an interoperable architecture between NEMO and 6LoWPAN by way of an extended LOAD routing scheme for MRs [23]. Chai et al. [24] proposed a network architecture that supports the integration of NEMO and 6LoWPAN which shows that the handoff signaling of NEMO is 1/*N* times (N is the number of MNs) smaller than that of MIPv6, hence this means that the consumed energy of NEMO is much smaller than that of MIPv6. However, nodes that are selected as sensor routers consume more energy thus they suggested the use of non-power aware devices as sensor routers or MRs in NEMO.

HWSN6 defines a protocol to carry out intra-WSN mobility to support medical sensor networks based on 6LoWPAN. In this protocol, the mobility management is delegated to Monere system as BR which monitors a mobile patient's vital data [25]. This mobility scenario looks very similar to the NEMO protocol, in which the mobility of the entire network is viewed as a single unit.

The state of the art in HWSN6 related with high performance solutions includes security and authentication of MN for movements, global IPv6 addressing, intra-mobility among the Monere systems, reduced overload in MNs with respect to Mobile IPv6, distributed storage of the information among all the Monere systems, and mobility control messages to avoid fragmentation. Node authorization and authentication must be supported to offer security capability, integrity and confidentiality of the information, ensure protection of the resources.

In [26], they overview available handover mechanisms used for wireless sensors mobility and proposes a new ubiquitous mobility solutions for Body Sensor Networks (BSNs) in healthcare monitoring. This paper [27] surveys the most recent intra-mobility solutions with special focus on handover approaches that can be used in HWSNs. It proposed open issues that can contribute to improving the performance of handover solutions when applied to hospitalized patients were highlighted.

Although HWSN6 and previous solutions consider mobility issues but the energy consumption optimization of mobile patient node remains an open issue. The adaptation of the current mobility methods to 6LoWPAN remains a serious problem, and the further researches on 6LoWPAN mobility is necessary [7].

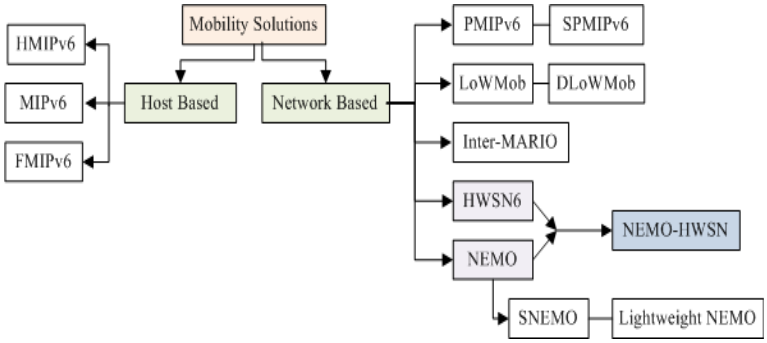



**Fig. 2.** Summary of IPv6 Mobility Solutions

## 3. System Architecture

The hospital system architecture is made up of patient nodes (MN with a set of sensors), Monere system (local gateway or BR), Internet gateway, Hospital Information System (HIS), and users (physicians, surgeons and nurses). As shown in figure 3, each part of the hospital such as operating theatre, observation rooms and wards are organised as a PAN which is under network coverage to keep the connectivity among the nodes and the Internet. Each PAN with all the nodes belong to the same domain deployed with a BR to connect to the Internet, HIS, and other PANs via the network backbone [4].

### 3.1. Gateway and HIS Node

A gateway manages its domain, establishes connections between networks, and interconnects with each other through wireless or wired links. HIS is a system based on Open Services Gateway Initiative (OSGi) technology for the management of all the other systems from the hospital. HIS saves the important monitoring information of all nodes and provides information and services to the other systems belonging to the hospital such as management of alarms from the Monere systems, electronic health record, health status, localization service, and directory service [4].

### 3.2. Monere System

Monere system [28] is a new BR device that has been suggested to cover each part (domain) of a hospital and also acts as a Mobile Data Collector (MDC) coming from the patient sensors, similar to a sink node in each PAN. It is equipped with several interfaces that establish connections with other networks technologies like Bluetooth, cellular networks, Ethernet and home automation (ZigBee, X10 and EIB) and standards such as CANBus, Ethernet and Serial Interface [29]. The area covered under the interconnected

BRs is referred to as a PAN or domain. 6LoWPAN BR plays two roles: it be identified as an HA responsible for buffering and forwarding packets to the MN, or as a Foreign Agent (FA) which coordinates visited network. Finally, it supports the security requirements like privacy and security that it can cipher the communications with AES-CBC cryptography (256bits key) [4].

### 3.3. Patient Node

This paper proposed the concept of mobile patient node which moves between multiple PANs in a hospital environment. A set of sensors acting as one unit fixed on the patient's body (6LoWPAN MN) measures and collects health data continuously such as heart rate, SpO2, peripheral and core body temperature, glucose etc [25]. From figure 4, two types of sensor nodes have been defined in IEEE 802.15.4: they are Full-Function Device (FFD) and Reduced-Function Device (RFD) respectively. FFDs are designed to support all network functionalities and participate in peer-to-peer topologies with multi-hop communications. On the other hand, RFD devices are limited mainly to perform measurements only of physical parameters and to processing non-complex tasks in star topologies since they do not support multi-hop communications. Normally each PAN coordinator controls a PAN, this is done by way of setting up and maintaining of the PAN. Hence, only a FFD device can assume the role of PAN coordinator [13].Two models are suggested for the patient mobile node: in the first model, there is a main FFD device with one IP address which collects data from a set of RFDs and also manages the patient node area as a coordinator. FFD acts as an MR and connects the BR to the patient through a 6LoWPAN node. All RFDs data are accessible from FFD, and this constitutes a bottleneck in the network. In the second model all 6LoWPAN sensors are considered as FFD devices with their own IPv6 addresses, they send their data directly to BR without any interface such as MR. Thus, it is clear that the second model is more expensive in terms of energy requirement and data exchange during mobility [25].

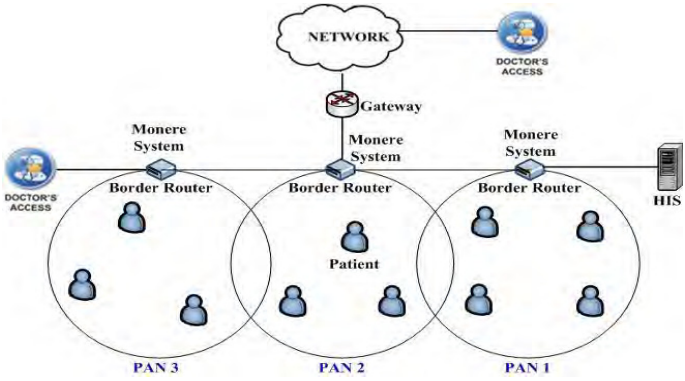



**Fig. 3.** Hospital Network Architecture

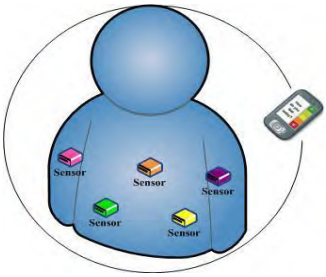



**Fig. 4**. Patient Node Sample Architecture and Topology

## 4. The HWSN6 Mobility Scenario

The WSNs mobility protocols proposed a large scope of applicability with the conjunction of the variety of case scenarios make it difficult to generate a standard mobility. To overcome this challenge, a specific scheme in mobility management for hospital WSNs has been proposed. The requirements of this scheme are continuous monitoring, low latency, no packet loss and low signaling. Figure 5 shows a movement scenario of a patient that moves between the home network and visited networks and then returns to base/home network. This kind of scenario is common at hospitals when the patients walk or move to other rooms to do medical tests. Phase 1 shows an initial state of the patient node which is in its home network and exchanges vital signs via the Monere system to maintain a continuous monitoring. In phase 2 and 3, it moves to a visited network and runs mobility protocol and handover mechanism, and finally it returns to the home network in phase 4.

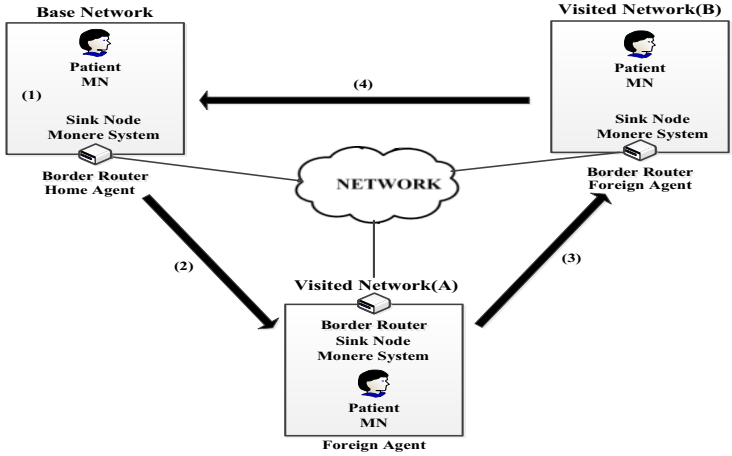



**Fig. 5.** Mobility Scenario of Mobile Patient Node

Figure 6 shows the HWSN6 mobility diagram with the messages exchanged in each step of mobility scenario as follows:

*Exchange of* messages *in home network:* The general frames (data, requests, responses and ACK frames) exchanged between sensors such as SPo2 level per each 5 seconds and BR.

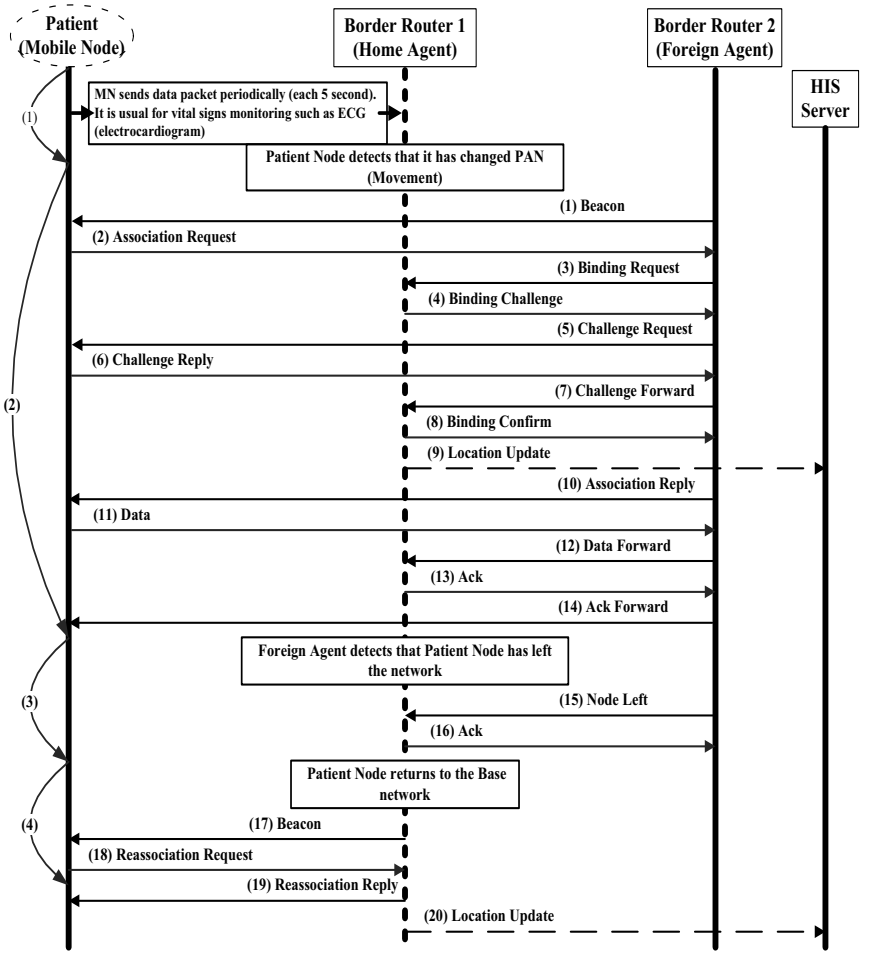



**Fig. 6.** Message Exchanges in a Mobility Scenario

*Movement detection time:* When an MN moves, it detects that its link quality has degraded beyond a certain threshold. This means that the existing router is no longer reachable, or a new access router is available [30].

*Entering the visited network:* Upon the mobile patient node entering the threshold or new network area (PAN), then it receives a Beacon message (message 1) which is

broadcasted periodically by 6LoWPAN BR acting as the coordinator (Monere system). Hence it detects the movement and sends Association Request (message 2).

*Confirmation of MN in visited network:* In order to authenticate the roaming MN, the following messages are exchanged: Binding Request (message 3), Binding Challenge (message 4), Challenge Request (message 5), Challenge Reply (message 6), Challenge Forward (message 7), Binding Confirm (message 8), Location Update (message 9), and Association Reply (message 10) message. These challenge messages are used to confirm that MN is a real node from its network. Patient node ciphers the challenge message and sends it to the FA. FA forwards to the HA. HA checks the challenge, if it is right, it sends a confirm message to the FA. In other case, it sends a deny message to avoid that the unauthenticated patient node receives or sends confidential information. Finally, the proposed mobility protocol supports security and authenticate MN with a challenge based on AES 128 bits when the MN changes its BR.

*Interchange of data frames in the visited network:* The messages from 11 to 14 show how a data frame and its Ack are exchanged.

*Returning to the base network:* Finally, as the patient node returns to its base network, it informs HA of its new location by sending a Re-association Request message (messages 17-20).

*Movement between visited networks:* When a patient node leaves the visited networks, FA informs the HA via Node Left and Ack messages (messages 15 and 16) of the event.

## 5. NEMO-HWSN Mobility Mechanism Scheme

As mentioned in section ȝ, the mobile patient node with its attached sensors is considered as a network or PAN that moves between different PANs like NEMO, because when the patient moves, all attached sensor nodes move together. Hence, it looks like the PAN or a group of mobile sensor nodes moves together and they also need a strong power device acting as an MR to coordinate and collect the PAN data. Hence, the partial of mobility cost have close relation to the PAN architecture such as type and number of sensors, message overhead, and the MR as a coordinator which manages the mobility in mobile PAN. The handoff and tunneling costs of patients in the mobility process depend on the number of attached sensors. As a result, the increased number of sensor node increases the complexity of fast handoff detection and decreases its efficiency, and finally increase the energy consumption. Hence, in the following methods, we will survey possible mobility scenarios to show the benefits of our proposed scheme. Figure 7 shows three mobility models that can be applied with mobile patient sensor nodes.

Figure 7 (a) presents the first model in which MR acts as a sink node, it controls, maintains PAN, collects data from body sensors and transmits to BR in the base network or visited network, and finally executes the mobility protocol. Although this model is similar to NEMO and has reduced handoff cost due to the use of MR that only it supports and runs the mobility process. However, the MR presents a bottleneck to the PAN because it should collect all data from attached sensors. This is a serious constraint in 6LoWPAN. As a result, the MR is made to work as a coordinator to handle the

mobility and collect data as a sink node. The benefit of this model is that it is less mobility complex and can perform fast handoff detection. The most serious problems are therefore bottleneck at the MR and end-to-end delay in tunneling process.

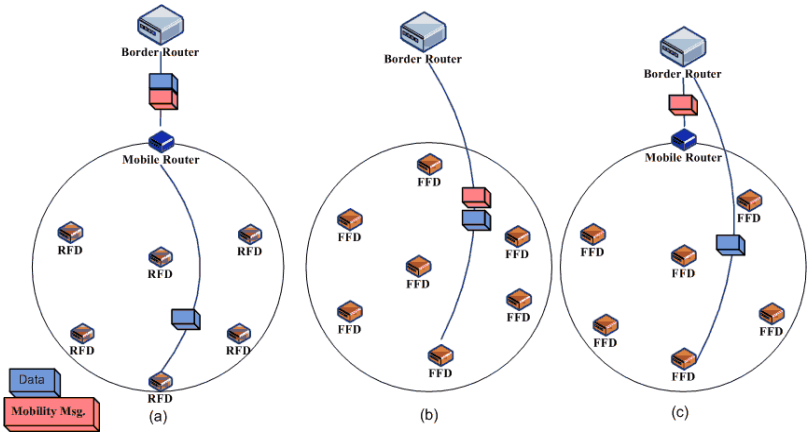



**Fig. 7.** The Messages Scheduling of Three Models. (a) RFD Devices with MR, (b) FFD Devices without MR, (c) FFD Devices with MR

Figure 7 (b) shows the second model in which all body sensors are FFD devices without any coordinator that attend to the mobility process. Accordingly, all FFDs repeat and execute the mobility scenario such as coordinator node (in previous model) and send their mobility messages to BR directly. This model is similar to individual mobile node that runs mobility scenario; it means the mobility protocol is supported with each individual node separately. The disadvantage with this model is that the handoff process will be increased based on the number of nodes, therefore the handoff complexity also will be increased [25]. With the benefit of this model is that each sensor node can leave its PAN and run mobility scenario separately and hence there is no bottleneck compared with the previous model. Finally, the MNNs send their data frames directly, thus the end-to-end delay in tunneling process will be optimized compared with previous method.

**Table 1.** Benefits of NEMO-HWSN Scheme

| Mobility Issues | NEMO | HWSN6 | NEMO-HWSN |
|---|---|---|---|
| End-to-End Delay | High | Low | Low |
| Bottleneck Node | MR Node | No | Optimized |
| Mobility Complexity | Low | High | Low |

NEMO-HWSN [15] is our mobility management solution which is designed to solve the serious challenges of previous mobility models to apply for group mobility in 6LoWPAN. We present a new scheme with low handoff cost like NEMO and light traffic on MR to optimize the PAN lifetime. Figure 7 (c) illustrates the proposed architecture which comprises of FFDs as sensor nodes with an MR as the coordinator. In

this model, the MR as a coordinator just runs the mobility process based on mobility diagram (Figure 6) to exchange the handover messages in movement situation; but data from sensors or MNNs are transmitted to BR directly. Consequently, the end-to-end delay in tunneling will be reduced due to remove one hop (MR node) in the direction of tunneling process. Hence, the duty of sensing data transmission is eliminated from MR, thus it leads to longer lifetime of MR during the tunneling process and sensor nodes can be located behind the MR without mobility message support. Finally, the MR registers all FFDs in the BR as an FA in order to create a connection with a new FA and transmit their data frames into networks. By way of this technique, we provide the best handoff cost and mobility scenario for MR. Hence, any increase in the number of FFD will not increase the cost of handoff during mobility. As a result, FFDs as members of patients' node send their data frames directly and the MR is set free of congestion at tunneling time. Thus, the bottleneck problem will be overcome by this scheme. Table 1 shows the previous challenges that are solved in NEMO-HWSN.

Figure 8, 9, and 10 show a comparison of the mobility diagram in terms of mobility and data messages scenario in three models. The dotted lines show handoff messages direction, when the MR or Mobile Network Nodes (MNNs) as mobile sensors run the mobility scenario which exchanges the handoff messages to follow the mobility process. The bold lines present the case when the tunneling scenario happens to exchange the sensing data from MNNs to destination like HA or CN. As shown in figure 10, the total signaling cost of our proposed scheme is better than the two previous mobility models. The NEMO-HWSN scheme that the MR mobility overhead is optimized by way of reduction in the MR traffic and the amount of mobility messages. As has been pointed out, the mobility cost is related to handoff and tunneling process time. Both of them have been surveyed by way of NEMO-HWSN solution through scheduling and managing the mobility functions of the MR.

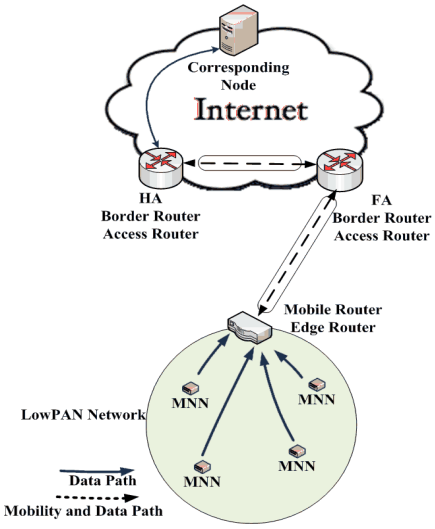



**Fig. 8.** Network Mobility Mechanism

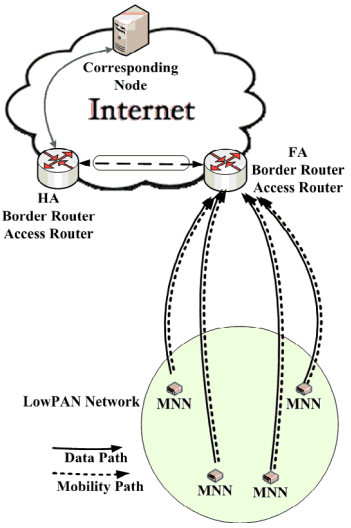



**Fig. 9.** Node Mobility for all MNNs

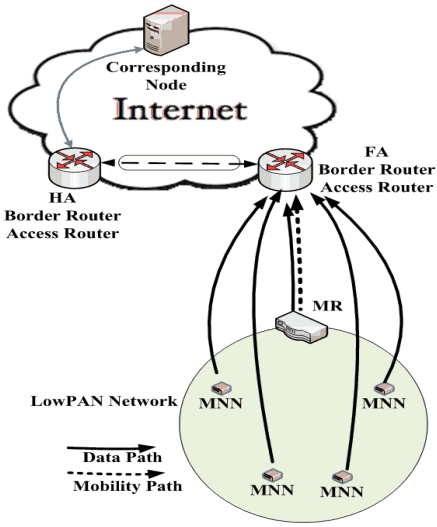



**Fig. 10.** NEMO-HWSN Mobility Mechanism

## 6. Simulation Results

To simulate our proposed scheme, we used OMNet++ simulator and the HWSN6 message diagram (Figure 6) which including binding update, challenging messages and etc. that exchange between MNNs, MR, HA, and FA in during the mobility scenario. In this scenario, the patient node consists of the five MNNs as mobile sensor nodes (attached sensors) to generate the sensing data and one MR node as a coordinator to manage the mobility mechanism.

The results from figure 11 shows the total mobility cost for tunneling and handoff process of the patient node with five attached sensor nodes (MNNs), i.e., the messages to exchange the data frames periodically from MNNs to HIS. It compares the total signaling cost of the NEMO-HWSN solution against that of the first model (NEMO) and the second model (node mobility) of the previous schemes. The graphs show that the total signaling cost in the NEMO-HWSN is very small in comparison to the second model (node mobility) and slightly smaller than NEMO protocol at minimum level. As mentioned before in figure 10, the proposed scheme (NEMO-HWSN) minimizes the handoff signaling of the MR; thus its total handoff cost is optimized as well as the NEMO protocol (first model). Consequently, the handoff signaling of NEMO-HWSN and NEMO are 1/N times (N is the number of MNNs) smaller than HWSN as node mobility. The graph shows the total signaling cost between NEMO-HWSN and NEMO is not very high, due to we exchange the low amount of data frames in tunneling process. In other words, the data frames start from MNNs are exchanged between MR and HA or HIS (as a CN in this scenario) without MR involvement in the tunneling direction.

Figure 12 compares the end-to-end delay between two network mobility models (HWSN6 and NEMO-HWSN). The end-to-end is optimized in our proposed scheme because it transfers MNNs data frames to HA without MR involvement in tunneling process. The NEMO-HWSN does not impose heavy traffic on the MR, and hence the bottleneck traffic is optimized. Therefore, the PAN lifetime is prolonged in mobility scenario process.

Finally, our proposed scheme reduces the mobility overhead of MR through reduction the tunneling messages to help extend the lifetime of PAN. This type of scenario is suitable for 6LoWPAN network mobility such as NEMO, which suffers from energy challenges such as energy constraint, limited battery or accessing to energy resources.

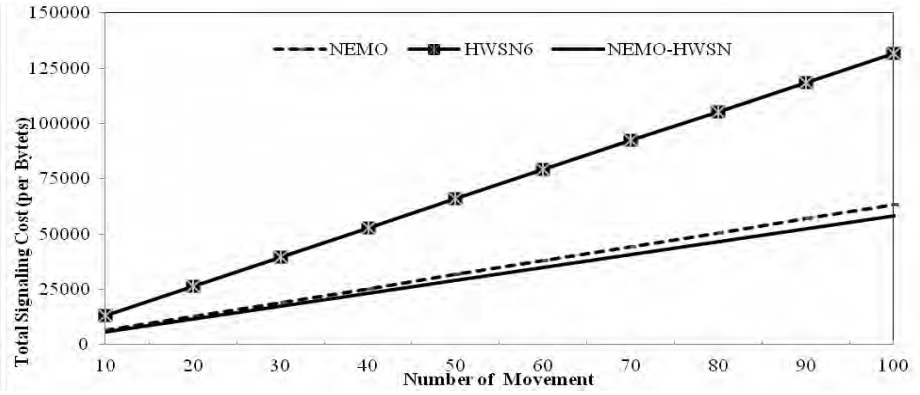



**Fig. 11.** Comparison of the Total Signaling Cost in Three Models

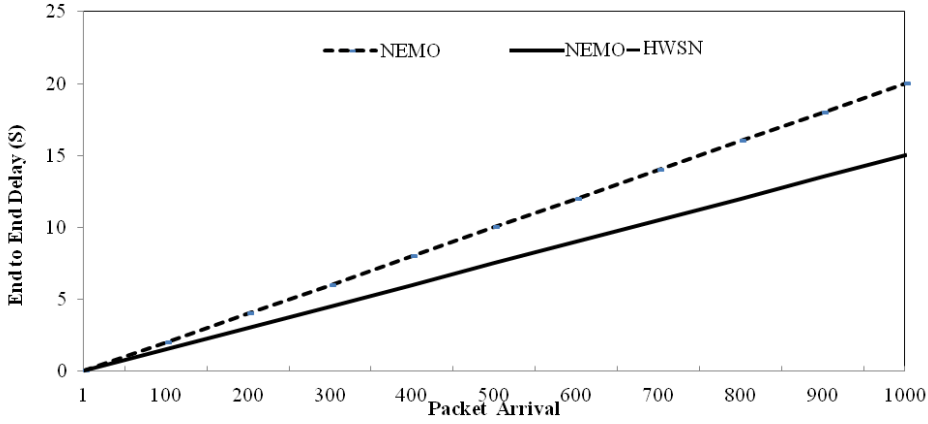



**Fig. 12.** Comparison of the End-to-End Delay in Two Group Mobility Models

## 7. Conclusion

This paper described a mobility solution for a group of 6LoWPAN mobile sensors like patient node with attached sensors in hospital settings to maintain the continuous connectivity between the patient nodes and hospital area network as a smart building. This solution considers the hospital architecture in order to define a solution that reduces the amount of messages exchanged between the mobile patient node and 6LoWPAN hospital network through the MR. This means that the signaling overload is decreased and also the lifetime of the MR is optimized due to the reduction in the total amount of mobility messages. The patient node should not run a costly configuration for new topology that causes the MR dies early due to congestion. Finally, it is shown that this scheme provides the low tunneling cost and light traffic on MR and BR regardless of the number of sensors attached to a patient node. Hence, the NEMO-HWSN mobility protocol for hospital architecture should be more feasible in a 6LoWPAN topology.

The article offers important insights for further studies on healthcare monitoring by using 6LoWPAN MNs as a part of IoT in movement. In the future, we will present the analytical model and real implementation to carry out a real test for performance evaluation in order to obtain the optimum handover solution along the mobility process.

**Acknowledgements.** The work is funded by the Ministry of Science, Technology and Innovation (MOSTI) of The GOVERNMENT OF MALAYSIA (No. 01-01-04-SF1218) and Universiti Putra Malaysia. The authors would like to thank all parties which have contributed towards the success of this project.

## References

1. Sahebi Shahamabadi, M., Mohd Ali, B. B., Varahram, P., Noura, M.: On Power Consumption in IPv6 Over Low Power Wireless Personal Area Network (6LoWPAN). In The First International Conference on Green Computing, Technology and Innovation. Kuala Lumpur, Malaysia, 21–28. (2013)
2. Dey, S., Shilpa, N.: Issues in IPv4 to IPv6 Migration. International Journal of Computer Applications in Engineering Sciences, Vol. I, No. I, 9–13. (2011)
3. Shelby, Z., Bormann, C.: 6LoWPAN: the wireless embedded internet. John Wiley & Sons, Ltd.United Kingdom. (2009)
4. Jara, A. J., Zamora, M. A., Skarmeta, A. F. G.: An Initial Approach to Support Mobility in Hospital Wireless Sensor Networks based on 6LoWPAN (HWSN6). Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, Vol. 1, No. 2/3, 107–122. (2010)
5. Huai, J. W., Culler, D.: Extending IP to Personal Area Networks. Internet Computing, IEEE, Vol. 12, No. 4, 37–45. (2008)
6. Jabir, A. J., Subramaniam, S. K., Ahmad, Z. Z., Hamid, N. A. W. A.: A cluster-based proxy mobile IPv6 for IP-WSNs. EURASIP Journal on Wireless Communications and Networking, Vol. 2012, No. 1, 1-17. (2012)
7. Wang, X., Zhong, S., Zhou, R.: A mobility support scheme for 6LoWPAN. Computer Communications, Vol. 35, No. 3, 392–404. (2012)
8. Jara, A. J., Blaya, F. J., Zamora, M. A., Skarmeta, A. F. G.: An Ontology and Rule Based Intelligent Information System to Detect and Predict Myocardial Diseases. In Proceedings of

the 9th International Conference on Information Technology and Applications in Biomedicine. Cyprus, 5–7. (2009)
9. Saha, D., Mukherjee, A., Misra, I. S., Chakraborty, M.: Mobility Support in IP: A Survey of Related Protocols. IEEE Network, Vol. 18, No. 34–40. (2004)
10. Soliman, H., Castelluccia, C., Bellier, L., Malki, K. El.: Hierarchical MIPv6 Mobility Management. In RFC 4140. (2005)
11. Kim, J. H., Hong, C. S., Shon, T.: A Lightweight NEMO Protocol to Support 6LoWPAN. ERTI Journal, Vol. 30, No. 5, 685–695. (2008)
12. Gundavelli, B. P. S., Leung, K., Devarapalli, V., Chowdhury K.: Proxy mobile IPv6. In IETF RFC 5213. (2008)
13. Oliveira, L. M. L., Sousa, A. F. D., Rodrigues, J. J. P. C.: Routing and mobility approaches in IPv6 over LoWPAN mesh networks. International Journal of Communication Systems, Vol. 24, No. 11, 1445–1466, (2011)
14. Caldeira, J. M. L. P., Rodrigues, J. J. P. C., Lorenz, P., Shu, L.: Intra-Mobility Handover Enhancement in Healthcare Wireless Sensor Networks. 14th International Conference on E-Health Networking, Applications and Services (IEEE Healthcom 2012), Beijing, China. (2012)
15. Sahebi Shahamabadi, M., Mohd Ali, B. B., Varahram, P., Jara, A. J.: A Network Mobility Solution Based on 6LoWPAN Hospital Wireless Sensor Network (NEMO-HWSN). The 7th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS-2013). Taiwan, 433–438. (2013)
16. Bag, G., Raza, M. T., Kim, K.-H., Yoo, S.-W.: LoWMob: Intra-PAN Mobility Support Schemes for 6LoWPAN. Sensors, Vol. 9, No. 7, 5844–77. (2009)
17. Perkins, C. E., Johnson, D. B.: Mobility support in IPv6. In Proceedings of the 2nd annual international conference on Mobile computing and networking (MobiCom '96). New York, USA, 27–37. (1996)
18. Camilo, T., Pinto, P., Rodrigues, A., Silva, J. S., Boavida, F.: Mobility management in IP-based Wireless Sensor Networks. In 2008 International Symposium on a World of Wireless, Mobile and Multimedia Networks. Newport Beach, CA, United States, 1–8. (2008)
19. Gao, T., Guo, N., Yim, K.: A Hybrid Approach to Secure Hierarchical Mobile IPv6 Networks. ComSIS Vol. 10, No. 2, 913-938. (2013)
20. Islam, M., Na, S., Lee, S., Huh, E.: A Novel Scheme for PMIPv6 Based Wireless Sensor Network. Lecture Notes in Computer Science, Vol. 6485, FGIT 2010, 429–438. (2010)
21. Ha, M., Kim, D., Kim, S. H., Hong, S.: Inter-MARIO: A Fast and Seamless Mobility Protocol to Support Inter-Pan Handover in 6LoWPAN. IEEE Global Telecommunications Conference GLOBECOM 2010. New York, USA,1–6. (2010)
22. Bag, G., Mukhtar, H., Shams, S. M. S., Kim, K. H., Yoo, S.: Inter-PAN Mobility Support for 6LoWPAN. In 2008 Third International Conference on Convergence and Hybrid Information Technology. Busan, South Korea, 787–792. (2008)
23. Kim, J. H., Hong, C. S., Okamura, K.: A Routing Scheme for Supporting Network Mobility of Sensor Network Based on 6LoWPAN. Lecture Notes in Computer Science, Vol. 4773, APNOMS 2007, Springer-Verlag Berlin Heidelberg New York, 155–164. (2007)
24. Chai, R., Zhao, Y.-L., Chen, Q.-B., Dong, T., Zhou, W.-G.: Group mobility in 6LoWPAN-based WSN. In 2010 International Conference on Wireless Communications & Signal Processing (WCSP). Suzhou, China,1–5. (2010)
25. Jara, A. J., Zamora, M. A., Skarmeta, A. F. G.: HWSN6: Hospital Wireless Sensor Networks Based on 6LoWPAN Technology: Mobility and Fault Tolerance Management. In International Conference on Computational Science and Engineering. Vancouver, BC, Canada, 879–884. (2009)
26. Caldeira, J. M. L. P., Rodrigues, J. J. P. C., Lorenz, P.: Towards Ubiquitous Mobility Solutions for Body Sensor Networks on HealthCare. IEEE Communications Magazine, Vol. 50, No. 5, 108-115. (2012)

27. Caldeira, J. M. L. P., Rodrigues, J. J. P. C., Lorenz, P.: Intra-Mobility Support Solutions for Healthcare Wireless Sensor Networks – Handover Issues. IEEE Sensors Journal, VOL. 13, NO. 11, 4339-4347, (2013)
28. Jara, A. J., Zamora-Izquierdo, M. A., Skarmeta, A. F.: Interconnection Framework for mHealth and Remote Monitoring Based on the Internet of Things. IEEE Journal in Selected Areas in Communications, Vol.31, No.9, 47-65. (2013)
29. Zamora-izquierdo, M. A., Gómez-skarmeta, A. F.: An Integral and Networked Home Automation Solution towards Indoor Ambient Intelligence. IEEE Pervasive Computing, Vol. 9, No. 4, 66–77. (2010)
30. Bag, G., Raza, M. T., Mukhtar, H., Akbar, A. H., Shams, S. M. S., Kim, K.-H., Yoo, S., Kim, D.: Energy-aware and bandwidth-efficient mobility architecture for 6LoWPAN. In MILCOM 2008 - 2008 IEEE Military Communications Conference. San Diego, CA, United States, 1–7. (2008)

**Mohammadreza Sahebi Shahamabadi** has received his bachelor degree in Computer Hardware Engineering from the Islamic Azad University, Maybod, Iran in 2001. Later, he received the Master degrees from the Islamic Azad University, Arak, Iran in 2008. He has started the Ph.D degree in wireless senseor networks in Universiti Putra Malaysia since 2010.

**Borhanuddin M Ali** received his PhD from the University of Wales (Cardiff) in 1985 and subsequently became a lecturer in Universiti Pertanian Malaysia (now renamed Universiti Putra Malaysia) since 1985. He served the university as a lecturer and various administrative positions -- Director of the Institute of Multimedia and Software; Director of Institute of Advanced Technology UPM (ITMA), and presently Director of the National Centre of Excellence in Sensor Technology (NEST), besides being the Head of Department twice. In 1996 he helped to realize the formation of Teman project a precursor to MYREN (a national research network), and later was made the Chairman of the MYREN Research Community till 2006. He is a Chartered Electrical Engineer and a member of IET and Senior Member of IEEE. He was Chair of IEEE Malaysia Section 2002-2004, and ComSoc Chapter twice, 1999-2002, and 2006-2008, and various portfolios in ComSoc AP Board culminating as Vice Director (2014-2016) and conference coordinator IEEE Region 10 in 2013. His research interest spans wireless communications in particular wireless sensor networks and broadband communications. He has authored and co authored some 100 journal papers and 200 conference papers under those areas of interests.

**Nor Kamariah Noordin** received her PhD at Universiti Putra Malaysia. She then became a lecturer in 1991 at the same department where she was later appointed as the Head from year 2000 to 2002. After that, she is assigned as a Deputy Dean (Academic, Student Affairs and Alumni) of the Faculty. Currently, she is the manager of Corporate Planning Division in Universiti Putra Malaysia. During her more than 15 years at the department, she has been actively involved in teaching, research, and administrative activities. She has supervised a number of undergraduate students as well as postgraduate students in the area of wireless communications, which led to receiving

some national and UPM research awards. Her research work also led her to publish more than 100 papers in journals and in conferences.

**Mohd Fadlee A. Rasid** is currently the deputy director for National Centre of Excellence for Sensor Technology (NEST) in Universiti Putra Malaysia (UPM). He received a Ph.D. in electronic and electrical engineering (mobile communications) from Loughborough University, U.K. He directs research activities within the Wireless Sensor Network (WSN) group and his work on wireless medical sensors is gaining importance in health care applications involving mobile telemedicine and has had worldwide publicity, including BBC news. He was a research consultant for a British Council UKIERI project on wireless medical sensors project. He was also part of the French Government STIC Asia Project on ICT-ADI: Toward a human-friendly assistive environment for Aging, Disability & Independence. He currently leads a few research projects on WSN, particularly for medical and agriculture applications. He is currently involved with a project under Qatar National Research Fund by Qatar Foundations on Ubiquitous Healthcare. He was nominated for IEE J A Lodge Award for Outstanding Work in Field of Medical Engineering, London, 2005 and was the proud recipient of State of Selangor Young Scientist Award 2006.

**Pooria Varahram** has obtained his bachelor degree in Electrical, Electronics Engineering from the Khaje Nasir University, Tehran, Iran in 2002. Later, he received the Master and Ph.D degrees in Wireless Communications from Tarbiat Modares, Tehran, Iran and Universiti Putra Malaysia, in 2005 and 2010. He is a member of IEEE since 2010. He has more than 5 years of experience in designing and developing a range of electronic and telecommunication related projects. He is currently senior lecturer in UPM.

**Antonio J. Jara**; Assistant Prof. PostDoc at University of Applied Sciences Western Switzerland (HES-SO) from Switzerland, vice-chair of the IEEE Communications Society Internet of Things Technical Committee, CTO and founder of the Wearable Computing and Personal Area Networks company HOP Ubiquitous S.L., CTO and co-founder of the Smart Cities company viBrain Solutions. He did his PhD (Cum Laude) at the Intelligent Systems and Telematics Research Group of the University of Murcia (UMU) from Spain. He was associated with the Department of Information and Communication Engineering, UMU, since 2007, where he has been working on several projects related to IPv6, WSNs. and RFID applications in building automation and healthcare. He is especially focused on the design and development of new protocols for security and mobility for Future Internet of things, which was the topic of his Ph.D. Nowadays, he continues working on IPv6 technologies for the Internet of Things in projects such as IoT6, and also Big Data and Knowledge Engineering for Smart Cities in collaboration with projects such as SmartSantander. He has also carried out a Master in Business Administration (MBA). He has published over 100 international papers, As well, he holds one patent. Finally, he participates in several Projects about the IPv6, Internet of Things, Smart Cities, and mobile healthcare.

*Received: September 17, 2013; Accepted: February 15, 2014.*

Computer Science and Information Systems 11(3):961–974 DOI: 10.2298/CSIS130926059M

# Long Distance Face Recognition for Enhanced Performance of Internet of Things Service Interface

Hae-Min Moon[1] and Sung Bum Pan[2]

[1] Dept. of Information and Communication Engineering, Chosun University,
375 Seoseok-dong, Dong-gu, Gwangju, Republic of Korea
bombilove@gmail.com
[2] Dept. of Electronics Engineering, Chosun University,
375 Seoseok-dong, Dong-gu, Gwangju, Republic of Korea
sbpan@chosun.ac.kr

**Abstract.** As many objects in the human ambient environment are intellectualized and networked, research on IoT technology have increased to improve the quality of human life. This paper suggests an LDA-based long distance face recognition algorithm to enhance the intelligent IoT interface. While the existing face recognition algorithm uses single distance image as training images, the proposed algorithm uses face images at distance extracted from 1m to 5m as training images. In the proposed LDA-based long distance face recognition algorithm, the bilinear interpolation is used to normalize the size of the face image and a Euclidean Distance measure is used for the similarity measure. As a result, the proposed face recognition algorithm is improved in its performance by 6.1% at short distance and 31.0% at long distance, so it is expected to be applicable for USN's robot and surveillance security systems.

**Keywords:** IoT, USN, surveillance, long distance face recognition.

## 1. Introduction

Until now, the Internet has been utilized as the optimal space by humans to share information as producers or consumers of information. In the future, not only information produced by humans but also everyday things will be connected to the Internet and will evolve so that the Internet of things can share the information of things via the Internet. Currently, industries, academics and governments from around the world are working on developing technologies and services for an intelligent network of things in various forms with Machine to Machine (M2M) or Internet of Things (IoT) [1]. Humans communicate with objects and services through IoT and objects and services communicate each other through IoT technology. As such, IoT interconnects human, objects and ambient environments including services and. It includes the traditional IoT services such as Smart Home/Security/Entertainment, Logistics/Distribution/Material Management/Security Management, Transportation/ Ambulance/Defense as well as various IT convergence services such as object recognition through location or motion and sensing information and situational awareness [2]. For example, when viewing from the human and service perspectives, a

mobile robot in a Ubiquitous Sensor Network (USN) or in a Smart Home environment recognizes family members and acts. Also the intelligent surveillance system continues to monitor the surveillance status and the acquired information may be provided to humans anytime and anywhere through network services. These technologies are implemented using actual IoT service interface and the IoT service interface plays the role of sensing, process/extracting/handling, storage, judgment, situational awareness, recognition, security, and human awareness of information [3], [4].

The purpose of this paper is to enhance the IoT service interface for uses human awareness by enhancing face recognition used in mobile robots and intelligent surveillance systems. Face recognition has a relatively lower recognition rate than fingerprint and iris recognition but because faces can be recognized from non-contact/non-cooperative environments and long distances, research studies on long distance human recognition using the face are currently underway [5], [6], [7]. Generally, the face recognition method is highly dependent on the quality of images obtained from the image sensor, so face recognition performance excels in short distance versus long distance. However, since the existing Linear Discriminant Analysis (LDA)-based face recognition technology works in short distance environment, if mobile robot or intelligent surveillance system of USN are applied literally to mobile robot or intelligent surveillance system, satisfactory service can't be expected. In order to provide seamless IoT service, it is necessary to have long distance face recognition technology that can recognize the target from various distances as well as at a short distance.

Therefore, this paper proposes a long distance face recognition algorithm that is applicable to mobile robots and intelligent surveillance systems. While the existing face recognition algorithm uses single distance image as training images, the proposed algorithm uses face images extracted from 1m to 5m as user training images. For face images at a distance of 1m to 5m, the size of face images extracted by distance is different, so it is normalized to the same size of face images using bilinear interpolation. In addition, Euclidean Distance measure is used for similarity measure. As a result, the face recognition rate of existing LDA-based face recognition was 85.8% in short distance and 44.0% in long distance, but the proposed face recognition method showed improved performance of 6.1% and 31.0%, respectively, for 91.9% in short distance and 75.0% in long distance. This paper is organized as follows. Section 2 introduces the concept of IoT service, face recognition technique and interpolation. Section 3 describes the proposed long distance face recognition algorithm and Section 4 analyzes the experiment results. Finally, Section 5 concludes the paper.

## 2. Background and Related Work

### 2.1. Concept of Internet of Things for Service Interface

By extending the traditional concept of the Internet, IoT is a next generation internet paradigm that encompasses networks of objects vs. object, and human vs. objects, which

various ambient objects are participating in the internet [8], [9], [10]. The definition of IoT can be generally divided into: Internet-based definition, Semantic- based definition and object-based definition. Firstly, the internet-based IoT definition is focused on network construction to therefore be able to connect with any objects, anywhere, and anyone such as the International Telecommunication Union (ITU) [11], [12]. Currently, the world is changing such that internet-based IoT is connecting a number of surrounding objects including mobile internet, Radio Frequency Identification (RFID), and sensor network and the objects are communicating with each other autonomously [13]. Secondly, the semantic-based definition is approaching IoT from the point of view of how to express, store, search and systemize many objects that will be included in IoT and the information which is produced from these objects [14], [15], [16]. Lastly, the RFID international standard organization, Global Standard 1 (GS1)/ Electronic Product Code (EPC) global defined IoT for the first time based on objects having the sole identifier-EPC. This made it possible to have object recognition and global location tracking by attaching a RFID tag with EPC to objects by reading these codes in real time through RFID readers installed all over the world and by storing and managing that information in IoT infrastructure distributed system [17]. Based on this, it is possible to: monitor and manage object information, which is part of IoT in real time and have various IoT services through a standardized interface. Recently, advances moving beyond simple identification studies are underway to provide various and intelligent IoT services through the development of an advanced interface including situation recognition and human recognition [3], [18], [19].

### 2.2. Algorithms of Face Recognition

Face recognition technology is examined in various studies ranging from still image-based face recognition in a controlled environment to video image-based face recognition from a crowded environment [20], [21], [22]. In this paper, we utilized LDA, which uses a feature extraction method using basis vector. In order to express two-dimensional face images, face shape and texture information are vectorized. For face shape information, physiographic features like the distance and ratio of face elements such as eye, nose and mouth are used. Texture information is expressed as brightness information itself in the face area. By arraying the brightness value of two dimensional face images in order, features are extracted by expressing one-dimensional vector. The feature extraction process in face recognition is to find the base vector for linear transition. LDA is to find the basis vector which reduces the scatter within the class and increases the distance between averages of each class [23], [24]. LDA use face images as a feature vector for face recognition by reflecting the face images to the basis vector.

Table 1 briefly shows the training process of the LDA technique. Table 2 briefly shows the recognition process of the LDA technique. In here, the most similar feature vector images are used as recognition result images by measuring the similarity of feature vectors between recognition images obtained and training images.

**Table 1.** Training process of LDA technique

**1.** Definition of $P$ number of training image vector

$$X = [x^1 \mid x^2 \mid ...x^P]$$

**2.** Definition of within-class scatter matrix of $i$-th

$$S_i = \sum_{x \in X_i} (x - mean_i)(x - mean)^T, \quad mean = \frac{1}{P_i} \sum_{x \in X_i}^{P} x$$

**3.** Definition of within-class scatter of matrix $S_w$

$$S_W = \sum_{i=1}^{C} S_i$$

,

**4.** Definition of between-class scatter of matrix $S_b$

$$S_B = \sum_{i=1}^{C} n_i (mean_i - mean)(mean_i - mean)^T, \quad mean = \frac{1}{P} \sum_{i=1}^{P} x^i$$

**5.** Definition of matrix that maximizes the ratio of $S_w$ and $S_b$

$$W_{opt} = \arg\max \frac{\left| W^T S_B W \right|}{\left| W^T S_W W \right|} = \left[ w_1, w_2, \cdots w_m \right]$$

$$S_B w_i = \lambda_i S_W w_i, \ i = 1, 2, \cdots m$$

- $C$ : number of classes, $n_i$ : number of images per class

**Table 2.** Recognition process of LDA technique

**1.** Definition of $P$ number of recognition image vector

$$Y = [y^1 \mid y^2 \mid ...y^P]$$

**2.** Difference of each image vector and average image vector

$$\overline{y^i} = y^i - mean, mean = \frac{1}{P} \sum_{i=1}^{P} y^i$$

**3.** Definition of feature vector for recognition image using $W_{opt}$

$$\widetilde{y^i} = W_{opt} \overline{y^i}$$

## 2.3. Interpolations for Image Normalization

For long distance face recognition, since the size of face images extracted according to the distance between camera and the subject is different, the size of face images to be verified should be normalized to fit to the size of training images. Therefore, interpolation is used to adjust the image size [25]. The nearest neighbor interpolation is the simplest method among interpolations and it refers to the pixel of nearest original images from the location that the output pixel is to be produced. Bilinear interpolation is a technique to produce the pixel to be interpolated using the adjacent four pixels. The interpolated pixel is determined by the sum of four pixels multiplied by a weighted value. At this time, weighted values are determined linearly and are inversely

proportional to the distance from each of the adjacent pixels. Figure 1 shows the bilinear interpolation using one-dimensional linear interpolation. To find the interpolated pixel *I*, bilinear interpolation is performed using the values of the adjacent four pixels (*A*, *B*, *C* and *D*). The bilinear interpolation provides a better image than nearest neighbor interpolation but it increases the computational complexity and the edge parts are not as smooth.

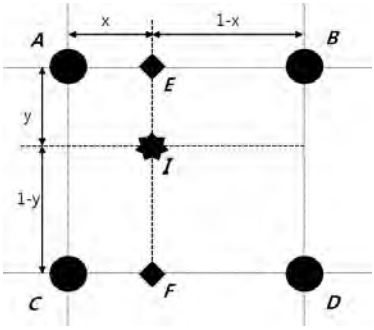
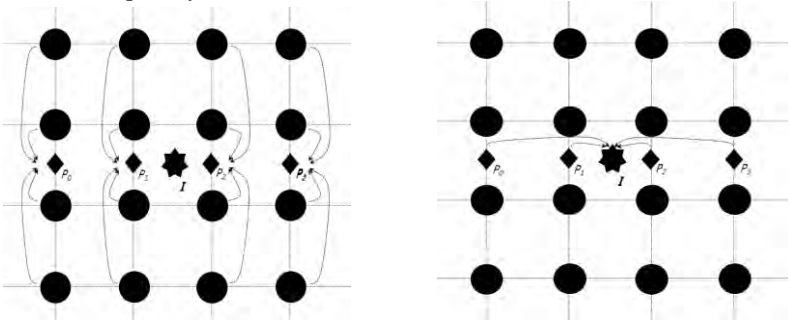



**Fig. 1.** Bilinear interpolation

Interpolation using a higher-order polynomial equation defines the function of weighted value and is a method to calculate the pixel values by adding all the values of neighboring pixel values of original images multiplied by weighted values. The representative method using a higher-order polynomial equation and includes cubic convolution interpolation [26]. Figure 2 shows the process of performing the two-dimensional cubic convolution interpolation using one-dimensional cubic convolution interpolation. Bicubic convolution interpolation produces new interpolated pixels using 16 pixels of original images. After four rounds of cubic convolution interpolation in the vertical direction as shown in Figure 2(a), four interpolated pixels ($P_0$, $P_1$, $P_2$ and $P_3$) are produced. Using the newly produced four interpolated pixels, when the cubic convolution interpolation is performed once horizontally, the final interpolated pixel I is produced as shown in Figure 2(b). Bicubic convolution interpolation refers to more pixels than bilinear interpolation so its image quality is good but it requires more computational complexity

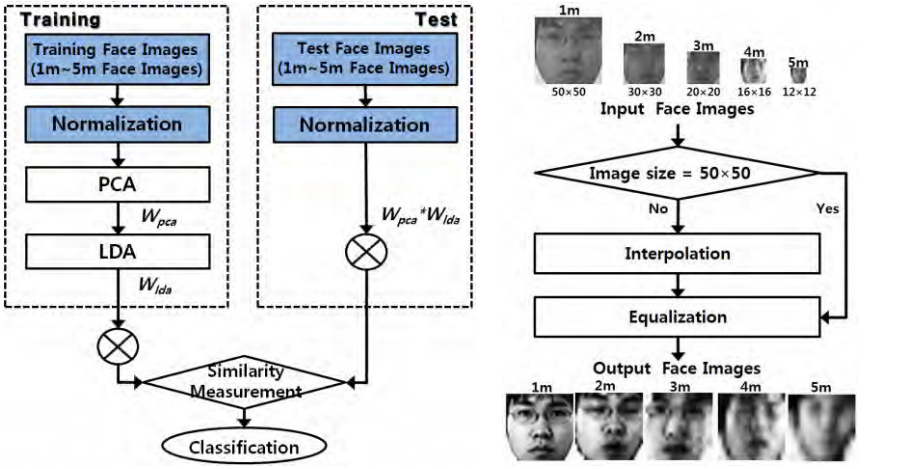



(a) Vertical interpolation (b) Horizontal interpolation

**Fig. 2.** Bicubic convolution interpolation

## 3. Proposed Long Distance Face Recognition System

Figure 3 is the flowchart of the proposed LDA-based long distance face recognition. Figure 3(a) shows the overall flow and Figure 3(b) presents the normalization process of face images being entered. The overall flow of the face recognition algorithm is the same as in existing face recognition algorithm. However, it has a difference in that the proposed algorithm uses face images at a distance of 1m to 5m as training images and adds a normalization process for face images based on distance.



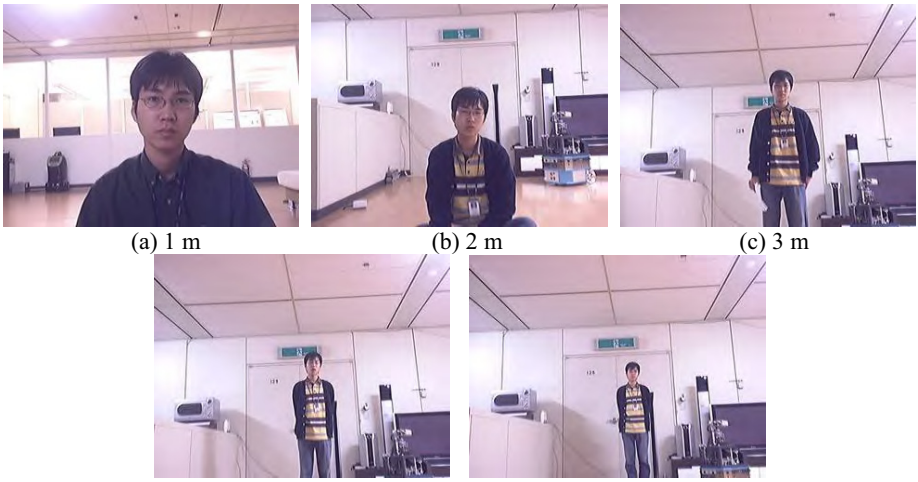

(a) The overall flow of proposed system (b) The normalization process of face images

**Fig. 3.** Long distance face recognition flowchart using LDA

The training process using face images from a distance is as follows. If face images at a distance of 1m to 5m are entered, the average face vector of the normalized face image is calculated through a normalization process. After calculating the difference of average face vectors in each face image, then find the covariance matrix. After finding the eigenvector and eigenvalue from the determined covariance matrix, finally $W_{pca}$ is generated. $W_{pca}$ generated through PCA is optimized by LDA again. To find $W_{lda}$ which is the data that the ratio of between-class scatter and within-class scatter in LDA is maximal. The test process of using face images from 1m to 5m distances is as follows. When the face image from 1m to 5m distances is entered, it is normalized through a normalization process. From the normalized face images, feature vectors are extracted through a difference of average face image vector and $W_{lda}$ projection. Finally, after comparing the feature vectors in the test area and the training areas, the face image that has the most similar value is classified.

The normalization of face image by distance is as follows. Once the face images for training are entered, the size of the input face images is judged. If the size of the image is 50×50, equalization will be conducted. However, if the size is smaller than 50×50 then equalization will be conducted after enlarging the size to 50×50 through interpolation. All face images entered through this process will be normalized into a 50×50 image size. Figure 4 is the original images at increasing distances and Figure 5 is the original face images extracted from person 1 according to the distance change of 1m

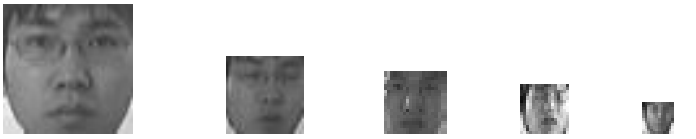
to 5m. The sizes of extracted face images are 50×50, 30×30, 20×20, 16×16 and 12×12 from 1m to 5m, respectively. The face images extracted by distance are normalized by four kinds of interpolation as shown in Figure 5.



(a) 1 m (b) 2 m (c) 3 m

(d) 4 m (e) 5 m

**Fig. 4.** Examples of original images at increasing distances



(a) 1m (b) 2m (c) 3m (d) 4m (e) 5m

**Fig. 5.** Examples of extracted face image at increasing distances

## 4. Experiment Result

Face recognition experiment uses ETRI face DB. As shown in Table 3, an ETRI face DB obtained 500 face images (1m to 5m: 100 images for each) per person from 10 different people in various lighting environments and at different distances [27]. Acquired face images were obtained using different distances ranging from 1m to 5m. In this paper based on the experimental images face images extracted from 1m to 2m were considered as short distance and face images extracted from 3m to 5m were considered as long distance. Face recognition, which is an 1:N search method rather than 1:1 authentication, classifies by comparing results for verified images of first face images which are the most similar among face images stored in the database. In addition, the experiment was carried out under the assumption that every face is extracted from input images regardless of distance. Additionally, a twisting or rotation of the face was not considered.

**Table 3.** ETRI face database

| |
|---|
| ● Total persons : 10 |
| ● Environment of obtained face images |
| - various lighting change |
| - 1m~5m distance change |
| - face position change |
| ● Face image size |
| - 1m : 50×50 2m : 30×30 3m : 20×20 4m : 16×16 5m : 12×12 |
| ● The number of total obtained face images : 5000 images |
| ● Obtained face images per a person : 500 images |

**Table 4.** Face recognition experiment according training images

| CASE | Training condition | Training time (sec) | Test time (sec) |
|---|---|---|---|
| 1 | Training image per person -1m : 20 images<br>Test image per person -1m~5m : 80 images each | 0.46 | 0.02 |
| 2 | Training image per person -1m~5m : 4 images each<br>Test image per person -1m~5m : 80 images each | 0.45 | 0.02 |
| 3 | Training image per person -1m~5m : 10 images each<br>Test image per person -1m~5m : 80 images each | 1.35 | 0.43 |

### 4.1. Face Recognition Rate Changes according to Interpolations

This experiment was carried out using Table 4 in order to find appropriate interpolations for the proposed algorithm. LDA was used as the face recognition method and Euclidean Distance was used for similarity measure. For normalization of the face image size by distance of 1m to 5m, the nearest neighbor, bilinear, bicubic convolution and Lanczos3 interpolations were used [28].

Figure 6 shows the results of LDA-based face recognition rate using normalized face images by distance through interpolation. In the experimental condition as shown in CASE 1 in Table 4 in order to get training images per person, 20 images of 1m face image were used and 80 images of face images at distances of 1m to 5m were used for verification images. As a result, when Lanczos3 interpolation was used for short distance the face recognition was 85.6%, which was the best performance. At long distance, when bicubic convolution and Lanczos3 were used, it showed similar performance at 44% and 44.1%, respectively. Figure 7 shows the results of LDA-based face recognition rate using normalized face images by distance through interpolation. For experiment condition as shown in CASE 2 in Table 4, a total of 20 face images for 1m to 5m distance by each 4 images were used to generate training images. As for test images, each of 80 face images at 1m to 5m distances was used. As a result, when Lanczos3 interpolation was used the face recognition was 92.9%, which showed the best face recognition performance. In long distance, the face recognition performance was excellent for 75.0% when bilinear interpolation was used. Figure 8 shows the results of LDA-based face recognition rate using normalized face images by distance through interpolation. For experiment condition as shown in CASE 3 in Table 4, a total of 50 face images for 1m to 5m distance by each 10 images were used for training images. As

for test images, each of 80 face images by 1m to 5m distances was used. As a result, when bilinear interpolation was used the face recognition was 93.8%, which showed the best face recognition performance. In long distance, the face recognition performance was excellent for 78.54% when nearest neighbor interpolation was used.

As a result, when the short distance face image is used as training, it is better to use Lanczos3 at the image normalization method in LDA-based face recognition. However, when using the face images at 1m to 5m distances as training, the face recognition performance was the best using the bilinear interpolation. When comparing CASE 1 and CASE 2 results, it was confirmed that it had better performance when using face images at 1m to 5 m distances than that of short distances. In addition, according to CASE 2 and CASE 3 results, as the number of training images per person increased, the recognition performance was improved. CASE 3 has better performance than CASE 2 but 50 pages of the training images per person are not used for general face recognition, so in this paper, the CASE 2 condition was considered.

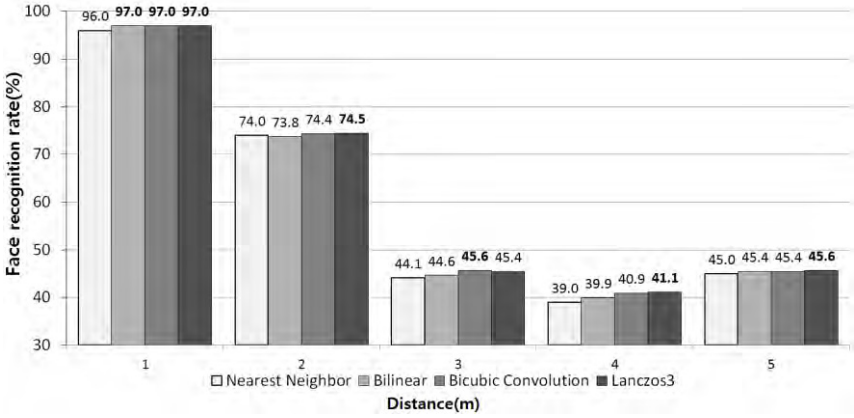



**Fig. 6.** Face recognition rate of CASE 1 according to interpolations

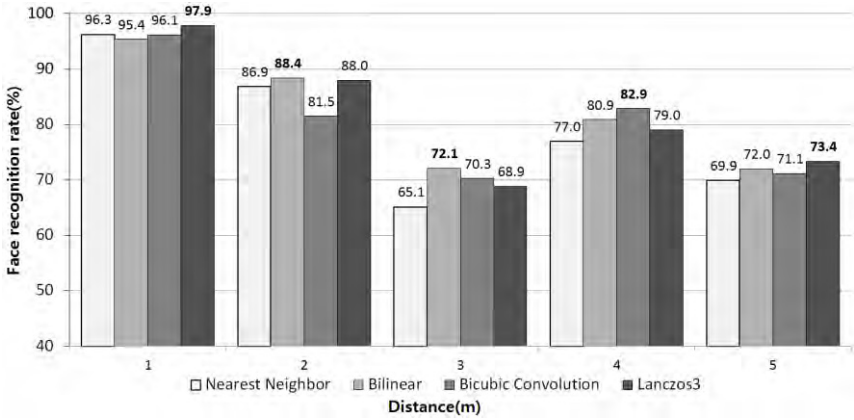



**Fig. 7.** Face recognition rate of CASE 2 according to interpolations

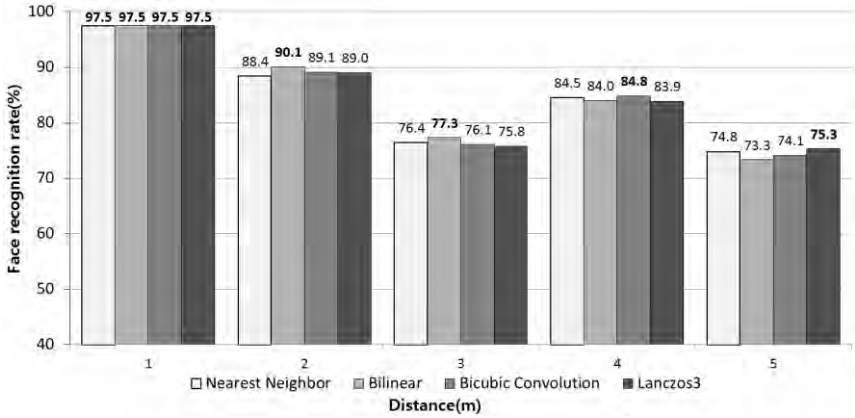



**Fig. 8.** Face recognition rate of CASE 3 according to interpolations

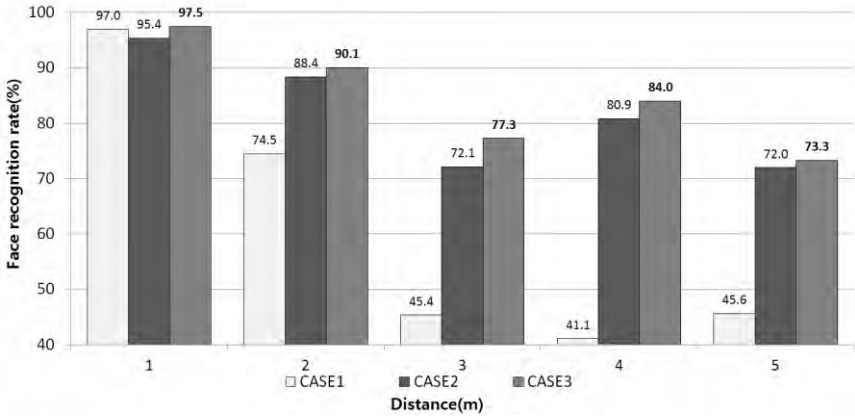



**Fig. 9.** Face recognition rate according to training images

## 4.2. Face Recognition Rate Change according to the Configuration of Training Images

Through this experiment, the effect on the face recognition rate of the configuration of the training image and the excellence of LDA-based face recognition when face images that are at a distance were used as training images are proved. Figure 9 shows the results of the configuration of training image effect on face recognition in LCA-based face recognition. In CASE 1, Lanczos3 interpolation was used and in CASE 2 and CASE 3 bilinear interpolation was used for normalization of the face image size. L2 was used for the similarity measure. As a result, when using a single distance for training images of CASE 2, the performance was 85.8% at short distances and 44.0% at long distances. When using face images at distances of 1m to 5m, the short distance had better performance for 91.9% than when using single distance for training images, which was

75.0%. Consequentially, when the same number of training images was used, the face recognition rate was improved if multi-distance face images were used rather than single distance face images.

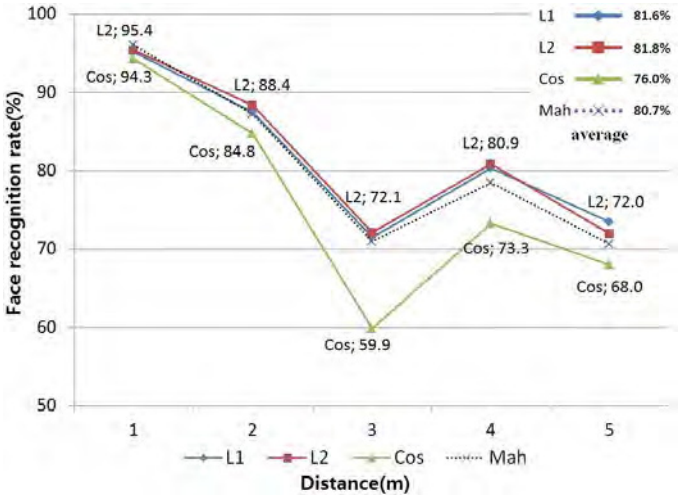



**Fig. 10.** LDA-based face recognition rate according to similarity measure

### 4.3. Face Recognition Rate Change according to Similarity Measure

Through this experiment, when the face images at 1m to 5m distances were used, the similarity measure that is appropriate to long distance face recognition is proposed. The configuration of training images were like in CASE 2 and LDA was used for the face recognizer. Bilinear interpolation was used as image normalization method. For similarity measure, Manhattan Distance (L1), Euclidean Distance (L2), Cosine Similarity (Cos), and Mahalanobis Distance (Mah) distance scale method were used [29]. Figure 10 shows the face recognition rate of LDA-based face recognition according to similarity measure. As a result, L2 was used for short distance and it showed the best performance at 91.9%. In long distance, L1 and L2 showed similar performance at 75.1% and 75%, respectively. The overall average face recognition rate of 1m to 5m were 81.6%, 81.8%, 76.0% and 80.7% respectively when using L1, L2, Cos and Mah and the recognition rate of L2 was the best. Consequently, in LDA-based long distance face recognition when multi-distance images were used as training, the face recognition performance was the best when the Euclidean Distance (L2) similarity measure was used.

## 5. Conclusion

The world considers IoT as a means of securing national competitiveness and develops efficient interface based on technology development and IoT dissemination by policy.

This paper proposes a long distance face recognition algorithm, which is applicable as an USN or MSM service-based technology. Face recognition, which has used the existing single distance face images as training images, has the disadvantage of lower recognition rate as the distance between the surveillance camera and the user increases. In this paper, an LDA-based long distance face recognition algorithm appropriate to the environment of the surveillance camera is proposed. Face images at distance were used in a proposed face recognition algorithm and the low resolution images at distance were normalized using a bilinear interpolation. For the similarity measure, Euclidean Distance measure method was used. A major result of this experiment showed that the proposed face recognition algorithm had improved face recognition rate for 6.1% in short distance and for 31.0% in long distance compared to the LDA-based face recognition using existing short distance face images.

In future, the proposed algorithm will be developed into a structure that is able to use minimization and low power processing of the proposed algorithm suitable for an object communication service environment. Additionally, technologies that can protect personal information effectively used in human recognition received on a mobile robot or terminal device will be developed.

**Acknowledgments.** This research was supported by a Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0023147)

## References

1. Strategy, I.T.U., Unit, P.: ITU Internet Reports 2005: The Internet of Things, International Telecommunication Union, Geneva. (2005)
2. Gonzalez-Miranda, S., Alcarria, R., Robles, T., Morales, A., Gonzalez, I., Montcada, E: An IoT-leveraged Information System for Future Shopping Environments. IT Convergence Practice, Vol. 1, No. 3, 49–65. (2013)
3. Sundmaeker, H., Guillemin, P., Friess, P., Woelfflé, S.: Vision and Challenges for Realising the Internet of Things. CERP-IoT–Cluster of European Research Projects on the Internet of Things, 1–230. (2010)
4. Perera, C., Zaslavsky, A., Christen, P., Georgakopoulos, D.: Context Aware Computing for the Internet of Things: a survey. IEEE Communications Surveys & Tutorials, Vol. 16, No. 1, 414–455. (2014)
5. Moon, H.M., Pan, S.B.: A New Human Identification Method for Intelligent Video Surveillance System. In Proceedings of 19th International Conference on Computer Communication and Networks, Zurich, Switzerland, 1–6. (2010)
6. Tsai, H.C., Wang, W.C., Wang, J.C., Wang, J.F.: Long Distance Person Identification using Height Measurement and Face Recognition. In Proceedings of 2009 IEEE Region 10 Conference, Singapore, 1–4. (2009)
7. Yao, Y., Abidi, B., Kalka, N.D., Schmid, N., Abidi, M.: High Magnification and Long Distance Face Recognition: Database Acquisition, Evaluation, and Enhancement. In Proceeding of 2006 Biometrics Symposium: Special Session on Research at the Biometric Consortium Conference, Baltimore, Maryland, 1–6. (2006)
8. Atzori, L., Iera, A., Morabito, G.: The Internet of Things: A Survey. Computer Networks, Vol. 54, No. 15, 2787–2805. (2010)

9. Gershenfeld, N., Krikorian, R., Cohen, D.: The Internet of Things. Scientific American, Vol. 291. No. 4, 76–81. (2004)
10. Ashton, K.: That 'Internet of things' Thing, RFID Journal, Vol. 22, 1–6. (2009)
11. ITU.: http://www.itu.int/en/ITU-T/techwatch/Pages/internetofthings.aspx
12. IPSO.: http://ipso-alliance.org/about
13. Weinstein, R.: RFID: A Technical Overview and Its Application to the Enterprise. IT Professional, Vol. 7, No. 3, 27–33. (2005)
14. Toma, I., Simperl, E., Hench, G.: A Joint Roadmap for Semantic Technologies and the Internet of Things. In Proceeding of 3rd STI Roadmapping Workshop, Helsinki, Greece. (2009)
15. Katasonov, A., Kaykova, O., Khriyenko, O., Nikitin, S., Terziyan, V.: Smart Semantic Middleware for the Internet of Things. In Proceedings of the Fifth International Conference on Informatics in Control, Automation and Robotics, Robotics and Automation, Madeira, Portugal. (2008)
16. Song, Z., Cárdenas, A.A., Masuoka, R.: Semantic Middleware for The Internet of Things. Internet of Things, 1–8. (2010)
17. San Jose, J.I., de Dios, J.J., Zangroniz, R., Pastor, J.M.,: WebServices Integration on An RFID-based Tracking System for Urban Transportation Monitoring. IT Convergence Practice, Vol. 1, No. 4, 1–23. (2013)
18. Miorandi, D., Sicari, S., Pellegrini, F.D., Chlamtachm, I.: Internet of Things: Vision, Applications and Research Challenges. Ad Hoc Networks, Vol. 10, No. 7, 1497–1516. (2012)
19. Domingo, M.C.: An Overview of the Internet of Things for People with Disabilities. Journal of Network and Computer Applications, Vol. 35, No. 2, 584–596. (2012)
20. Wiskott, L., Fellous, J.M., Kr¨uger, N., von der Malsburg, C.: Face Recognition by Elastic Bunch Graph Matching. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, 775–779. (1997)
21. Chellappa, R., Wilson, C.L., Sirohey, S.: Human and Machine Recognition of Faces: A Survey. Proceedings of IEEE, Vol. 83, No. 5, 705–741. (1995)
22. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face Recognition: a Literature Survey. ACM Computing Surveys, Vol. 35, 399–458. (2003)
23. Turk, M., Pentland, A.: Eigenfaces for Recognition. Journal of Cognitive Neuroscience, Vol. 3, No. 1, 71–86. (1991)
24. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces vs. Fisherfaces: Recognition using Class Specific Linear Projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, 771–720. (1999)
25. Parker, J.A. , Kenyon, R.V., Troxel, D.E.: Comparison of Interpolating Methods for Image Resampling. IEEE Transactions on Medical Imaging, Vol. 2, No. 1, 31–39. (1983)
26. Keys, R.G.: Cubic Convolution Interpolation for Digital Image Processing. IEEE Transactions on Acoustic, Speech, and Signal Processing, Vol. asp-29, No. 6, 1153–1160. (1981)
27. Kim, D.H., Lee, J.Y., Yoon, H.S., Cha, E.Y., A Non-Cooperative User Authentication System in Robot Environments. IEEE Transactions on Consumer Electronics, Vol. 53, No.2, 804–810. (2007)
28. Duchon, C.E.: Lanczos Filtering in One and Two Dimensions. Journal of Applied Meteorology, Vol. 18, No. 8, 1076–1022. (1979)
29. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, John Wiley & Sons, USA. (2004)

**Hae-Min Moon** received the B.S. degree in Control, Instrumentation, and Robot Engineering in 2009 from Chosun University, Gwangju, Korea. He received the M.S. degrees in Information and Communication Engineering in 2010 from Chosun University, Gwangju, Korea. He is currently working toward the Ph.D. degree. His research interests include image interpolation, video surveillance, and video compression.

**Sung Bum Pan** is the corresponding author of this paper. He received the B.S., M.S., and Ph.D. degrees in Electronics Engineering from Sogang University, Korea, in 1991, 1995, and 1999, respectively. He was a team leader at Biometric Technology Research Team of ETRI from 1999 to 2005. He is now professor at Chosun University. His current research interests are biometrics, security, and VLSI architectures for real-time image processing.

*Received: September 26, 2013; Accepted: January 6, 2014.*

Computer Science and Information Systems 11(3):975–999 DOI: 10.2298/CSIS130917060G

# PPS: A Privacy-Preserving Security Scheme for Multi-operator Wireless Mesh Networks with Enhanced User Experience

Tianhan Gao[1], Nan Guo[2], Kangbin Yim[3], and Qianyi Wang[4]

[1] Faculty of Software College, Northeastern University,
110819 Shenyang, China
gaoth@mail.neu.edu.cn
[2] Faculty of Information Science and Engineering College, Northeastern University,
110819 Shenyang, China
guonan@ise.neu.edu.cn
[3] Faculty of Information Security Engineering, Soonchunhyang University,
336745 Asan, Korea
yim@sch.ac.kr
[4] Faculty of Economics and Administration, University of Malaya,
50603 Kuala Lumpur, Malaysia
qianyiyouyou@sina.com

**Abstract.** Multi-operator wireless mesh networks (WMNs) have attracted increasingly attentions as a low-cost accessing approach for future large-scale mobile network. Security and privacy are two important objectives during the deployment of multi-operator WMNs. Despite the necessity, limited literature research takes both privacy and user experience into account. This motivates us to develop PPS, a novel privacy-preserving security scheme, for multi-operator WMNs. On one hand, most of the privacy needs are satisfied with the hybrid utilization of a tri-lateral pseudonym and a ticket based on proxy blind signature. On the other hand, the sophisticated unlinkability is implemented where mobile user is able to keep his pseudonym unchanged within the same operator in order to gain better user experience. PPS is presented as a suite of authentication and key agreement protocols built upon the proposed three-tire hierarchical network architecture. Our analysis demonstrates that PPS is secure and outperforms other proposal in terms of communication and computation overhead.

**Keywords:** Multi-operator wireless mesh network, privacy preservation, mutual authentication, security, user experience.

## 1. Introduction

Wireless mesh networks (WMNs) have recently emerged as a promising and competitive technology to cope with the challenges in next generation mobile network due to the features of self-organization, self-maintenance, as well as low upfront investment [1]. It can also be envisioned that the future large scale WMNs will be composed of a majority of autonomous domains managed by different operators as

opposed to few ones today [2]. Typically, in the multi-operator WMNs scenario as Fig.1, each operator maintains its own mesh backbone including mesh gateway and mesh routers, or shares some of the infrastructure components with other operators to provide network services to the mesh clients. Whereas mesh client may be associated with one or more operators by contractual means and has the ability to roam to the rest of the cooperating operators, if necessary. Different operators in a given geographical area will cooperate with each other in order to obtain large scale coverage and more consecutive user experience. However, security issues inherited from the intrinsically dynamic and open nature of wireless networks are still the main obstacle for the wide deployment of WMNs since it is unappealing to subscribers to obtain access and service without security guarantees. In addition, different operators may hold different security management policies, which will make the security control more complicated in the multi-operator WMNs. To this end, some proposals on WMNs security [3-4] have been presented recent years. In [3], the authors developed a broker-based attack-resilient security architecture (ARSA) for WMNs to address a wide range of particular attacks. We [4] proposed a localized efficient mutual authentication scheme (LEAS) with identity-based proxy signature [5] for access security in multi-operator WMNs. Despite the necessity and importance, security of WMNs is still in its early stage and has gained little attention so far [6].

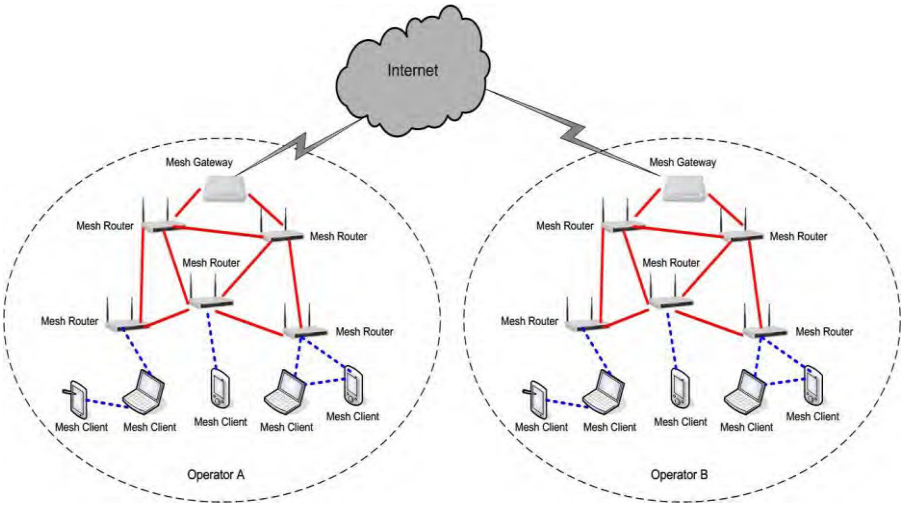



**Fig.1**. A typical architecture of multi-operator wireless mesh networks

Another big challenge for actually deploying WMNs with a multi-operator manner is how to provide adequate protection over user privacy since the communications contain various kinds of sensitive user information like personal identities, location information, financial information, social connections, and so on. Once disclosed to malicious attackers, the sensitive information could be illegally utilized or further be correlated together to compromise user privacy. Besides, the dynamic network architecture, hop-by-hop open wireless link, as well as autonomous yet cooperating operators render

WMNs highly vulnerable to various privacy-oriented attacks. Hence, privacy-preserving is of paramount practical importance in multi-operator WMNs.

The most important requirement of user privacy is anonymity that is concerned with hiding the real identity information of a user from his activities unless it is intentionally disclosed by himself. Different communication sessions associated with the same user should also be unlinkable to prevent association analysis. In reality, anonymity is conflicted with authentication or access control. With perfect anonymity, a user can misbehave arbitrarily and avoid being traced even to the identity issuer. Therefore, accountability is highly desirable for detecting and tracing malicious users in case of disputes and frauds. In terms of the above privacy requirements, several schemes have been proposed recently that are surveyed by [1] to meet the privacy-preserving needs for WMNs. However, limited literature research has been conducted to multi-operator context where operators are geographically distributed yet cooperating with each other. While user roaming across different operator WMNs, novel security architecture should be set up and conscious tradeoffs must be made to achieve both privacy-preserving authentication and fine user experience. According to [7], a new plan declared by Disney World will track visitors with wireless bracelets. Imagine walking through Disney World, Snow White walks up to you and wishes your child a happy birthday by name. Something like that could make an already memorable trip even more amazing. The cost of such a program is that your privacy, such as name, age, or even the credit card information, will be encoded in the bracelets. So Disney is able to track you during your trip or later. How to make a balance between privacy and user experience, is really a new challenge in multi-operator WMNs.

In this paper, we propose a privacy-preserving security scheme for large-scale multi-operator WMNs upon a three-tire hierarchical security architecture. Broker, acts as the root trust on the top tire, is responsible for the security management of all the involved entities. Based on such architecture, a novel mutual authentication scheme equipped with key agreement ability is achieved that takes inter-operator and intra-operator roaming scenarios into account. The combination of pseudonym and ticket is introduced as the authentication credential in our scheme. In light of the privacy requirements, on one hand a tri-lateral pseudonym approach is presented to meet anonymity need without key escrow. On the other hand, a ticket based on proxy blind signature (PBS) [8] is designed for mobile user against being traced from operator and broker. Both the pseudonym and the ticket can be altered by mobile user at his will when roaming across different operators. Thus the sophisticated unlinkability is implemented where mobile user is able to keep his pseudonym unchanged within the same operator in order to gain better user experience. In addition, the accountability is also satisfied due to the salient features borrowed from e-cash system on PBS. The system analysis demonstrates that our scheme is secure and outperforms similar one in terms of communication and computation overhead.

Specifically, our contributions are 3-folded as follows:

- The variable tri-lateral pseudonym approach and PBS-based ticket are designed to deal with the anonymity and untraceability needs;
- Sophisticated unlinkability is achieved through the bind of pseudonym and operator-level ticket in order to gain enhanced user experience;
- Accountability property is incorporated with the idea inherited from e-cash system to detect malicious users.

To sum up, our research is mainly focus on the security and privacy issues in multi-operator WMNs. It should be noted that the implementation of routing security and anonymity is out of the scope of this paper, which is left as the future works.

The rest of this paper is organized as follows. Section 2 reviews the identity-based primitives. Section 3 presents the system model including the hierarchical network architecture. We propose the mutual authentication scheme in terms of different roaming scenarios in Section 4. In Section 5, we provide security and performance analysis of our scheme. Section 6 discusses the related work. Finally, we conclude the paper in Section 7.

## 2. The Cryptographic Background

### 2.1. Bilinear Pairing

Let G be an additive group and $G_T$ be a multiplicative group of the same prime order q, $I_G$ and $I_{GT}$ is the generator of G and $G_T$ respectively. Assume that the discrete logarithm problem [9] is hard on both G and $G_T$. A mapping $\hat{e}: G \times G \to G_T$ which satisfies the following properties is called bilinear pairing:

(1) Bilinear: For all $P, Q \in G$ and $a, b \in Z_q^*$, $\hat{e}(a \cdot P, b \cdot Q) = \hat{e}(b \cdot P, a \cdot Q) = \hat{e}(P, Q)^{ab}$;

(2) Non-degenerate: $\hat{e}(P, Q) \neq I_{GT}$;

(3) Computable: For all $P, Q \in G$, there is an efficient approach to compute $\hat{e}(P, Q) \in G_T$.

The Weil and Tate [10] associated supersingular elliptic curve can be modified to construct such bilinear pairing.

### 2.2. Short Signature (BLS)

Boneh et al. [11] proposed short signatures (BLS) from the Weil pairing in 2001, which is a simple but efficient signature scheme. It is designed for systems where signatures are sent over a low-bandwidth channel. The scheme is specified as following algorithms.

***Setup.***

PKG chooses additive group $G_1$ and multiplicative group $G_2$, as well as a bilinear pairing $\hat{e}: G_1 \times G_1 \to G_2$; PKG chooses arbitrary $P \in G_1$ and a hash function $H_1: \{0,1\}^* \to G_1$.

***Key Generation.***

User selects random $x \in Z_q^*$ and computes $R = x \cdot P$. $R$ is public key and $x$ is private key.

***Sign.***

To sign a message m, signer computes $V = x \cdot H_1(m)$. $V$ is the signature.

***Verify.***

To verify $V$, verifier checks whether $\hat{e}(R, H_1(m)) == \hat{e}(P, V)$.

### 2.3. Identity-based Proxy Signature

The concept of proxy signatures was first introduced by Mambo et al. [12] in 1996. A proxy signature scheme permits an original signer to delegate its signing rights to a proxy signer so that it can sign on behalf of the original signer within a given context. Holding a proxy signature, anyone can verify both the delegation of original signer and the digital signature from proxy signer. Bo Gyeong Kang et al. [5] constructed a concrete identity-based proxy signature (IBPS) which is derived from BLS and CBE [13] as below.

***Setup.***

Assume Alice (original signer) and Bob (proxy signer) have private/public key pairs ($s_A$, $s_A \cdot P$) *and* ($s_B$, $s_B \cdot P$) respectively and the common system parameters $PARA = (G_1, G_2, \hat{e}, P, H_1, H_2)$, where two hash functions $H_1 : \{0,1\}^* \to G_1$ *and* $H_2 : \{0,1\}^* \times G_1 \to Z_q^*$ *are defined*.

***Delegation.***

In order to delegate signing right to Bob, Alice sends to Bob a warrant $\omega$ together with a BLS signature $Cert_B = s_A \cdot P_B, where\ P_B = H_1(PK_A \| PK_B \| \omega)$. The corresponding proxy signing key of Bob is $SKP_B = Cert_B + s_B \cdot P_B$.

***Sign.***

To sign message $m$ on behalf of Alice, Bob selects secrect random $r \in Z_q^*$ and computes $\sigma = (U, V)$, where $U = r \cdot P_B$, $h = H_2(m, U)$, and $V = (r + h)SKP_B$.

***Verify.***

To verify signature $\sigma$, verifier checks whether $\hat{e}(PK_A + PK_B, U + h \cdot P_B) == \hat{e}(P, V), where\ h = H_2(m, U)$.

## 3. System Model

Our concrete privacy-preserving security scheme is based on the following system model which contains network architecture, trust model, as well as privacy model. After some definitions of handover types and credentials, a three-tire hierarchical network architecture is first presented to support different kinds of handovers in multi-operator WMNs. Both trust and privacy model are then illustrated making the trust hypothesis and privacy needs explicit. The system is also initialized to develop the later proposed security scheme.

### 3.1. Definitions and Notations

**Definitions.** Some definitions that are frequently used in this paper are given in this subsection.

*Inter-operator handover.* Inter-operator handover occurs when mesh client roams from one operator WMNs to another under the same trust broker.

*Intra-operator handover.* Intra-operator handover refers that mesh client handoffs from one mesh router to another within the same operator WMNs.

*Certificate.* The certificate here is different from the X.509 public key certificate in PKI [14] which manifests the binding of owner's identity and public key. In contrast, our certificate is a delegation from issuer to owner and used in IBPS.

*Pseudonym.* Pseudonym, generated by some cryptographic primitives, is one of user's authentication credentials whereas contains no essential identity information (e.g. SSN or driver's license) of user.

*Ticket.* Ticket is the other authentication credential hold by mesh router or mesh client. We define three types of ticket for the later proposed authentication scheme.

(1) RTK: Mesh router's ticket which has long-term validity throughout multi-operator WMNs.

(2) CTK: Mesh client's ticket which has long-term validity throughout multi-operator WMNs.

(3) OTK: Mesh client's operator-level ticket which has short-term validity within operator WMNs.

*Double deposit.* A type of misbehavior that refers to mesh client's double depositing his CTKs at the same visiting mesh router.

**Notations.** To simplify the hereafter descriptions, we make some notations in Table 1.

**Table 1.** Notations and explanations

| Notation | Meaning |
|---|---|
| B | Broker |
| OM (O) | operator manager |
| MR (R) | mesh router |
| MC (C) | mesh client |
| ID_X | real identity of entity X |
| PS_X | pseudonym of entity X |
| Cert_X | certificate of entity X |
| RTK_X | ticket of mesh router X |
| CTK_X | ticket of mesh client X |
| OTK_X | operator-level ticket of mesh client X |
| $A_M$ | account of mesh client |
| $PK_X$ /$SK_X$ | public and secret key of entity X |

| | |
|---|---|
| $\overrightarrow{PK_X}$ / $\overrightarrow{SK_X}$ | self-generated public and secret key of entity X from PS_X |
| PARA | system parameters |
| $X_{INFO}$ | Related information of entity X |
| $P_X$ | Hash value of $X_{INFO}$ |
| $\{M\}_{\alpha\_Sign\_SK}$ | sign message M with algorithm α and secret key SK |
| $\{\sigma\}_{\beta\_Verify\_PK}$ | verify signature σ with algorithm β and public key PK |
| $K_{X-Y}$ | shared key between entity X and entity Y |
| $SEK_{X-Y}$ | session key between entity X and entity Y |
| $SKP_X$ | proxy signing key of entity X |
| TS | timestamp |
| Exp | expiration time of ticket or certificate |
| X→Y:[M] | entity X sends message M to entity Y |
| M1‖M2 | concatenation of two messages M1 and M2 |

### 3.2. Network Architecture

The three-tire hierarchical network architecture in Fig.2 is set up for multi-operator large-scale WMNs where each operator WMNs is taken as an administrative domain.

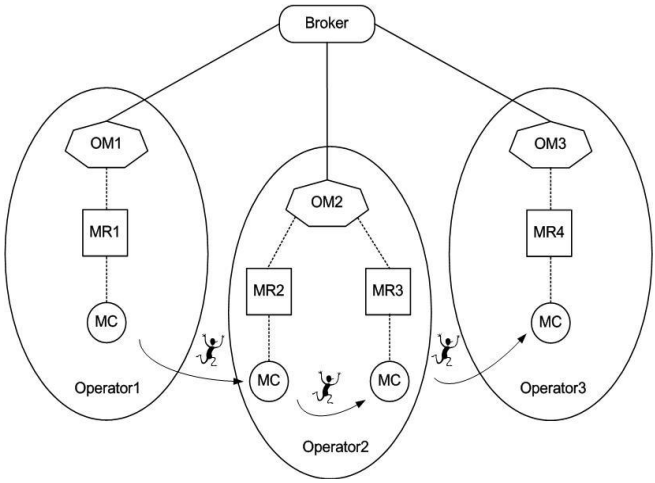



**Fig.2** Hierarchical network architecture for multi-operator WMNs, which is composed of three administrative domains

Broker on the top tier of the hierarchical architecture is introduced as a trusted anchor for all domains. The second tier of the architecture is composed of OMs who take the role of connectors between operator domain and broker and is in charge of the registration and trust management for MRs, as well as MCs inside operator domain. In reality, the functionalities of OM can be achieved into mesh gateway who shares reachability to all MRs through either direct or multi-hop wireless links as shown in Fig.1. MRs form the third tire of our security architecture and can provide access

service for both local and roaming MCs. MC associated with certain operator may take arbitrary handover across different operator domains under the hierarchical architecture.

From the collaboration point of view, any operator domain in our architecture is able to create relationship with others in order to provide larger-scale coverage and more access opportunities through signing service level agreements (SLAs) by the OMs.

### 3.3. Trust Model

In the context of multi-operator WMNs, the main security goals include:

- Mutual authentication. Users and visiting network should authenticate each other before user's access to avoid both malicious users and rouge routers.
- Confidentiality. After a successful user access, the subsequent communications between user and entities in the visiting network should be further protected to prevent different attacks such as eavesdropping and modification.

Due to the above security goals and the intrinsically open and collaborative features of multi-operator WMNs, it is essential to establish trust relationships among entities against free riders and malicious attackers.

As shown in Fig.3, our trust model is constructed in terms of the proposed hierarchical network architecture. The trust relationships among entities are defined and elaborated as follows:

- Broker, functions as a trustworthy administrator, is the root trust for all operator domains.
- OMs have long-term trust relationship with broker. Meanwhile, two OMs may also trust each other if they have signed SLA before. The SLA contains all the credible public keys of OM and MRs ($PK_O$ and $PK_R$) in the other operator domain.
- MRs have long-term trust relationship with the OM in the same operator domain.
- MCs have long-term trust relationship with the OM in their home operator domain.
- There is no trust relationship between MC and MR before MC's access. Two MCs do not trust each other.

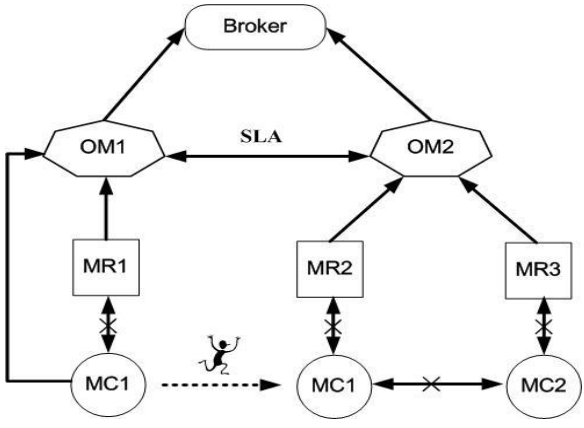



**Fig.3** Broker-based trust model

The trust relationship above means that there is a pre-established secure channel between two entities. The later proposed mutual authentication scheme is based on this trust model and the objective is how to build trust relationship between MC and access MR as well as the trust relationships amongst MCs.

It is also worth noting that our trust model is different from the one presented in [3], where broker and operator issue authentication credentials to MCs and MRs separately. Different trust anchors make the trust management more implicit. In contrast, broker takes the role of root trust in our trust model. Any operator could not issue credentials to MRs or MCs without broker's permission and delegation. The trust management is thus more explicit and is suitable for the security control in multi-operator WMNs.

### 3.4. Privacy Model

In addition to keep access and communication secure, privacy provision is another critical issue to be considered for WMNs deployment. However, privacy is difficult to achieve even if traffics are protected since users' activities can be easily monitored or traced with regard to their movement, which may cause the exposure of the sensitive information. Therefore, the establishment of a practical privacy model is necessary to provide adequate privacy concerns and detect malicious users simultaneously.

**Anonymity.** User's activities, during the roaming procedure, should not be correlated to his real identity (e.g. SN or driver's license). In our privacy model, we utilize pseudonym and ticket as hybrid authentication credential to achieve user anonymity. Neither pseudonym nor ticket contains real identity of user so that user can roam anonymously in multi-operator WMNs.

**Untraceability.** For untraceability, it is required that the credential issuer can't trace user's activity during the roaming procedure. Thus both the pseudonym and the ticket should be alerted by user while roaming.

**Sophisticated unlinkability.** On one hand, from the privacy-preserving point of view, different communication sessions from the same user should not be linked against association analysis. On the other hand, from the user experience point of view, the recognizable credential is preferable in the same operator WMNs or collaborative operator WMNs. For such sophisticated unlinkability, user is equipped with variable pseudonym and temporary operator-level ticket in our privacy model to keep balance between privacy and user experience.

**Accountability.** Unconditional anonymity may result in perfect crimes since misbehaving users are no longer traceable. Therefore, accountability is highly desirable for detecting and tracing malicious users. We borrow the idea from e-cash system to form a novel ticket management scheme. The real identity of misbehaving user, who double deposits his CTK at the same MR, could be disclosed with the help of broker and OM.

In summary, our privacy model aims at the above privacy guarantees meanwhile takes user experience into account. It's a trade off: giving up some privacy in return for an enhanced user experience.

### 3.5. System Initialization

In order to support the proposed security framework, our system must be initialized to distribute indispensable system parameters, certificates, as well as key materials to involved entities. Specifically, the following system initialization steps should be performed when the network bootstrapped.

System parameter generation

**(1)** Broker generates parameter tuple $(G_1, G_2, \hat{e}, P, Q, H, H_1, H_2)$, where $P$ and $Q$ are generators of $G_1$, $\hat{e}$ is a billiner pairing, hash functions $H: G_2 \to Z_q^*$, $H_1: \{0,1\}^* \to G_1$, $H_2: \{0,1\}^* \times G_1 \to Z_q^*$.

**(2)** Broker randomly selects a master secrect key $SK_B = S_B \in Z_q^*$ and calculates the public key $PK_B = S_B \cdot P$, then publishes the system parameter $PARA=(G_1, G_2, \hat{e}, P, Q, H, H_1, H_2, PK_B)$.

OM certificate insurance

**(1)** Each OM randomly selects a secrect key $SK_O = S_O \in Z_q^*$ and calculates its public key $PK_O = S_O \cdot P$ according to *PARA*.

**(2)** $O \to B: [PK_O]$

**(3)** Broker generates certificate for OM: $Cert\_O = S_B \cdot P_O, where\ P_O = H_1(PK_B \| PK_O \| Exp)$ .

**(4)** $\mathrm{B} \to O: [Cert\_O]$

**(5)** OM calculates the proxy signing key: $SKP_O = Cert\_O + S_O \cdot P_O = (S_B + S_O)P_O$.

MR ticket insurance

**(1)** Each MR randomly selects a secrect key $SK_R = S_R \in Z_q^*$ and calculates its public key $PK_R = S_R \cdot P$ according to *PARA*.

**(2)** $\mathrm{R} \to \mathrm{O}: [PK_R]$

**(3)** OM generates certificate and RTK for managed MR: $Cert\_R = S_O \cdot P_R, where\ P_R = H_1(PK_O \| PK_R \| Exp)$; $RTK\_R =< Exp, PK_B, PK_O, PK_R, \sigma >$, where $\sigma = \{Exp \| PK_B \| PK_O \| PK_R\}_{IBPS\_Sign\_SKP_O}$.

**(4)** $\mathrm{O} \to \mathrm{R}: [Cert\_R,\ RTK\_R]$

**(5)** MR calculates the proxy signing key $SKP_R = Cert\_R + S_R \cdot P_R = (S_O + S_R) \cdot P_R$.

Through the above system initialization, OMs and MRs obtain their certificates and proxy signing keys with the delegated right from broker. Besides, MRs are also equipped with the RTKs which will be applied into the following proposed mutual authentication scheme.

## 4. PPS: The Proposed Scheme

To address the security and privacy concerns in multi-operator WMNs with enhanced user experience, we propose a privacy-preserving mutual authentication scheme, upon the security system, together with accountability capability. The scheme is based on the hybrid employment of pseudonym and ticket to achieve anonymity, untraceability, as

well as sophisticated unlinkability. In light of the handover types defined in section 3.1, we take two authentication scenarios (as shown in Fig.2) into account: inter-operator authentication and intra-operator authentication. Shared key establishment is also integrated into PPS to protect subsequent communications in the air and gain more efficiency. In addition, we also consider MC-MC authentication and user accountability issues in multi-operator WMNs. In this section, we will give the details of PPS.

### 4.1. Pseudonym Generation

The pseudonym is used to hide the real identity of user during the roaming procedure, which is necessary for both anonymity and user experience. Moreover, in order to meet the sophisticated unlinkability need, the pseudonym should also be variable in our design. The widely adopted way to achieve that is to assign a batch of pseudonyms to user and showing one each time [15, 16]. However, the communication and update cost are the main obstacles. In [6], the authors presented a more efficient method. The pseudonym is generated with the help of an authority while can be alerted by user whenever needed. As such, user is able to frequently update his pseudonym to enhance unlinkability. Unfortunately, the authority may learn user's secret key which is derived from the pseudonym, thus results in the key-escrow problem and violates the untraceability requirement.

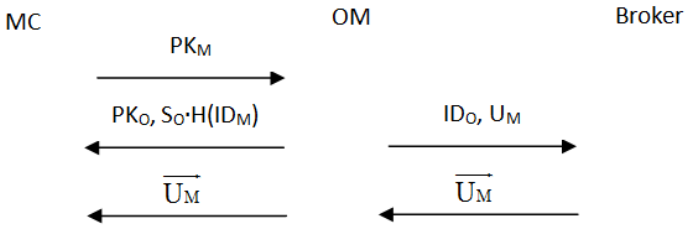



**Fig. 4** Workflow of tri-lateral pseudonym generation among MC, OM and Broker

To address the above issues, we propose a tri-lateral pseudonym generation approach as shown in Fig.4. Before the approach bootstrapping, MC first registers the real identity ($ID_M$) to the home domain OM through either offline method or the pre-established secure channel. Afterwards the following steps are executed for the pseudonym generation.

**(1)** MC randomly selects a secrect key $SK_M = S_M \in Z_q^*$ and calculates its public key $PK_M = S_M \cdot P$ according to *PARA*.

**(2)** $\mathrm{M} \rightarrow \mathrm{O}: [PK_M]$

**(3)** OM computes $K_{O-M} = \hat{e}(S_O \cdot Q, PK_M)$, then derives $k_M = H(K_{O-M})$, $U_M = k_M \cdot H_1(ID_M)$, $A_M = S_O \cdot U_M$, where $A_M$ is MC's account at OM. OM further stores the binding relation $<ID_M, A_M, k_M, U_M>$ for MC.

**(4)** $\mathrm{O} \rightarrow \mathrm{M}: [PK_O, A_M, S_O \cdot H_1(ID_M)]$

**(5)** MC computes $K_{M-O} = \hat{e}(S_M \cdot Q, PK_O)$, then derives $k_M = H(K_{M-O})$.

**(6)** $\mathrm{O} \rightarrow B:\left[ID_O,\ U_M\right]$

**(7)** Broker computes $\overrightarrow{U_M} = S_B \cdot U_M$

**(8)** $\mathrm{B} \rightarrow \mathrm{O} \rightarrow \mathrm{M}:\left[\overrightarrow{U_M}\right]$

**(9)** MC computes $\overrightarrow{U_M} \cdot k_M^{-1} = S_B \cdot H_1(ID_M)$, then generates the pseudonym $PS_M = S_M \cdot H_1(ID_M)$ and the correspoding key pair: $\overrightarrow{PK_M} = (S_B + S_O) \cdot H_1(ID_M)$, $\overrightarrow{SK_M} = S_M \cdot \overrightarrow{PK_M} = (S_B + S_O) \cdot PS_M$.

A pairing-based key agreement method is incorporated into the above procedure. It can be easily proved that:

$$K_{O\text{-}M} = \hat{e}(S_O \cdot Q, PK_M) = \hat{e}(Q,P)^{S_O \cdot S_M} = \hat{e}(S_M \cdot Q, PK_O) = K_{M\text{-}O}$$

The agreed key ($K_{O\text{-}M}$/$K_{M\text{-}O}$) and the relevant key material ($k_M$) are the building blocks of our tri-lateral pseudonym generation approach. Such keys are the secret knowledge shared between MC and OM. We can also find that the pseudonym is self-generated by user with his own secret ($S_M$) thus can be altered at his will. Meanwhile, the secret key with regard to the pseudonym is composed of broker's secret ($S_B$) and OM's secret ($S_O$). The key escrow problem is averted as neither broker nor OM knows the secret key of the other party. Moreover, any MC can sign a message ($m$) with the generated $\overrightarrow{SK_M}$ using BLS: $\sigma = \{m\}_{\mathrm{BLS\_Sign\_}\overrightarrow{SK_M}} = \overrightarrow{SK_M} \cdot H_1(m)$. Any party may verify $\sigma$ using BLS: $\{\sigma\}_{\mathrm{BLS\_Verify\_}PS_M\&PK_B\&PK_O}$.

## 4.2. Ticket Insurance

Ticket is the other authentication credential in PPS. The insurance of RTK has been presented by in section3.5. We will elaborate CTK's insurance procedure in this section.

PBS is borrowed for the generation and insurance of CTK. The insurance procedure can be carried out locally between MC and OM who owes the delegation from broker. The detailed procedure is demonstrated through the following steps.

**(1)** OM randomly selects $r \in Z_q^*$ and calculates $R = r \cdot P$.

**(2)** $\mathrm{O} \rightarrow \mathrm{M}:\left[R,\ O_{INFO}\right]$, where $O_{INFO} = <PK_B,\ PK_O>$.

**(3)** MC randomly selects $a,\ b,\ \alpha,\ \beta \in Z_q^*$ and $\omega$, where $\omega$ is an agreement between MC and OM such as *Exp* or other restrictions on the CTK.

**(4)** MC calculates: $d1 = \alpha \cdot PK_O,\ d2 = \beta \cdot PK_O,\ d = \beta \cdot A_M,\ P_O = H_1(PK_B \,||\, PK_O),\ t = \hat{e}(R + a \cdot P, PK_O) \cdot \hat{e}(b \cdot P_O,\ PK_B + PK_O),\ C' = H_2(d \,||\, d1 \,||\, d2 \,||\, \omega \,||\, O_{INFO} \,||\, t) + b$

**(5)** $\mathrm{M} \rightarrow \mathrm{O}:\left[C'\right]$.

**(6)** OM caculates $S' = C' \cdot SKP_O + r \cdot PK_O$.

**(7)** $\mathrm{O} \rightarrow \mathrm{M}:\left[S'\right]$.

**(8)** MC first calculates: $S = S' - a \cdot PK_O,\ C = C' - b,\ t' = \hat{e}(S, P) \cdot \hat{e}(-C \cdot P_O,\ PK_B + PK_O)$, then checks whether $C == H_2(d \| d1 \| d2 \| \omega \| O_{INFO} \| t')$, if the equation holds, MC obtains the CTK=<$d$, $d$1, d2,$\omega$, $O_{INFO}$, S, C>; Otherwise, MC quits the procedure.

Actually, <S, C> in CTK is the signature result of PBS on < $d$, $d$1, d2,$\omega$, $O_{INFO}$ > and step (8) is the PBS verification process. CTK together with pseudonym will be utilized during the authentication between MC and visiting MR in PPS.

### 4.3. Inter-operator Authentication

In Fig.2, while a MC (M), registered with OM1 (O1) in operator1 domain, entering operator2 domain managed by OM2 (O2) and accessing MR2 (R2), inter-operator authentication protocol is executed between MC and MR2 as below.

**(1)** $R2 \rightarrow M: [RTK\_R2 =< Exp, PK_B, PK_{O2}, PK_{R2}, \sigma 1 = \{Exp \| PK_B \| PK_{O2} \| PK_{R2}\}_{IBPS\_Sign\_SKP_{O2}} >]$ through beacon message.

**(2)** MC executes the following operations:
- Check the validity of $Exp$ in RTK_R2;
- Verify $\sigma 1$ with $PK_B$ and $PK_{O2}$ : $\{\sigma 1\}_{IBPS\_Verify\_PK_B\&PK_{O2}}$;
- Computes $K_{M-R2} = \hat{e}(\overline{SK_M}, PK_{R2})$.

**(3)** $M \rightarrow R2: [PS_M,\ CTK\_M =< d, d1, d2, \omega = Exp, O1_{INFO} =< PK_B, PK_{O1} >, S, C >, t1, \sigma 2 = \{CTK\_M \| t1\}_{BLS\_Sign\_\overline{SK_M}}]$, where $t1$ is the current timestamp.

**(4)** MR2 executes the following operations:
- Check the validity of $Exp$ in CTK_M and the freshness of $t$1;
- Verify $\sigma 2$ with $PK_B$, $PK_{O1}$, $PS_M$: $\{\sigma 2\}_{BLS\_Verify\_PK_B\&PK_{O1}\&PS_M}$;
- Verify $< S, C >$ in CTK_M with $PK_B$, $PK_{O1}$: $\{S, C\}_{PBS\_Verify\_PK_B\&PK_{O1}}$;
- Compute $K_{R2-M} = \hat{e}(S_{R2} \cdot PS_M, PK_B + PK_{O1})$.

**(5)** $R2 \rightarrow M: [e,\ t2,\ \sigma 3 = \{e \| t2\}_{HMAC\_Sign\_K_{R2-M}}]$, where $e$ is a challenge selected from $\{0,1\}^*$ and $t2$ is the current timestamp.

**(6)** MC executes the following operations:
- Check the freshness of t2;
- Verify $\sigma 3$ with $K_{M-R2}$: $\{\sigma 3\}_{HMAC\_Verify\_K_{M-R2}}$. If the verification success, MC regards MR2 as a legitimate MR.
- Computes: $u = H_2(CTK\_M \| e \| d2), v = \beta + \alpha \cdot u$.

**(7)** $M \rightarrow R2: [u,\ v,\ t3,\ \sigma 4 = \{u \| v \| t3\}_{HMAC\_Sign\_K_{M-R2}}]$, where $t3$ is the current timestamp.

**(8)** MR2 executes the following operations:
- Check the freshness of t3;
- Verify $\sigma 4$ with $K_{R2-M}$: $\{\sigma 4\}_{HMAC\_Vverify\_K_{R2-M}}$;
- Compute $t' = \hat{e}(S, P) \cdot \hat{e}(-C \cdot P_{O1},\ PK_B + PK_{O1})$;
- Check whether $u == H_2(CTK\_M \| e \| v \cdot PK_{O1} - u \cdot d1)$ and $C == H_2(d \| d1 \| v \cdot PK_{O1} - u \cdot d1 \| Exp \| O1_{INFO} \| t')$.
- If all the equations hold, MR2 regards MC as a legitimate user and stores <CTK_M, e, u, v > for MC.

**(9)** $R2 \rightarrow M: [OTK\_M =< PS_M, R2_{INFO}, PK_{O1}, Exp, t4, \sigma 5 = \{PS_M || R2_{INFO} || Exp || t4\}_{IBPS\_Sign\_SKP_{R2}} >]$, where $R2_{INFO} =< PK_{O2}, PK_{R2} >$ and $t4$ is the current timestamp.

**(10)** MC does the followings:
- Check the freshness of t4;
- Verify $\sigma 5$ in OTK_M with $PK_B$, $PK_{O2}$, and $PK_{R2}$ : $\{\sigma 5\}_{IBPS\_Verify\_PK_{O2}\&PK_{R2}}$. If the verification success, MC obtains OTK_M as a legitimate OTK.

After the inter-operator authentication, MC and MR2 are able to generate their session key $SEK_{M\text{-}R2}=H(K_{M\text{-}R2}||t1||t2)$ respectively to protect the subsequent communications.

It should be noted that the HMAC [17] operations introduced above are symmetric-key method which is much more efficient than the public-key ones as BLS, IBPS, as well as PBS. In addition, in order to achieve untraceability and unlinkability across operators, the pseudonym should be altered by MC each time when accessing a new operator domain. After successful mutual authentication between MC and MR2 through steps (1)-(8), MR2 directly issues OTK to MC (by step (9)) with the proxy signing key ($SKP_{R2}$) delegated from OM2 and broker. This OTK will be utilized as an authentication credential during the following intra-operator authentication scheme.

### 4.4. Intra-operator Authentication

Intra-operator authentication occurs while MC (M) moves from MR2 (R2) to MR3 (R3) within operator2 domain managed by OM2 (O2) as shown in Fig.2. The authentication protocol is as below.

**(1)** $R3 \rightarrow M: [RTK\_R3 =< Exp, PK_B, PK_{O2}, PK_{R3}, \sigma 6 = \{Exp || PK_B || PK_{O2} || PK_{R3}\}_{IBPS\_Sign\_SKP_{O2}} >]$ through beacon message.

**(2)** MC verifies $\sigma 6$ in RTK_R3 with $PK_B$ and $PK_{O2}$ : $\{\sigma 6\}_{IBPS\_verify\_PK_B\&PK_{O2}}$, then computes $K_{M\text{-}R3} = \hat{e}(\overrightarrow{SK_M}, PK_{R3})$.

**(3)** $M \rightarrow R3: [OTK\_M =< PS_M, PK_{O1}, R2_{INFO}, Exp, t4, \sigma 5 = \{PS_M || PK_{O1} || R2_{INFO} || Exp || t4\}_{IBPS\_Sign\_SKP_{R2}} >, t5, \sigma 7 = \{OTK\_M || t5\}_{HMAC\_Sign\_K_{M\text{-}R3}}]$, where $t5$ is the current timestamp.

**(4)** MR3 executes the following operations:
- Check the validity of $Exp$ in OTK_M and the freshness of $t5$;
- Verify $\sigma 5$ with $PK_{O2}$ and $PK_{R2}$: $\{\sigma 5\}_{IBPS\_Verify\_PK_{O2}\&PK_{R2}}$;
- Compute $K_{R3-M} = \hat{e}(S_{R3} \cdot PS_M, PK_B + PK_{O1})$;
- Verify $\sigma 7$ with $K_{R3-M}$: $\{\sigma 7\}_{HMAC\_verify\_K_{R3\text{-}M}}$;
- If all the above verifications hold, MR3 regards MC as a legitimate user.

**(5)** $R3 \rightarrow M: [t6, \sigma 8 = \{t6\}_{HMAC\_Sign\_K_{R3\text{-}M}}]$, where $t6$ is the current timestamp.

**(6)** MC executes the following operations:
- Check the freshness of t6;
- Verify $\sigma 8$ with $K_{M\text{-}R3}$ : $\{\sigma 8\}_{HMAC\_Verify\_K_{M\text{-}R3}}$. If the verification success, MC regards MR3 as a legitimate MR.

When the intra-operator authentication finished, MC and MR3 are able to generate their session key $SEK_{M\text{-}R3}=H(K_{M\text{-}R3}||t5||t6)$ respectively to protect the subsequent communications.

OTK is effective within the operator domain. We can see from the above intra-operator authentication that the deposit and verification of OTK are based on HMAC and IBPS operations which are more efficient than the PBS process on CTK. Besides, MC may keep the pseudonym unchanged in the same operator domain in order to gain better user experience. However, from the unlinkability point of view, MC could also choose to show CTK and new generated pseudonym at accessing MR to allow frequent update of OTK.

Another issue should be considered is that MC handoffs across two cooperated operator domains. In our trust model, the OM of the two domains shares the trusted public keys ($PK_O$, $PK_R$) in the other domain through SLA. Theses public keys are further distributed to the managed MRs periodically by OM. For example, if operator2 and operator3 (in Fig.2) are cooperated, then OM3 will record the trusted $PK_{O2}$, $PK_{R2}$, $PK_{R3}$ and broadcast them to MR4, vice versa. In light of this, OTK_M is still effective in operator3 domain though MC makes an inter-operator domain handover from MR3 to MR4, since MR4 is able to verify such OTK_M with $PK_{O2}$ and $PK_{R2}$ using the same operations in intra-operator authentication scheme.

### 4.5. MC-MC Authentication

There is no pre-established trust relationship between two MCs. As a consequence, privacy-preserving MC-MC authentication and key agreement are critical. Fortunately, with the help of the above proposed authentication schemes, MC-MC authentication can be easily implemented.

Suppose that two MCs (M1, M2) registered to different OMs (O1, O2) hold their CTKs (CTK_M1, CTK_M2) respectively. Mutual authentication between M1 and M2 is achieved as the inter-operator authentication scheme along with following steps.

**(1)** $M1 \rightarrow M2:[PS_{M1},\ \overline{SK_{M1}}\cdot P,\ CTK\_M1 =< d,\ d1,\ d2,\ \omega=Exp,\ O1_{INFO} =< PK_B,\ PK_{O1} >,\ S,\ C>,\ t7, \sigma 9 = \{CTK\_M1 \,||\, \overline{SK_{M1}}\cdot P \,||\, t7\}_{BLS\_Sign\_\overline{SK_{M1}}}]$, where t7 is the current timestamp.

**(2)** M2 executes the following operations:

- Check the validity of *Exp* in CTK_M1 and the freshness of t7;
- Verify $\sigma 9$ with $PK_B$, $PK_{O1}$, $PS_{M1}$: $\{\sigma 9\}_{BLS\_Verify\_PK_B\&PK_{O1}\&PS_{M1}}$;
- Verify $< S,\ C >$ in CTK_M1 with $PK_B$, $PK_{O1}$: $\{S, C\}_{PBS\_Verify\_PK_B\&PK_{O1}}$;
- If all the verification success, M2 regards M1 as a legitimate MC;
- Compute $K_{M2-M1} = \hat{e}(\overline{SK_{M2}}\cdot Q,\ \overline{SK_{M1}}\cdot P)$.

**(3)** $M2 \rightarrow M1:[PS_{M2},\ \overline{SK_{M2}}\cdot P,\ CTK\_M2 =< d',\ d1',\ d2',\ \omega=Exp',\ O1_{INFO} =< PK_B,\ PK_{O2} >,\ S',\ C'>,\ t8, \sigma 10 = \{CTK\_M2 \,||\, \overline{SK_{M2}}\cdot P \,||\, t8\}_{BLS\_Sign\_\overline{SK_{M2}}}]$, where t8 is the current timestamp.

**(4)** M1 executes the following operations:

- Check the validity of Exp' in CTK_M2 and the freshness of t8;
- Verify $\sigma 10$ with $PK_B$, $PK_{O2}$, $PS_{M2}$: $\{\sigma 10\}_{BLS\_Verify\_PK_B\&PK_{O2}\&PS_{M2}}$;
- Verify $<S', C'>$ in CTK_M2 with $PK_B$, $PK_{O2}$: $\{S', C'\}_{PBS\_Verify\_PK_B\&PK_{O2}}$;
- If all the verification success, M1 regards M2 as a legitimate MC;
- Compute $K_{M1-M2} = \hat{e}(\overline{SK_{M1}} \cdot Q, \overline{SK_{M2}} \cdot P)$.

It is obvious from the above operations that

$$K_{M1-M2} = \hat{e}(\overline{SK_{M1}} \cdot Q, \overline{SK_{M2}} \cdot P) = \hat{e}(Q,P)^{\overline{SK_{M1}} \cdot \overline{SK_{M2}}} = \hat{e}(\overline{SK_{M2}} \cdot Q, \overline{SK_{M1}} \cdot P) = K_{M2-M1}.$$

After the MC-MC authentication finished, M1 and M2 generate their session key $SEK_{M1\text{-}M2}=H(K_{M1\text{-}M2}||t7||t8)$ respectively to protect the subsequent communications.

### 4.6. User Accountability

PPS achieves fine user privacy through the combination of pseudonym and ticket, while still maintaining user accountability. In PPS, MC authenticates himself as a legitimate service subscriber to the OM ($O_H$) in the home operator domain. The real identity of MC ($ID_M$) and his account ($A_M$) are only known by himself and $O_H$. Neither the visiting OM ($O_V$) nor the broker has knowledge of MC's privacy information during his roaming. However, from the accountability point of view, it is necessary to detect malicious MCs. As described in our system model, MC's misbehavior is defined as his double depositing the CTKs at the same visiting mesh router ($R_V$).

Assume that a MC accesses a foreign operator WMNs and double deposits his CTKs (CTK1, CTK2) to a $R_V$. Then two authentication records will be left at $R_V$ according to the proposed inter-operator authentication scheme: Record1 <CTK1, e1, u1, v1> and Record2 <CTK2, e2, u2, v2>. In order to disclose the identity of such malicious MC, the following operations are executed with the collaboration of $R_V$, $O_V$, $O_H$, as well as broker.

**(1)** $R_V \rightarrow O_V \rightarrow B : [Record1, Record2]$.

**(2)** Broker deduces between Record1 and Record2 to compute $\beta = \frac{u2 \cdot e1 - u1 \cdot e2}{e1 - e2}$, broker further obtains $A_M = \beta^{-1} \cdot d$, where d is in MC's CTK.

**(3)** $B \rightarrow O_H : [A_M]$.

**(4)** $O_H$ obtains $U_M = S_{O_H}^{-1} \cdot A_M$, thus to discolse $ID_M$ through the binding relation <$ID_M$, $A_M$, $k_M$, $U_M$> stored during the pseudonym generation phase.

The implementation of the above user accountability function is due to the features of e-cash system based on PBS.

# 5. System Analysis

## 5.1. Security and privacy analysis

**Authenticity.** Mutual authentication is achieved in PPS to avert both free riders and bogus service providers. MC is equipped with pseudonym and ticket issued by OM under the delegation from broker. Owing such authentication credentials, MC is able to roam securely across multi-operator WMNs in light of the root trust to broker. In addition, the proposed inter-operator authentication scheme and intra-operator authentication scheme are implemented locally between MC and visiting MR for better efficiency.

**Confidentiality.** Communicating entities establish a shared symmetric key and the corresponding session key to secure their subsequent communications after authentication. In PPS, we adopt pairing-based key agreement approach to construct such keys between MC and the visiting MR. The symmetric key is also used in the mutual authentication protocols together with HMAC operations in order to mitigate the computation burden on both MC and MR sides.

**Anonymity.** MC takes pseudonym and CTK as the authentication credentials during the roaming procedure. While the pseudonym is composed of MC's own secret and the hash value of MC's identity information: $PS_M = S_M H_1(ID_M)$.

The $CTK =< d, d1, d2, \omega, O_{INFO}, S, C >$ contains some cryptographic results derived from MC's account (AM) and public keys of OM and broker, as well as the PBS signature on them. Neither pseudonym nor CTK comprises real identity of MC so that the anonymity is guaranteed during MC's roaming. Moreover, MC is also unable to know the real identity of the visiting MR since such information is not included in the $RTK =< E\exp, PK_B, PK_O, PK_R, \sigma >$. Thus the anonymity is bidirectional.

**Untraceability.** Untraceability requires that the credential issuer can't trace MC's activity when he is roaming. On one hand, the pseudonym in PPS can be alerted by MC at his will to avoid the traceability from OM and broker. On the other hand, MC's CTK is also different between the insurance phase and the showing phase due to the non-key escrow feature of PBS. Consequently, OM cannot trace MC's activity through the CTK.

**Sophisticated unlinkability.** Sophisticated unlinkability is preferable in order to give consideration to both privacy-preserving and user experience. In PPS, when MC roaming across different operator WMNs, although the CTK remains unchanged, while the pseudonym is required to be alerted by MC. Thus the adversary is unable to link different communication sessions to the same user. In addition, MC will obtain a temporary OTK after the inter-operator authentication procedure. Owing such OTK and a constant pseudonym, MC can gain better user experience within the same operator WMNs.

**User accountability.** User accountability is so important in PPS for detecting malicious users. To achieve this, for a legitimate MC, none of the entities, including broker, OMs, as well as MRs, could disclose the real identity of MC in terms of the above anonymity and untraceability features. However, if MC double deposits his CTK at a visiting MR

which is defined as misbehavior, upon the collusion of visiting OM, broker, and home OM, the real identity of malicious MC can be exposed with the help of the accountability function borrowed from PBS-based e-cash system.

## 5.2. Performance Analysis

In this section, the performance analysis of our scheme, PPS, in terms of communication and computation overhead is presented compared with the similar security approach of SAT [6] which also utilizes pseudonym and ticket as hybrid authentication credentials. Our analysis takes both inter-operator and intra-operator authentication scenarios into account. In addition, since the resource-constraint mesh client is the performance bottleneck of the whole system, our performance analysis is thus mainly focus on the mesh client side.

Without loss of generality, we borrow the parameters from [6] and [18] in the following analysis, resulting in the elements length in $G_1$ ($|G_1|$) and $G_2$ ($|G_2|$) to be roughly 171 bits and 1024 bits respectively. We also assume that SHA-1[19] is used in our HMAC operations, that yields a 160-bit output.

**Communication Overhead.** Communication overhead refers to the communication cost incurred by MC during the authentication procedure. The overhead is mainly composed of the pseudonym, ticket, signature, as well as HMAC result transmitted from MC side, where the shorter components are out of consideration compared with the above ones, such as the *Exp* and *TS*.

*Inter-operator Communication Overhead.* In SAT, a tree-based hierarchical security architecture and pseudonym approach is proposed. Both hierarchical pseudonym ($PST_M$) and client pseudonym ($PS_M$) should be transmitted by MC during inter-operator authentication. SAT introduces a ticket based on restrictive partially blind signature [20]. The total ticket length is $5|G_1|+2|G_2|$. In contrast, only one self-generated pseudonym is involved in inter-operator authentication in PPS contributed to our delegated trust model. Moreover, the CTK in PPS is signed with PBS, which makes the total ticket length $6|G_1|$. As a consequence, our ticket length is greatly reduced compared with SAT since $|G_2|$ is much longer than $|G_1|$. In light of the above analysis, we can observe from Table 2 that the inter-operator communication overhead of PPS outperforms SAT greatly over 59%.

*Intra-operator Communication Overhead.* There is no need of hierarchical pseudonym during the intra-operator authentication in SAT. However the same ticket ($5|G_1|+2|G_2|$) as in inter-operator authentication is still necessary. As described in section 4.4, an OTK ($6|G_1|$) is transmitted by MC instead of CTK ($6|G_1|$) plus pseudonym ($1|G_1|$) during intra-operator authentication in PPS, which will further reduces the communication overhead.

As shown in Table 3, the intra-operator communication overhead of PPS drops down almost 67% compared with that of SAT.

**Table 2.** Analysis results of inter-operator communication overhead

| Scheme | Inter-operator communication overhead | Total bits |
|---|---|---|
| **SAT** | $PST_M$: $1\|G_1\|$<br>$PS_M$: $1\|G_1\|$<br>Key material: $1\|G_1\|$<br>$\sigma_{HIBS}$: $1\|G_1\|$<br>Ticket: $5\|G_1\|+2\|G_2\|$<br>$\sigma_{HMAC}$: $1\|HMAC\|$<br>**Total: $9\|G_1\|+2\|G_2\|+\|HMAC\|$** | 3747 |
| **PPS** | $PS_M$: $1\|G_1\|$<br>CTK: $6\|G_1\|$<br>$\sigma_{BLS}$: $1\|G_1\|$<br>$\sigma_{HMAC}$: $1\|HMAC\|$<br>**Total: $8\|G_1\|+\|HMAC\|$** | 1528 |

**Note:** $\sigma_{HIBS}$, $\sigma_{HMAC}$, $\sigma_{BLS}$ denote the signature results from HIBS [21], HMAC, and BLS respectively.

**Table 3.** Analysis results of intra-operator communication overhead.

| Scheme | Intra-operator communication overhead | Total bits |
|---|---|---|
| **SAT** | $PS_M$: $1\|G_1\|$<br>$2\sigma_{BLS}$: $2\|G_1\|$<br>Ticket: $5\|G_1\|+2\|G_2\|$<br>$\sigma_{HMAC}$: $1\|HMAC\|$<br>**Total: $8\|G_1\|+2\|G_2\|+\|HMAC\|$** | 3576 |
| **PPS** | OTK: $6\|G_1\|$<br>$\sigma_{HMAC}$: $1\|HMAC\|$<br>**Total: $6\|G_1\|+\|HMAC\|$** | 1186 |

**Table 4.** Computational cost of the operations on MC side during authentication.

| | SM | PA | BP | MG | MTP | Hash |
|---|---|---|---|---|---|---|
| $\mathbf{BLS_s}$ | 1 | N/A | N/A | N/A | 1 | N/A |
| $\mathbf{BLS_v}$ | N/A | N/A | 2 | N/A | 1 | N/A |
| $\mathbf{HIBS_s}$ | 1 | 1 | N/A | N/A | 1 | N/A |
| $\mathbf{HIBS_v}$ | N/A | N/A | 3 | 2 | 1 | N/A |
| $\mathbf{IBPS_s}$ | 2 | N/A | N/A | N/A | N/A | 1 |
| $\mathbf{IBPS_v}$ | 1 | 2 | 2 | N/A | N/A | 1 |
| $\mathbf{HMAC_s}$ | N/A | N/A | N/A | N/A | N/A | 1 |
| $\mathbf{HMAC_v}$ | N/A | N/A | N/A | N/A | N/A | 1 |
| **KA** | N/A | N/A | 1 | N/A | N/A | N/A |

**Note:** $BLS_{s/v}$, $HIBS_{s/v}$, $IBPS_{s/v}$ denote the signing and verifying operations of each schemes respectively. KA denotes the key generation operation.

**Computation Overhead.** Communication overhead refers to the computation cost experienced at MC side during the authentication procedure, which mainly caused by the signing, verifying, as well as key generating operations. The involved operations consist of bilinear pairing (BP), scale multiplication (SM), point addition (PA), multiplication in group (MG), map to point function (MTP), and hash function (Hash). We first report the cost of these operations in Table 4 for the consequent analysis.

*Inter-operator Computation Overhead.* Table 5 shows the computation operations involved in the inter-operator authentication of SAT and PPS. With the correlated observation from Tab.4 and Table 5, we can draw the following conclusions:

$$IRCO_{SAT}=4BP+2MTP+1SM+1PA+2MG+1Hash \quad (1)$$

$$IRCO_{PPS}=3BP+1MTP+2SM+2PA+3Hash \quad (2)$$

where $IRCO_{SAT}$ and $IRCO_{PPS}$ represent the inter-operator computation overhead of SAT and PPS respectively.

**Table 5.** Analysis results of inter-operator computation overhead.

| Scheme | $BLS_s$ | $HIBS_s$ | $HIBS_v$ | $IBPS_v$ | $HMAC_s$ | $HMAC_v$ | KA |
|---|---|---|---|---|---|---|---|
| **SAT** | N/A | 1 | 1 | N/A | 1 | N/A | 1 |
| **PPS** | 1 | N/A | N/A | 1 | 1 | 1 | 1 |

Let $t_x$ denote the computational cost of operation x. According to [22-23], $t_{PA}$, $t_{MG}$, and $t_{Hash}$ are negligible compared with $t_{BP}$ ,$t_{MTP}$, and $t_{SM}$. In addition, based on the analysis results in [24], we also get the following conclusions:

$$t_{BP} = 2\, t_{MTP} = 3t_{SM} \quad (3)$$

Through equations (1)-(3), we obtain that $IRCO_{PPS}$ is about 78% of $IRCO_{SAT}$ since less BP operations are involved in PPS than in SAT.

*Intra-operator Computation Overhead.* The computation operations deal with the intra-operator authentication of SAT and PPS are shown as Table 6. The following conclusions are able to be obtained through the combination of Table 4 and Table 6.

$$IACO_{SAT}=4BP+2MTP+1SM+2MG+1Hash \quad (4)$$

$$IACO_{PPS}=3BP+1SM+2PA+3Hash \quad (5)$$

where $IACO_{SAT}$ and $IACO_{PPS}$ represent the intra-operator computation overhead of SAT and PPS respectively.

**Table 6.** Analysis results of intra-operator computation overhead

| Scheme | $BLS_s$ | $HIBS_s$ | $HIBS_v$ | $IBPS_v$ | $HMAC_s$ | $HMAC_v$ | KA |
|---|---|---|---|---|---|---|---|
| **SAT** | 1 | N/A | 1 | N/A | 1 | N/A | 1 |
| **PPS** | N/A | N/A | N/A | 1 | 1 | 1 | 1 |

Through equations (3)-(5), $IACO_{PPS}$ is only 62.5% of $IACO_{SAT}$ as the computation consuming operations in PPS are further mitigated during intra-operator authentication.

Though PPS owes better computation overhead compared with SAT from the above analysis. We can still see some computation intensive BP operations in PPS. However, many literature efforts have been made to speedup BP computation either by software or hardware means. For example, in [32], the authors propose a set of software optimizations for BP computation and demonstrate the feasibility of integrating BP-based security approaches into wireless network. The performance results show that it only take 0.14s for BP computation even on Imote2 embedded platform [33]. The authors of [34] also present the FPGA implementation of BP on mobile device which only needs 1.07ms for the computation. Such realizations are able to make PPS more practical in multi-operator WMNs against the heavy computation overhead.

## 6. Related work

Security and privacy issues in WMNs have gained considerable research focus in the literature. Most of these efforts fall in the scope of addressing the general security and privacy issues or establishing cross-domain security architecture.

Some efforts depend on identity manipulation approaches to satisfy the security and privacy requirements in WMNs. [18] organizes mobile users into different groups, the identity information is only known to the user and the group manager. The anonymity and unlikability are achieved through the variant short group signature [25] and late binding scheme. In terms of the feature of group signature, user accountability is also implemented with the collusion of domain manager and group manager. However, the key escrow problem is still existed and high computation cost is obligatory on user side. Ahmet Onur Durahim et al. [26-27] introduce an authority that is responsible of issuing pseudonym for mobile user as authentication credential. They utilize DAA [28] to achieve the anonymity and untraceability during user's roaming. Furthermore, the malicious users can be tracked by the collusion of the authority and domain manager. While the scheme suffers from the public key management problem inherited from PKI.

Other efforts take cross-domain authentication issues into account. Wang Z. et al. [29] propose a security architecture and trust model regards to cross-domain scenarios. The hierarchical credential is designed for user anonymity and cross-domain authentication. In addition, the certificateless cryptographic approach [30] is adopted in the authentication procedure to avert the key-escrow problem. Unfortunately, the other privacy requirements beyond anonymity, such as unlinkability, untraceability, accountability, are not involved in the scheme. [6] brings another cross-domain hierarchical security architecture for WMNs based on HIBS scheme. Most of the privacy requirements are also satisfied due to the usage of partially blind signature

scheme. However, some drawbacks in accountability procedure of [6] have been pointed out by [31].

In summary, the literature research are mainly focus on the security and privacy issues of WMNs, few of them take multi-operator scenarios and user experience into the design account. These are the motivations for us to provide our privacy-preserving security scheme with fine user experience for multi-operator WMNs.

## 7. Conclusion

In this paper, we propose PPS, a privacy-preserving security scheme for multi-operator WMNs, which addresses the conflicting privacy requirement of unlinkability and fine user experience. By hybrid utilization of the tri-lateral variable pseudonym approach and different kinds of tickets under identity-based proxy signature (IBPS) and proxy blind signature (PBS), anonymity, untraceability, as well as sophisticated unlinkability are satisfied during MC's roaming. User accountability is also achieved through PBS-based e-cash system that is incorporated into our mutual authentication protocols equipped with key agreement features. Our analysis shows that PPS is able to implement desired security objectives and high efficiency.

As a future work, intensive simulations of PPS, e.g. on NS3 [35], should be made to further demonstrate its feasibility. We also plan to develop location privacy approach and anonymous routing scheme for multi-operator WMNs upon our hierarchical security architecture.

**Acknowledgements.** This work was supported by Major National Scientific & Technological Projects of China under Grant No. 2013ZX03002006.

## References

1. Jaydip Sen: Secure and Privacy-Preserving Authentication Protocols for Wireless Mesh Networks. Book Chapter in Applied Cryptography and Network Security, 3-34. (2012)
2. Ze Wang, Maode Ma, Wenju Liu, Xixi Wei: A Unified Security Framework for Multi-domain Wireless Mesh Networks, Lecture Notes in Computer Science, Vol. 7043, 319-329. (2011)
3. Yanchao Zhang, Yuguang Fang: ARSA: An Attack-Resilient Security Architecture for Multihop Wireless Mesh Networks. IEEE journal on selected areas in communications, 24(10): 1916-1928. (2006)
4. Tianhan Gao, Nan Guo, Kangbin Yim: LEAS: Localized Efficient Authentication Scheme for Multi-operator Wireless Mesh Network with Identity-based Proxy Signature. Mathematical and Computer Modeling, Volume 58, Issues 5–6, 1427-1440. (2013)
5. Bo Gyeong Kang, Je Hong Park, Sang Geun Hahn: A Certificate-Based Signature Scheme. In Proceedings of The Cryptographer's Track at RSA Conference - CT-RSA, 99-111. (2004)
6. Jinyuan Sun, Chi Zhang, Yanchao Zhang: SAT: A Security Architecture Achieving Anonymity and Traceability in Wireless Mesh Networks. IEEE Transactions on Dependable and Secure Computing, Vol. 8, No. 2, 295-307. (2011)
7. When Privacy and Enhanced User Experience Collide Online. [Online]. Available: http://inklingmedia.net/2013/01/10/when-privacy-and-enhanced-user-experience-collide-online/(current July 2014)

8. Zuowen Tan: An E-Cash Scheme Based on Proxy Blind Signature from Bilinear Pairings. Journal of Computers, Vol.5, No. 11, 1638-1645. (2010)
9. Joseph H. Silverman: The Arithmetic of Elliptic Curves. Springer. (2009)
10. Antoine Joux: The Weil and Tate Pairings as Building Blocks for Public Key Cryptosystems Survey. In Proceedings of the 5th International Symposium on Algorithmic Number Theory (ANTS-V), LNCS 2369, 11-18. (2002)
11. Dan Boneh, Ben Lynn, Hovav Shacham: Short Signatures from the Weil Pairing. In Proceedings of ASIACRYPT - ASIACRYPT , 514-532. (2001)
12. M. Mambo, K. Usuda, E. Okamoto. Proxy signatures: delegation of the power to sign messages. Transactions on Fundamentals of Electronic Communications and Computer Science, vol. E79-A, 1338-1354. (1996)
13. Craig Gentry: Certificate-Based Encryption and the Certificate Revocation Problem. In Proceedings of Theory and Application of Cryptographic Techniques - EUROCRYPT , 272-293. (2003)
14. C. Adams, S. Farrell, T. Kause, T. Mononen: Internet X.509 Public Key Infrastructure Certificate Management Protocol (CMP). RFC4210. (2005)
15. M. Raya, J-P. Hubaux: Securing Vehicular Ad Hoc Networks,Journal of Computer Security. special issue on security of ad hoc and sensor networks, Vol. 15, No. 1, 39-68. (2007)
16. G. Ateniese, A. Herzberg, H. Krawczyk, G. Tsudik: Untraceable Mobility or How to Travel Incognito. Computer Networks,Vol. 31, No. 8, 871-884. (1999)
17. H. Krawczyk, M. Bellare, R. Canetti: HMAC: Keyed-Hashing for Message Authentication. RFC2104. (1997)
18. Kui Ren, Shucheng Yu, Wenjing Lou, Yanchao Zhang: PEACE: A Novel Privacy-Enhanced Yet Accountable Security Framework for Metropolitan Wireless Mesh Networks. IEEE Transactions on Parallel and Distributed Systems, Vol. 21, No. 2, 203-215. (2010)
19. D. Eastlake, P. Jones: US Secure Hash Algorithm 1 (SHA1). RFC3174. (2001)
20. X. Chen, F. Zhang, and S. Liu: ID-Based Restrictive Partially Blind Signatures and Applications. Journal of Systems and Software, vol. 80(2), 164-171. (2007)
21. Craig G, Alice S: Hierarchical ID-based cryptography. In: Proc. of the 8th Int'l Conf. on the Theory and Application of Cryptology and Information Security. LNCS 2501, 548-566. (2002)
22. Sandip Vijay, Subhash C. Sharma: Threshold signature cryptography scheme in wireless ad-hoc computing. Contemporary Computing, 40 (7), 327–335. (2009)
23. Mohamed Abid, Songbo Song, Hassnaa Moustafa, Hossam Afifi: Integrating identity-based cryptography in IMS service authentication. International Journal of Network Security and Its Applications, 1–13. (2010)
24. Tianhan Gao, Nan Guo, Kangbin Yim: A Hybrid Approach to Secure Hierarchical Mobile IPv6 Networks. Computer Science and Information Systems, Vol. 10 No. 2, 913-938. (2013)
25. D. Boneh and H. Shacham: Group Signatures with Verifier-Local Revocation. In Proc. ACM Conf. Computer and Comm. Security (CCS), 168-177. (2004)
26 Ahmet Onur Durahim, Erkay Savas: A-MAKE: An Efficient, Anonymous and Accountable Authentication Framework for WMNs. In Proceedings of Fifth International Conference on Internet Monitoring and Protection, 54-59. (2010)
27. Ahmet Onur Durahim, Erkay Savas: A2-MAKE: An efficient anonymous and accountable mutual authentication and key agreement protocol for WMNs. Ad Hoc Networks 9(7): 1202-1220. (2011)
28. E. F. Brickell, J. Camenisch, and L. Chen: Direct anonymous attestation. In Proc. of ACM CCS 04, 132–145. (2004)
29. Ze Wang, Maode Ma, Wenju Liu, Xixi Wei: A Unified Security Framework for Multi-domain Wireless Mesh Networks, Lecture Notes in Computer Science, Vol. 7043, 319-329. (2011)
30. Al-Riyami, S.S., Paterson, K.G: Certificateless Public Key Cryptography. In: Laih, C.-S.(ed.) ASIACRYPT 2003, LNCS, vol. 2894, 452–473. (2003)

31. Huaqun Wang,Yuqing Zhang: On the Security of a Ticket-Based Anonymity System with Traceability Property in Wireless Mesh Networks. IEEE Transactions on Dependable and Secure Computing, Vol. 9, No. 3, 443-446. (2012)
32. Leonardo B. Oliveira, Diego F. Aranha, Conrado P.L. Gouvêa, Michael Scott, Danilo F. Câmara, Julio López , Ricardo Dahab: TinyPBC: Pairings for authenticated identity-based non-interactive key distribution in sensor networks, Computer Communications, Vol. 34, No. 3, 485–493. (2011)
33. L. Nachman, R. Kling, R. Adler, J. Huang, V. Hummel: The intel mote platform: abluetooth-based sensor network for industrial monitoring. In Proceedings of Fourth International Symposium on Information Processing in Sensor Networks, 437–442. (2005)
34. Sylvain D, Nicolas G: A FPGA pairing implementation using the residue number system. Cryptology ePrint Archive. (2011). [Online]. Available: http://eprint.iacr.org/2011/176 (current July 2014)
35. George F. Riley, Thomas R. Henderson: The ns-3 Network Simulator, Modeling and Tools for Network Simulation, 15-34. (2010)

**Tianhan Gao** received the BE in Computer Science & Technology, the ME and the PhD in Computer Application Technology, from Northeastern University, China, in 1999, 2001, 2006, respectively. He joined Northeastern University in April 2006 as a lecture of Software College. He obtained an early promotion to an associate professor in January 2010. He has been a visiting scholar at department of Computer Science, Purdue, from February 2011 to February 2012. He is the author or co-author of more than 30 research publications. His primary research interests are next generation network security, MIPv6/HMIPv6 security, wireless mesh network security, Internet security, as well as security and privacy in ubiquitous computing.

**Nan Guo** received the BE in Computer Science & Technology, the ME and the PhD in Computer Application Technology, from Northeastern University, China, in 1999, 2001, 2005, respectively. She joined Northeastern University in September 2005. She has been an associate professor since 2008. She has been a visiting scholar at department of Computer Science, Purdue, from August 2010 to August 2011. She is the author or co-author of more than 20 research publications. Her primary research interests are security and privacy in service computing and digital identity management.

**Kangbin Yim** received his B.S., M.S., and Ph.D. from Ajou University, Suwon, Korea in 1992, 1994 and 2001, respectively. He is currently an associate professor in the Department of Information Security Engineering, Soonchunhyang University. He has served as an executive board member of Korea Institute of Information Security and Cryptology, Korean Society for Internet Information and The Institute of Electronics Engineers of Korea. He also has served as a committee chair of the international conferences and workshops and the guest editor of the journals such as JIT, MIS, JISIS and JoWUA. His research interests include vulnerability assessment, code obfuscation, malware analysis, leakage protection, secure hardware, and systems security. Related to these topics, he has worked on more than fifty research projects and published more than a hundred research papers.

**Qianyi Wang** received her B.S. from Troy University, U.S.A in 2012, and working on her M.S in university of Malaya, Malaysia. She is currently a research student of Department of Economics and Administration, Faculty of Economics and Administration, University of Malaya. Her primary research interest is rural land consolidation in China. She is also involved in researches of China rural economy development, spatial planning as well as technology development.

*Received: September 17, 2013; Accepted: January 16, 2014.*

Computer Science and Information Systems 11(3):1001–1016 DOI: 10.2298/CSIS130915061J

# A Computer Remote Control System Based on Speech Recognition Technologies of Mobile Devices and Wireless Communication Technologies

Hae-Duck J. Jeong, Sang-Kug Ye, Jiyoung Lim, Ilsun You, and WooSeok Hyun

Department of Computer Software
Korean Bible University
Seoul, South Korea
hdjjeong@gmail.com, pajamasi726@hanmail.net
{jylim, isyou, wshyun}@bible.ac.kr

**Abstract.** This paper presents a computer remote control system using speech recognition technologies of mobile devices and wireless communication technologies for the blind and physically disabled population as assistive technology. These people experience difficulty and inconvenience using computers through a keyboard and/or mouse. The purpose of this system is to provide a way that the blind and physically disabled population can easily control many functions of a computer via speech. The configuration of the system consists of a mobile device such as a smartphone, a PC server, and a Google server that are connected to each other. Users can command a mobile device to do something via speech; such as writing emails, checking the weather forecast, or managing a schedule. These commands are then immediately executed. The proposed system also provides blind people with a function via TTS(Text To Speech) of the Google server if they want to receive contents of a document stored in a computer.

**Keywords:** speech recognition technology, mobile device, Android, wireless communication technique.

## 1. Introduction

Speech recognition technology, which is able to recognize human speech and change to text, or to perform a command, has emerged as the 'Next Big Thing' of the IT industry. Speech recognition is technology that uses desired equipment and a service which can be controlled through voice without using items such as a mouse or keyboard. It also appeared as part of ongoing research in progress in 1950s, but was not popularized until the mid-2000s, with low voice recognition. Presently, related speech recognition technologies, which have been previously used limitedly for special-purposes, have been rapidly evolving because of the proliferation of portable computing terminals such as smartphones interconnected with the expansion of the cloud infrastructure [8].

One of the most prominent examples of a mobile voice interface is *Siri*, the voice-activated personal assistant that comes built into the latest iPhone. But voice functionality is also built into Android, the Windows Phone platform, and most other mobile systems, as well as many applications. While these interfaces still have considerable limitations, we are inching closer to machine interfaces we can actually talk to [7].

This paper presents a computer remote control system using speech recognition technologies of mobile devices and wireless communication technologies for the blind and physically disabled population [5], [6], [13]. These people experience difficulty and inconvenience using computers through a keyboard and/or mouse. The purpose of this system is to provide a way the blind and physically disabled population can easily control many functions of a computer via speech. The configuration of the system consists of a mobile device such as a smartphone, a PC server, and a Google server that are connected to each other. Users command a mobile device to do something via speech such as directly controlling computers, writing emails and documents, calculating numbers, checking the weather forecast, or managing a schedule. These commands are then immediately executed. The proposed system also provides blind people with a function via TTS (Text To Speech) of the Google server when they want to receive contents of a document stored in a computer.

In Section 2, a few related works and technologies of the proposed remote computer control system are discussed. Section 3 describes comparison of speech recognition rates of current speech recognition systems. Section 4 presents how the proposed system using speech recognition technologies is designed and implemented, and finally the conclusions are described in Section 5.

## 2. Related Works and Technologies

Related works and technologies of the proposed computer remote control system using speech recognition technologies of mobile devices and wireless communication technologies are Android, and speech recognition algorithms as follows.

### 2.1. Android

Android is a Linux-based open mobile platform for mobile devices such as smartphones and tablet computers. It is composed of not only an operating system, but also middleware, user interface (UI), browser, and application. It also includes C/C++ libraries that are used in components of various Android systems [3]. Figure 1 shows that Android system architecture is divided into five hierarchical categories: applications, application framework, libraries, Android runtime, and Linux kernel [1], [2], [9]. The proposed application was designed and developed on Android.

### 2.2. Speech Recognition Algorithms

**Google Speech Recognition.** Google uses artificial intelligence algorithms to recognize spoken sentences, stores voice data anonymously for analysis purposes, and cross matches spoken data with written queries on the server. Key problems of computational power, data availability and managing large amounts of information are handled with ease using `android.speech.RecognizerIntent` package [1]. Client application starts up and prompts user to input using Google Speech Recognition. Input data is sent to the Google server for processing and text is returned to client. Input text is passed to the natural language processing (NLP) server for processing using HTTP (HperText Transfer

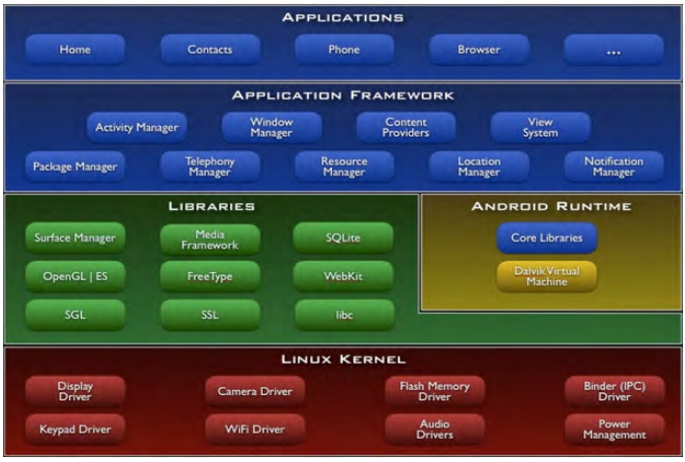



**Fig. 1.** Android system architecture.

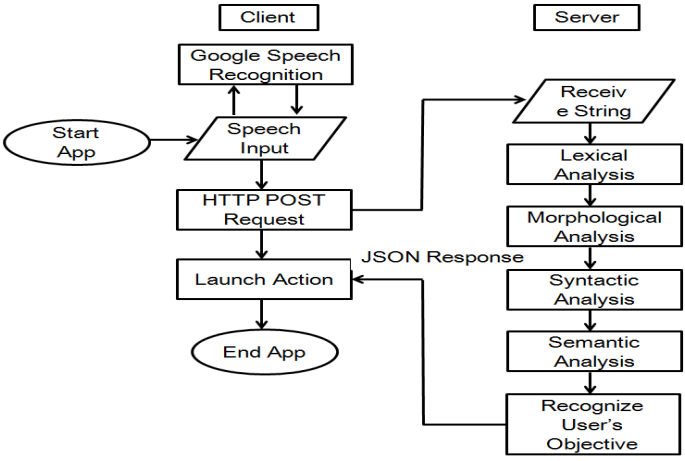



**Fig. 2.** Data flow diagram of speech recognition.

Protocol) POST [1]. Then the server performs NLP. Data flow diagram of speech recognition in Figure 2 shows that there are several steps involved in NLP as in the following:

1. **Lexical Analysis** converts sequence of characters into a sequence of tokens.
2. **Morphological Analysis** identifies, analyzes, and describes the structure of a given language's linguistic units.

[1] POST request is used to send data to a server. The string detected by speech recognizer is passed to the server using this method. It accomplishes this using in-built `HttpCore` API (i.e., `org.apache.http` package). The server performs processing and returns a JSON (JavaScript Object Notation) response. JSON is a lightweight data-interchange format, is based on a subset of the JavaScript programming language, and is completely language independent. In Java, `org.json.JSONObject` is used to parse strings [1].

3. **Syntactic Analysis** analyzes texts, which are made up of a sequence of tokens, to determine their grammatical structure.
4. **Semantic Analysis** relates syntactic structures from the levels of phrases and sentences to their language-independent meanings.

**Hidden Markov Model.** Modern general-purpose speech recognition systems are based on Hidden Markov Models (HMM). HMM is a doubly stochastic process with an underlying stochastic process that is not observable (it is hidden), but can only be observed through another set of stochastic processes that produce the sequence of observed symbols [4], [11]. HMMs are statistical models that output a sequence of symbols or quantities, and are used in speech recognition because a speech signal can be viewed as a piecewise stationary signal or a short-time stationary signal. In a short time-scales (e.g., 10 milliseconds), speech can be approximated as a stationary process. Speech can be thought of as a Markov model for many stochastic purposes [15]. Another reason why HMMs are popular is because they can be trained automatically and are simple and computationally feasible to use. In speech recognition, the hidden Markov model would output a sequence of n-dimensional real-valued vectors (with n being a small integer, such as 10), outputting one of these every 10 milliseconds. The vectors would consist of cepstral coefficients, which are obtained by taking a Fourier transform of a short time window of speech and decorrelating the spectrum using a cosine transform, then taking the first (most significant) coefficients. The hidden Markov model will tend to have in each state a statistical distribution that is a mixture of diagonal covariance Gaussians, which will give a likelihood for each observed vector. Each word, or (for more general speech recognition systems), each phoneme, will have a different output distribution; a hidden Markov model for a sequence of words or phonemes is made by concatenating the individual trained hidden Markov models for the separate words and phonemes.

The following notations for a discrete observation HMM are defined.
Let $T = \{1, 2, \cdots, T\}$ be the observation sequence (i.e., number of clock times), and $T$ is length of the observation sequence. Let $Q = \{q_1, q_2, \cdots, q_N\}$ be states, where $N$ is the number of states, $V = \{v_1, v_2, \cdots, v_M\}$ be discrete set of possible symbol observations, where $M$ is the number of possible observations, $A = \{a_{ij}\}$ be state transition probability distribution, where $a_{ij} = Pr(q_i \text{ at } t+1 | q_i \text{ at } t)$, $B = \{b_j(k)\}$ be observation symbol probability distribution in state $j$, where $b_j(k) = Pr(v_k \text{ at } t | q_j \text{ at } t)$, and $\pi = \{\pi_i\}$ be initial state distribution, where $\pi_i = Pr(q_i \text{ at } t = 1)$ [11].

The mechanism of the HMM is explained in the following:

- Step-1. Choose an initial state, $i_1$, according to the initial state distribution, $\pi$.
- Step-2. Set $t = 1$.
- Step-3. Choose $O_t$, according to $b_{i_t}(k)$, the symbol probability distribution in state $i_t$.
- Step-4. Choose $i_{t+1}$ according to $\{a_{i_t i_{t+1}}\}, i_{t+1} = 1, 2, \cdots, N$, the state transition probability distribution for state $i_t$.
- Step-5. Set $t = t + 1$; return to Step-3 if $t < T$; otherwise terminate the procedure.

We use the compact notation $\lambda = (A, B, \pi)$ to represent an HMM. For every fixed state sequence $I = i_1 i_2 \cdots i_\tau$, the probability of the observation sequence $O$ is $Pr(O|I, \lambda)$, where

$$Pr(O|I, \lambda) = b_{i_1}(o_1) b_{i_2}(o_2) \cdots b_{i_\tau}(o_\tau). \tag{1}$$

In other words, the probability of such a state sequence $I$ is

$$Pr(I|\lambda) = \pi_{i_1} a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_{\tau-1} i_\tau}. \tag{2}$$

The joint probability of $O$ and $I$ is simply the product of the above two terms,

$$Pr(O, I|\lambda) = Pr(O|I, \lambda) Pr(I|\lambda). \tag{3}$$

Then the probability of $O$ is obtained by summing this joint probability over all possible state sequences:

$$Pr(O|\lambda) = \sum_{allI} Pr(O|I, \lambda) Pr(I|\lambda) \tag{4}$$

$$= \sum_{i_1, i_2, \cdots, i_\tau} \pi_{i_1} b_{i_1}(o_1) a_{i_1 i_2} b_{i_2}(o_2) \cdots a_{i_{\tau-1} i_\tau} b_{i_\tau}(o_\tau). \tag{5}$$

**Neural Networks.** Neural networks emerged as an attractive acoustic modeling approach in automatic speech recognition (ASR) in the late 1980s. Since then, neural networks have been used in many aspects of speech recognition such as phoneme classification, isolated word recognition, and speaker adaptation [12], [15]. In contrast to HMMs, neural networks make no assumptions about feature statistical properties and have several qualities making them attractive recognition models for speech recognition. When used to estimate the probabilities of a speech feature segment, neural networks allow discriminative training in a natural and efficient manner. Few assumptions on the statistics of input features are made with neural networks. However, in spite of their effectiveness in classifying short-time units such as individual phones and isolated words, neural networks are rarely successful for continuous recognition tasks, largely because of their lack of ability to model temporal dependencies. Thus, one alternative approach is to use neural networks as a pre-processing e.g. feature transformation, dimensionality reduction, for the HMM based recognition.

**Other Speech Recognition Systems.** Modern speech recognition systems use various combinations of a number of standard techniques in order to improve results over the basic approach described above. A typical large-vocabulary system would need context dependency for the phonemes (so phonemes with different left and right context have different realizations as HMM states). It would use cepstral normalization to normalize for different speaker and recording conditions. For further speaker normalization it might use vocal tract length normalization (VTLN) for male-female normalization and maximum likelihood linear regression (MLLR) for more general speaker adaptation. The features would have so-called delta and delta-delta coefficients to capture speech dynamics and in addition might use heteroscedastic linear discriminant analysis (HLDA); or might skip the delta and delta-delta coefficients and use splicing and a linear discriminant analysis (LDA)-based projection followed perhaps by heteroscedastic linear discriminant analysis or a global semi-tied covariance transform (also known as maximum likelihood linear transform, or MLLT). Many systems use so-called discriminative training techniques that dispense with a purely statistical approach to HMM parameter estimation and instead optimize some classification-related measure of the training data. Examples are maximum mutual information (MMI), minimum classification error (MCE) and minimum phone error (MPE) [10], [15].

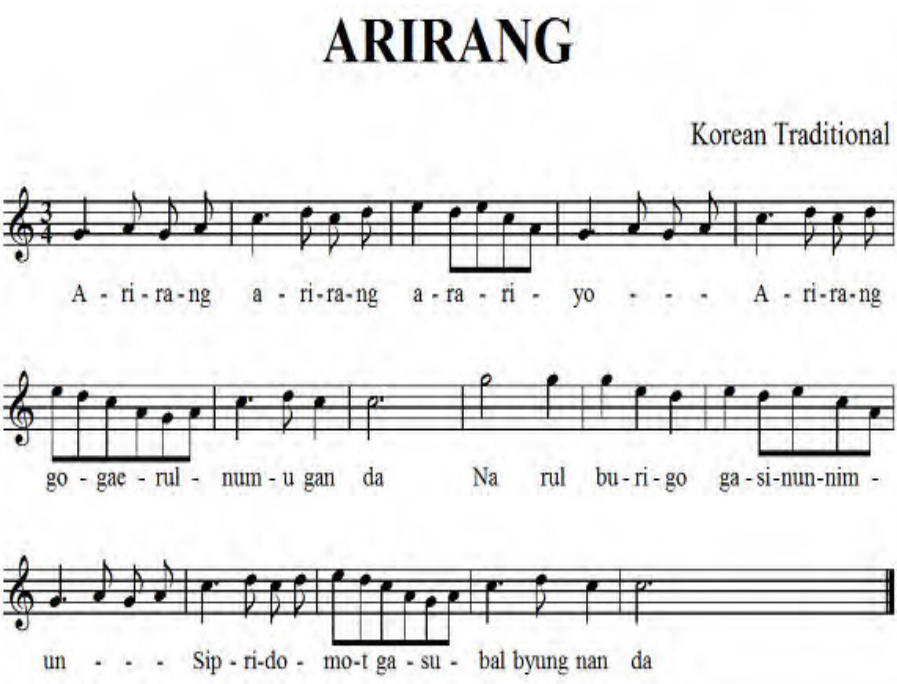



**Fig. 3.** Arirang note that is lyrical folk song in the Republic of Korea [14].

## 3. Comparison of Speech Recognition Rate

We have investigated how much recognition rates of current speech recognition systems, including Google speech recognition, NHN (Naver), Q Voice, S Voice, and Siri, are with Arirang [2], lyrical folk song in the Republic of Korea; and also see Arirang note in Figure 3.

One hundred replications in Korean were tested for each speech recognition system. According to our investigation, Table 1 shows that Google speech recognition system is the best of five speech recognition systems. Thus, it was used to design and implement our proposed system.

## 4. Implementation and Results

Figure 4 shows the architecture of the proposed system and command transmission methods among a mobile device, a Google server, and a personal computer server. The roles of each number are in the following:

1. A user commands using the speech recognition application of the mobile device.

[2] Arirang is a popular form of Korean folk song and the outcome of collective contributions made by ordinary Koreans throughout generations. Essentially a simple song, it consists of the refrain 'Arirang, arirang, arariyo' and two simple lines, which differ from region to region [14].

**Table 1.** Comparison of speech recognition rate for speech recognition systems.

| Speech recognition system | Recognition rate (%) | Smartphone type | Smartphone version | Techniques used |
|---|---|---|---|---|
| Google speech recognition | 100 | Galaxy III | Android4.1.2 | Google's own technology |
| NHN(Naver) | 51 | Galaxy III | Android4.1.2 | Link |
| Q Voice | 98 | Optimus G | Android4.1.2 | 1st step: Google 2nd step: Wernicke |
| S Voice | 96 | Galaxy III | Android4.1.2 | Vlingo |
| Siri | 94 | IPhone 5 | IOS 6.1 | Nuance |

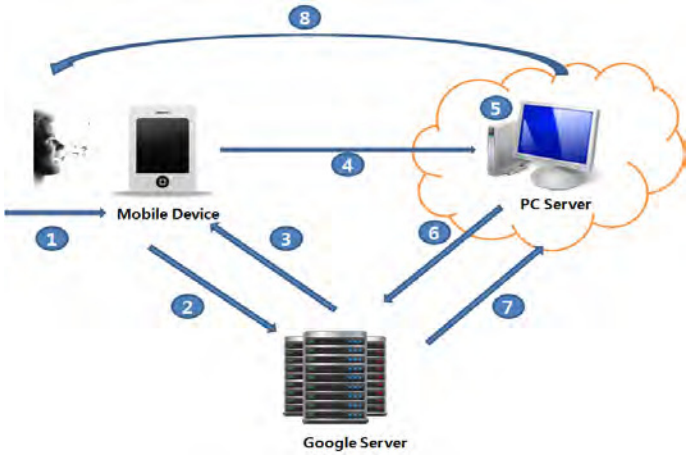



**Fig. 4.** Command transmission methods among a mobile device, a Google server, and a personal computer server.

2. Execute STT (speech to text) through the Google server.
3. Transmit results obtained from STT to the mobile device.
4. Transmit results obtained from STT to the personal computer server via wireless communications such as 3G, WIFI, and Bluetooth.
5. The personal computer server analyzes corresponding commands, and executes to distinguish between information which is sent to the Google server, and information which is executed on the personal computer server.
6. Transmit information to the Google server if there is information to use the Google server among commands.
7. The Google server returns corresponding values after analyzing corresponding services.
8. Give the user information received from the Google server with voice messages or execute.

Figure 5 shows overall use case diagram of the proposed system that contains more than five main functions such as speech recognition, keyboard control, mouse control, simple mode, and text transmission.


**Fig. 5.** Overall use case diagram of the proposed system.

Our proposed computer remote control system using speech recognition technologies of mobile devices and wireless communication technologies was implemented by Java programming language. The proposed application was designed and developed on Android as well.

### 4.1. Speech Recognition Mode

The below program code shows Java code of speech recognition for the proposed application. `startVoiceRecognitionActivity` fires an intent to start the speech recognition activity and `onActivityResult` handles the results from the recognition activity.

```
private void startVoiceRecognitionActivity() {
   Intent intent = new Intent(RecognizerIntent.ACTION_RECOGNIZE_SPEECH);
   intent.putExtra(RecognizerIntent.EXTRA_LANGUAGE_MODEL,
   RecognizerIntent.LANGUAGE_MODEL_FREE_FORM);
   intent.putExtra(RecognizerIntent.EXTRA_PROMPT, "Speech recognition demo");
   startActivityForResult(intent, VOICE_RECOGNITION_REQUEST_CODE);
 }
@Override
protected void onActivityResult(int requestCode, int resultCode, Intent data) {
  if (requestCode == VOICE_RECOGNITION_REQUEST_CODE && resultCode == RESULT_OK) {
  // Fill the list view with the strings the recognizer thought it could have heard
  ArrayList<String> matches =
                        data.getStringArrayListExtra(RecognizerIntent.EXTRA_RESULTS);
  mList.setAdapter(new ArrayAdapter<String>(this, android.R.layout.simple_list_item_1,
                                                                    matches));
  }
 super.onActivityResult(requestCode, resultCode, data);
}
```

Figure 6 shows speech recognition by touching the mobile device screen. When executing speech recognition by touching the top of the mobile device screen, all speech contents are typed and saved on the computer. When executing speech recognition by touching the bottom, corresponding service is executed by recognizing all speech contents. For example, a user commands the mobile device to do 'what is today's weather?' and then the remote system answers 'Today is 20 degrees Celsius and the weather is fine.' Another example is that a user from the outside commands his/her mobile device to do 'Send meeting document in the document folder.' and then the system finds it in the folder and transmits it to user's mobile device or a personal computer that he/she wants.

### 4.2. Keyboard Control Mode

Figure 7 demonstrates computer keyboard control by touching the smartphone screen. A computer's keyboard is controlled by a method that the key value entered by the user is transmitted from smartphone (client) to PC (server) through socket communication. The QWERTY keyboard, which is the most common modern-day keyboard layout, consists of XML. Each button has an independent `OnClickListener`, and depending on the state of the keyboard, transmitted values are different.

User-entered key values with the specified protocol ("$") are sent to PC (server). The received values are stored on the PC (server) using `keypress ()` and `keyRelease ()` methods of the `Robot` class in Java.

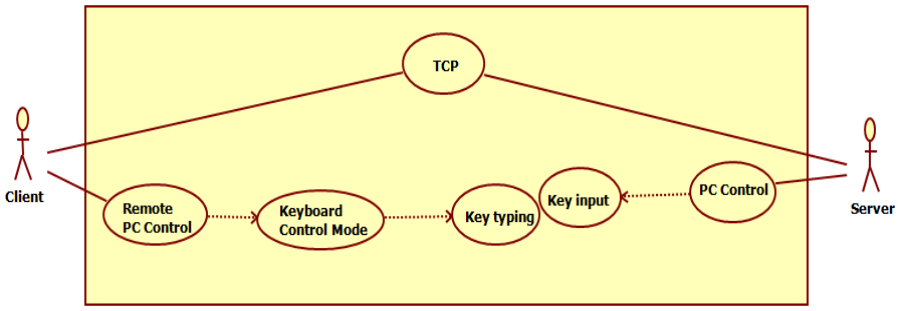

**Fig. 6.** Speech recognition by touching the smartphone screen.

### 4.3. Mouse Control Mode

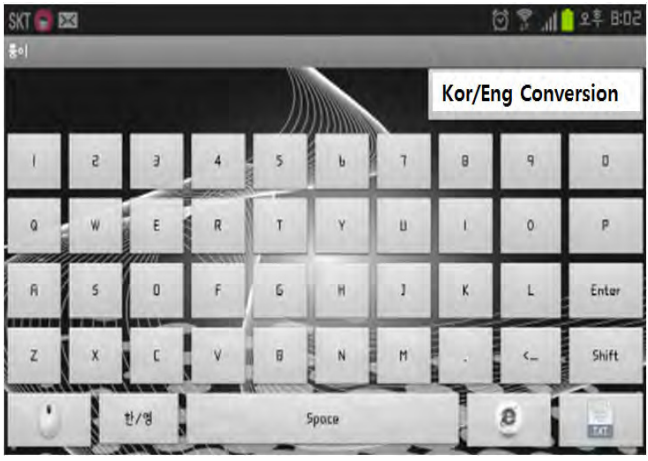
Figure 8 presents computer mouse control by touching the smartphone screen. There are double click, left click, and right click buttons. In order to control the mouse, using the touch screen of the smartphone (client), with UDP, the remote computer control system transmits the first coordinate and an actuated coordinate. In case of the mouse control, with UDP, speed rather than accuracy is prioritized because the system has to quickly transmit data. Using the `mouseMove ()` method of the `Robot` class in Java, the system remotely controls user's PC mouse pointer on PC (server) that was received the transmitted coordinates.

### 4.4. Simple Mode

Execution of applications users want on the simple mode is shown in Figure 9. While using a computer, there are programs that you often use, such as explorer, notepad, Hangul (Korean) word processor, GOM Player, and messenger. The Simple mode is the mode of execution that these programs are executed with a single click from a remote location. When the button is clicked on smartphone (client), the commands will be sent to the PC (server) through TCP communication. Using the `exec ()` method of the `Runtime` class in Java, with the touch of a button, the program that you want will be easily executed on PC received the commands through the external command.




(a) Use case diagram of keyboard control

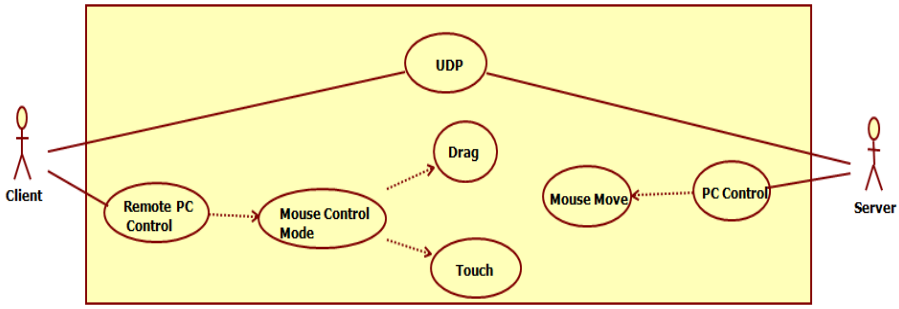



(b) Screenshot of keyboard control

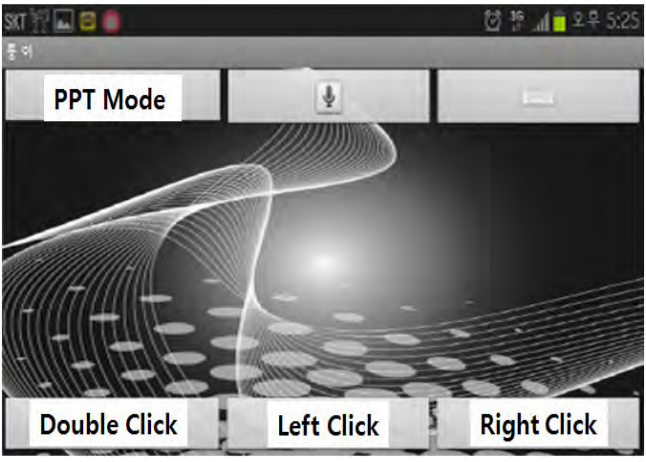
**Fig. 7.** Computer keyboard control by touching the smartphone screen.

### 4.5. Sending a Text Message Mode

The existing service method, which has transmitted texts through voice, does not read texts entered by the user and send back to the user. The proposed system, however, using the STT technology, provides the function that can correctly deliver the information since when the user inputs his/her voice on smartphone, it re-reads what you enter through the TTS function. When you have made all your input through `SmsManager`, the system sends a text message to the other party; and also see that Figure 10 shows a flowchart how to send a text message.




(a) Use case diagram of mouse control

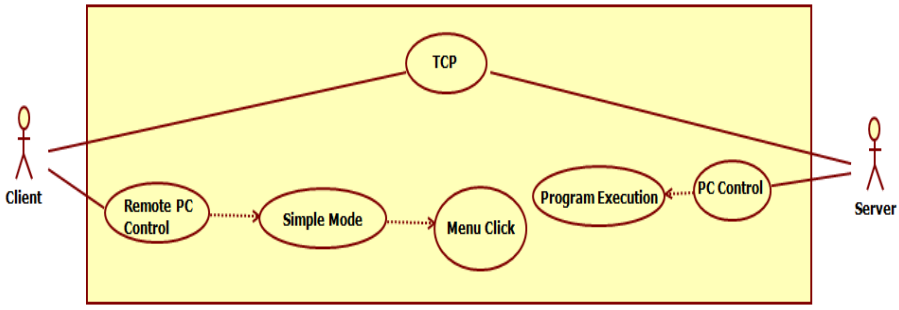



(b) Screenshot of mouse control

**Fig. 8.** Computer mouse control by touching the smartphone screen. There are double click, left click, and right click buttons.

### 4.6. Other Modes: File Search Mode

File search function is the ability to look at contents in the hard drive of the PC Server on smartphone. When smartphone users (Client) request a list of files in the PC (server), using the File class in Java, the proposed system distinguishes files and folders, and sends the list to the smartphone. This list with the folders and files shows on the smartphone screen through `ListView`. When the user clicks a folder, its contents shows in `ListView`. When the user clicks a file, the file is run through the `exec ()` method of `Robot` class on the PC Server. For example, when requesting to send `test.pdf` file from your smartphone, the `test.pdf` file, which is sent to your smartphone, can be found.




(a) Use case diagram of simple mode


(b) Screenshot of simple mode

**Fig. 9.** Execution of applications that users want to on the simple mode.

### 4.7. Other Modes: PC Screen Check Mode

When smartphone users request the transfer of your PC screen, the proposed system captures the current screen using the Robot class on PC, and transmits the screen to the smartphone through TCP communication. The smartphone receives the file and shows it on the `ImageView` screen. The multi-touch is possible, zooming in and out is feasible, and the system can check what the current PC's screen is. Commands with speech recognition are available, and a remote control mode in real time is possible by making sure the PC's screen.

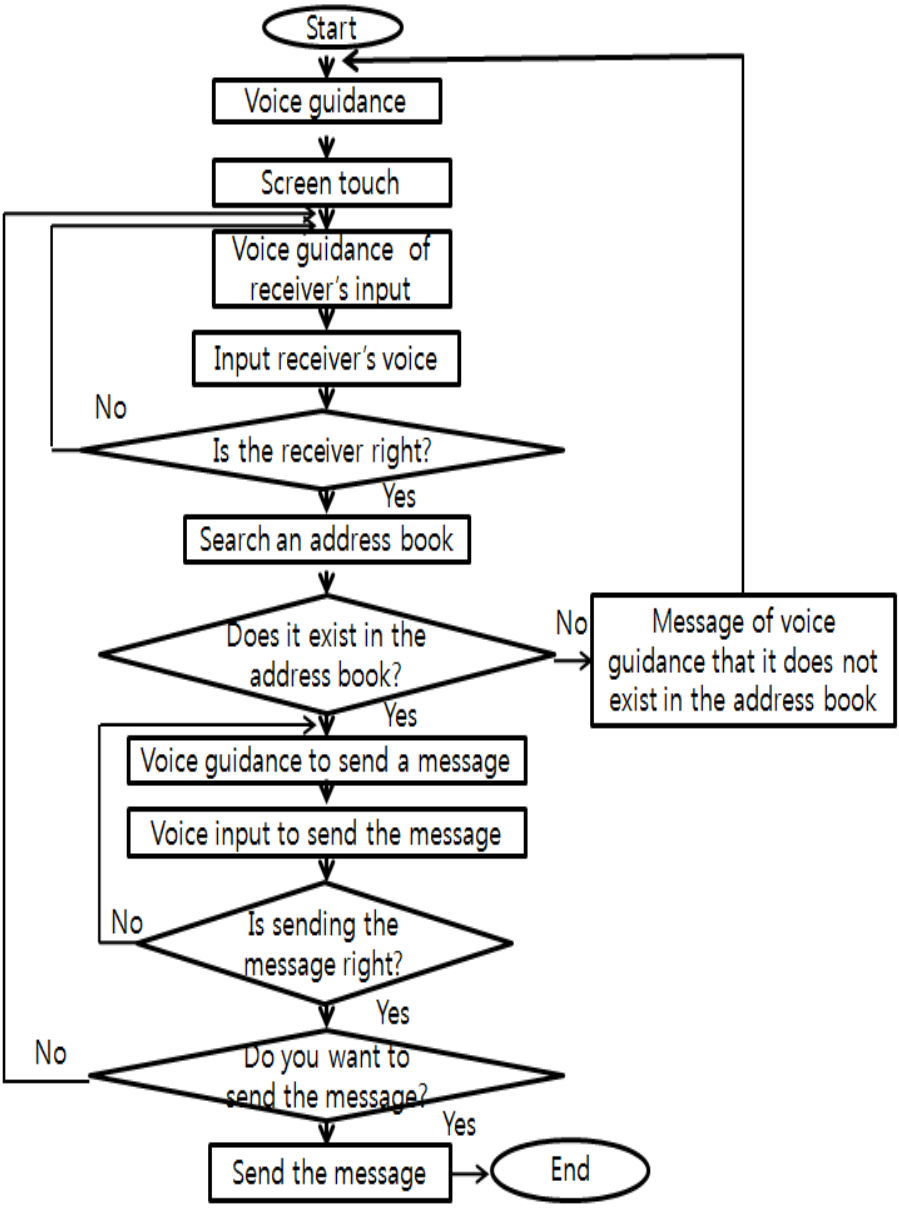



**Fig. 10.** Flowchart of sending a text message.

## 5. Conclusion

A computer remote control system using speech recognition technologies of mobile devices and wireless communication technologies for the blind and physically disabled population has been proposed. These people experience difficulty and inconvenience in using

computers through a keyboard and/or mouse. The major purpose of this system was to provide a system so that the blind and physically disabled population can easily control many functions of a computer via speech. The system is very useful for the general population as well. Users command a mobile device to do something via speech such as directly controlling computers, writing emails and documents, calculating numbers, checking the weather forecast, or managing the schedule. These commands are then immediately executed. The proposed system also provides blind people with a function via TTS (text to speech) of the Google server if they want to receive contents of a document stored in a computer.

**Acknowledgments.** The authors would like to give thanks to the funding agencies for providing financial support. Parts of this work were supported by a research grant from Korean Bible University. The authors also thank Robert Hotchkiss and three referees for their constructive remarks and valuable comments.

## References

1. Agarwal, A., Wardhan, K., Mehta, P.: JEEVES - A Natural Language Processing Application for Android. http://www.slideshare.net (2012)
2. Aguero, J., Rebollo, M., Carrascosa, C., Julian, V.: Does Android Dream with Intelligent Agents? Advances in Soft Computing 50, 194–204 (2009)
3. Android: Android Operating System, Wikipedia. http://en.wikipedia.org/wiki/Android OS
4. Jarng, S.S.: Analysis of HMM Voice Recognition Algorithm. Journal of Advanced Engineering and Technology 3(3), 241–249 (2010)
5. Jeong, H.D., Lim, J., Hyun, W., An, A.: A Real-time Location-based SNS Smartphone Application for the Disabled Population. Computer Science and Information Systems (ComSIS) 10(2), 747–765 (2013)
6. Jeong, H.D., Ye, S.K., Lim, J., You, I., Hyun, W., Song, H.K.: A Remote Computer Control System Using Speech Recognition Technologies of Mobile Devices. In: The Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing: Future Internet and Next Generation Networks (FINGNet-2013). pp. 595–600. Taichung, Taiwan (2013)
7. Knight, W.: Where Speech Recognition Is Going. MIT Technology Review, technologyreview.com (2012)
8. Korea Creative Contents Agency: Trends and Prospects of Speech Recognition Technologies (2011)
9. Lee, C.Y., An, B., Ahn, H.Y.: Android based Local SNS. Institute of Webcating, Internet Television and Telecommunication 10(6), 93–98 (2010)
10. Mao, Q.R., Zhan, Y.Z.: A Novel Hierarchical Speech Emotion Recognition Method Based on Improved DDAGSVM. Computer Science and Information Systems (ComSIS) 7(1), 211–221 (2010)
11. Rabiner, L., Juang, B.: An Introduction to Hidden Markov Models. IEEE ASSP Magazine pp. 4–16 (1986)
12. Tan, Z.H., Varga, I.: Network, Distributed and Embedded Speech Recognition: An Overview. Advances in Patterns Recognition (2008)
13. Torrente, J., Á.d. Blanco, Á. Serrano-Laguna, Vallejo-Pinto, J., Moreno-Ger, P., Fernández-Manjón, B.: Towards a Low Cost Adaptation of Educational Games for People with Disabilities. Computer Science and Information Systems (ComSIS) 11(1), 369–391 (2014)

14. UNESCO: Arirang, lyrical folk song in the Republic of Korea. http://www.unesco.org/culture/ich/RL/00445
15. Wikipedia: http://en.wikipedia.org/wiki/Speech_recognition

**Hae-Duck Joshua Jeong** is an associate professor in the Department of Computer Software at Korean Bible University, Seoul, South Korea. He received his Ph.D. in Computer Science and Software Engineering from the University of Canterbury, New Zealand. He is the author or co-author of more than 85 research publications, including more than twenty patents. Dr. Jeong is on the editorial board and a reviewer for various domestic and international journals. He is the corresponding guest editor or guest editor of ComSIS and MCM. His research interests include teletraffic modeling, stochastic simulation, multimedia telecommunication networks, intrusion detection system, social networking service, and real-time system. Member of IEEE NZ, KIPS, KSII, and ORSNZ.

**Sang-Kug Ye** is a student in the Department of Computer Software at Korean Bible University, Seoul, South Korea. He is the author or co-author of more than 5 research publications, including two patents. He received the President's Computer Software Competition Award four times including the grand prize twice from Korean Bible University from 2011 to 2013. His research interests include network security, intrusion detection system, and mobile applications.

**Jiyoung Lim** received her Ph.D. degree in Computer Science and Engineering from Ewha Womans University in 2001. She is currently an associate professor of Computer Software at Korean Bible University, Seoul, South Korea. Her research interests include wireless/sensor network security, and M2M network security.

**Ilsun You** received his Ph.D. degree in Computer Science from Dankook University, Seoul, South Korea in 2002. Also, he obtained his second Ph.D. degree from Kyushu University, Japan in 2012. In 2005, he joined Korean Bible University, South Korea as a full time lecturer, and he is now working as an associate professor. Dr. You is on the editorial board for various domestic and international journals. Also, he has served as a guest editor of several journals. His main research interests include Internet security, authentication, access control, formal security analysis, MIPv6, and ubiquitous computing.

**Wooseok Hyun** is the corresponding author of this paper. She is an associate professor in Computer Software at Korean Bible University, Seoul, South Korea. She received her Ph.D. in Computer Science from Gyeongsang National University, South Korea. She is the author or co-author of more than 30 research publications, including five patents; reviewer of various domestic and international journals. Her research interests include ubiquitous computing, intelligent system, fuzzy system, information retrieval system, and artificial intelligence. Member of KIISE, KIPS, KMMS.

*Received: September 15, 2013; Accepted: January 21, 2014.*

Computer Science and Information Systems 11(3):1017–1035 DOI: 10.2298/CSIS130926062C

# A New Hybrid Architecture with an Intersection-Based Coverage Algorithm in Wireless Sensor Networks

Young-Long Chen[1], Mu-Yen Chen[2], Fu-Kai Cheung[3], and Yung-Chi Chang[1]

[1] Department of Computer Science and Information Engineering, National Taichung University of Science and Technology, Taichung 404, Taiwan
ylchen66@nutc.edu.tw
[2] Department of Information Management, National Taichung University of Science and Technology, Taichung 404, Taiwan
mychen@nutc.edu.tw
[3] Institute of Communication Engineering, National Yunlin University of Science and Technology, Yunlin 640, Taiwan
jackalrice@hotmail.com

**Abstract.** Energy is limited in wireless sensor networks (WSNs) so that energy consumption is very important. In this paper, we propose a hybrid architecture based on power-efficient gathering in sensor information system (PEGASIS) and low-energy adaptive clustering hierarchy (LEACH). This architecture can achieve an average distribution of energy loads, and reduced energy consumption in transmission. To further extend the system lifetime, we combine the intersection-based coverage algorithm (IBCA) with LEACH architecture and the hybrid architecture to prolong the system lifetime that introducing sensor nodes to enter sleep mode when inactive. This step can save more energy consumption. Simulation results show that the performance of our proposed LEACH architecture with IBCA and the hybrid architecture with IBCA perform better than LEACH architecture with PBCA in terms of energy efficiency, surviving nodes and sensing areas.

**Keywords:** Hybrid architecture, PEGASIS, LEACH, System lifetime, IBCA, PBCA.

## 1. Introduction

Wireless sensor networks (WSNs) [1] are typically consist of a large number of resource-constrained sensor nodes which are randomly scattered to collect environmental condition data from an area [2], [3], such as humidity, temperature, solar radiation, concentration of carbon dioxide [4], and risky places for humans. Each sensor node is able to independently manage operations [5] and deliver the data to a base station (BS) by radio wave, infrared rays, or optical fiber transmission [6]. The lifetime of a node is dependent on its energy consumption rate in WSNs. Because the sensors are usually located in dangerous or inaccessible areas, and the battery cannot be replaced. The control of energy efficiency is a primary issue [7], [8], [9] in WSNs.

A Low-energy adaptive clustering hierarchy protocol (LEACH) for a distributed network proposed by Heinzelman et al. [10]. Sensor nodes are divided into several clusters in the LEACH architecture. Each cluster choosing a sensor node as the cluster head (CH), which delivers an advertisement (ADV) to every neighboring sensor node in accordance with carrier-sense multiple access (CSMA) [11], [12], [13] protocol in the MAC layer. Each sensor node identifies itself as cluster member with the CH [14], [15] through the strength of the ADV message. Cluster members transmit data to the CH using the time division multiple accesses (TDMA) [12], [13] mechanism. This TDMA protocol enables mechanisms to avoid and resolve collisions, because TDMA has a designated time slot for each node in which only that particular node transmits. Finally, the CH transmits the data to the BS which performs data aggregation.

In power-efficient gathering in sensor information system (PEGASIS) architecture [16], the nodes transmit data to the next closest neighboring nodes, and receive data from previous closest neighboring nodes, with all of the sensor nodes based on the greedy algorithm, to form a connected chain; each node collects from the previous node and transmits to the next one, until the node closest to the BS, called the chain header, is reached; it performs data aggregation and transmission to the BS [16], [17], [18]. The role of chain header for each round will rotate between nodes.

There are two problems with LEACH. First, there is an excess of CHs. Second, the distance between the CHs and cluster nodes is too great. However, PEGASIS architecture also has problems. In the first place, the transmission distance may not be the shortest between sensing nodes. Next, the aggregated number of data packets is larger than in LEACH architecture. To overcome these problems, we developed a combined architecture based mainly on PEGASIS, but utilizing the advantages of the LEACH structure.

In WSNs, sensor nodes are randomly deployed, as a result, the sensor field is uneven density, and this will cause coverage issues [19]. If a node's sensing range is overlapped, it is called a redundant node. The redundant node is entered into sleep mode, which does not affect the overall sensor field or connectivity. Sleep mode is a way to prolong the network lifetime. In order to upgrade LEACH system performance, there is a phase-based coverage algorithm (PBCA) [20], [21] for locating all redundant nodes. The results of LEACH architecture with PBCA show that provides excellent system efficiency compared with the original LEACH architecture. On the other hand, an intersection-based coverage algorithm (IBCA) is proposed in [22], [23], that is capable of locating all redundant nodes. The number of redundant nodes of IBCA will be larger than that of PBCA. Thus, IBCA can improve the network lifetime more than PBCA.

In order to achieve better network performance, we propose novel schemes involving the application of IBCA to the LEACH [24] and hybrid architectures in this paper. Our schemes can identify redundant nodes to improve energy consumption.

The rest of this paper is organized as follows: Section 2 describes the IBCA. In Section 3, we combine of IBCA and the LEACH architecture. Section 4 introduces our proposed hybrid topology architecture, and the combination of IBCA with the hybrid architecture. In Section 5, we simulate and analyze LEACH, LEACH with PBCA, LEACH with IBCA, and the hybrid architecture with IBCA that compare the performance of four architectures. The last section is the conclusion of this paper.

## 2. Related Work

PBCA [20], [21] is employed as a criterion to determine whether the target node is k-coverage. PBCA is utilized to identify whether the sensing range of target node is fully covered by neighbor nodes with a time complexity $O(N \log N)$, where $N$ represents the number of neighbor nodes around the target sensor node.

IBCA [22], [23] uses intersections to judge if certain sensor nodes can enter sleep mode. Both intersections of a target sensor node's sensing range and other nodes' sensing ranges and intersections on the perimeter of the total sensing range, are k-coverage. The result is that the target sensor node will be k-coverage. This algorithm requires the coordinate information of all sensor nodes and the computational complexity is $O(N^3)$, where $N$ is the number of neighbors of the target sensor node. Figure 1 shows an intersection of two sensor nodes. The target sensor node $p_y$ and the neighboring sensor node $p_x$ will intersect at two points, which are intersection 1 and intersection 2.

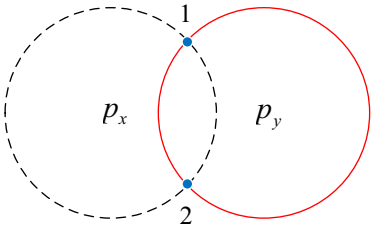



**Fig. 1.** The intersection of sensing ranges

The IBCA has a wider judgment condition ($d < 2R_s$) than the PBCA [20], [21], where $d$ is the distance from a sensor node to the target sensor node, $R_S$ is the sensing range of each sensor node. Therefore, the selected number of redundant nodes will be larger than that of PBCA. The number of redundant nodes that can be used for judgment will be larger. In other words, there are more neighbor node of the target sensor node that can be used for judgment. The calculation will become more complicated, and the implementation time will become longer.

IBCA provide redundancy rules for nodes as shown in [22], [23]. If both conditions have been established, the target node $p_y$ is considered a redundant node. Assuming a set of overlapped neighbor nodes $N(p_y)$, neighbor nodes $p_x \in N(p_y)$, and $d$ is the distance from neighbor node $p_x$ to target node $p_y$.

**Condition 1.** The distance from $p_x$ to $p_y$ should be less than or equal to twice the sensing range $R_S$. We have as follows:

$$d(p_y, p_x) \le 2R_S, \forall p_x \in N(p_y) \quad \textbf{(1)}$$

**Condition 2.** The intersections are generated by node $p_y$ and its neighbor nodes $p_x$.The target node $p_y$ must check all the intersections within its sensing range. Each intersection overlaps coverage by each neighbor node's sensing range at least once.

It is a redundant node which can enter sleep mode which intersections of a sensor node $p_y$ are covered by other sensor nodes' sensing ranges.

LEACH has two point problems. One is generate the excess of CHs. Another is distance between the CHs and cluster nodes. However, PEGASIS architecture has one problem that the aggregated number of data packets is larger than in LEACH architecture. To improve above problems, we proposed a combined architecture mainly based on PEGASIS, but utilizing the advantages of the LEACH architecture.

Our proposed architecture has two advantages. One is using PBCA to identify the sensing range of target node is fully covered by neighbor nodes. Another is using IBCA to find the redundancy nodes, and then let these nodes get into sleep mode.

## 3. LEACH Architecture with Intersection-Based Coverage Algorithm

In this section, we combine IBCA with LEACH topology architecture [10] and enter redundant sensor nodes into sleep mode. In this way, we improve system energy efficiency and extend the system lifetime.

### 3.1. The LEACH Architecture

In [10], Heinzelman et al. proposed the LEACH architecture operation. Each cluster member delivers data directly to the cluster head, rather than to the distant base station. As a result, the energy consumed by the cluster members is merely the amount required during data transmission between cluster members and the cluster head, although the cluster head requires a larger amount of energy to perform data aggregation and implement data transmission to the base station. It should be noted that in LEACH architecture, the system is composed of variable clusters for each round.

LEACH protocol is mainly divided into two phases. The first is the setup phase; there is a probability $P_i(t)$ that each sensor node will be specified as the cluster head in the initial round. The average expected value of the cluster head number is given by:

$$E[CH] = \sum_{i=1}^{K} P_i(t) \times 1 = \alpha \qquad (2)$$

Where $E[CH]$ is the average expected value of the number of cluster heads, $K$ is the total deployed number of sensor nodes in a WSN, $P_i(t)$ is the probability that node $i$ will decide it is to become a cluster head at time $t$ and $\alpha$ is the cluster quantity.

In order to prevent nodes serving as a cluster head in consecutive rounds, (3) determines the probability that each node will become the cluster head:

$$P_i(t) = \begin{cases} \dfrac{\alpha}{K - \alpha \times (r \mod \dfrac{K}{\alpha})} & , C_i(t) = 1 \\ 0 & , C_i(t) = 0 \end{cases} \quad (3)$$

Where $r$ is the current round and $t$ is increased by unity in the event that all the cluster members have acted as the head. $C_i(t) = 0$ indicates that node $i$ has been the cluster head this round. $C_i(t) = 1$ indicates that a node has not yet been the cluster head this round. LEACH architecture requires that each cluster member serve as the head once for each $K / \alpha$ round. Each node $i$ should choose to become a cluster head with probability $P_i(t)$ at round $r$.

### 3.2. The LEACH Architecture with IBCA

To extend the system lifetime, we combine the LEACH architecture with IBCA [24]. First, we use IBCA to find the redundant sensor nodes whose intersections are covered by other nodes' sensing ranges. Let these redundant sensor nodes enter sleep mode in order to reduce the energy consumption of the WSN. The remaining nodes of the WSN then form LEACH architecture. The flow chart is shown in Fig. 2.

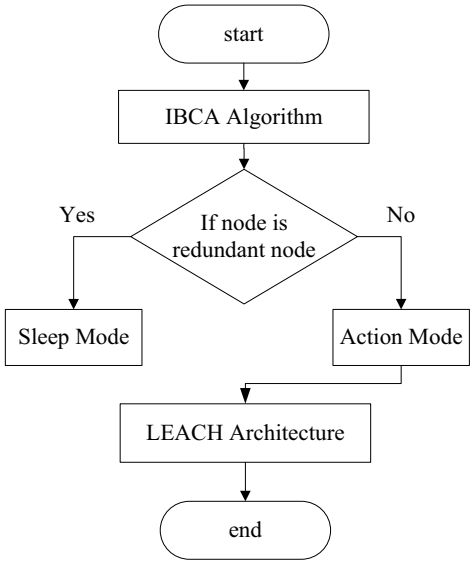



**Fig. 2.** The flow chart of LEACH architecture with IBCA

# 4. An Intersection-Based Coverage Algorithm for the Novel Hybrid Architecture

In this paper, we propose novel hybrid architecture with IBCA in order to obtain the advantages of both the PEGASIS and LEACH architectures. The hybrid architecture combined with IBCA improves the system energy efficiency and extends the system lifetime.

## 4.1. Our Novel Hybrid Architecture

In our proposed method, the hybrid architecture appoints a sensor node nearest the BS in each round, called the leader. The leader is responsible for the final aggregated data transmission to the BS. The leader is selected in turn from all the sensor nodes, so as to prevent the premature death of any one leader. When all of the sensor nodes have served as the leader, a new round begins.

The BS implements algorithms and receives data from sensor nodes. It has a leader list $L$ which records the status of all the sensor nodes that have served as the leader. Those nodes that have been appointed leader this round and dead nodes will be removed from this list. After deleting the deceased nodes and former leaders, the BS designates the node closest to the BS as the current leader. If two or more sensor nodes are the same distance from the base station, the node with the highest energy will serve as leader. The purpose of this selection is to update leaders; the leader list composed of all active nodes is updated after all the nodes have served as leader once each round.

After determining the leader, our new network architecture begins operating. The leader is defined as the only member in level 0. To illustrate this idea by levels, the members of level $k$ are generated by members in level $k-1$. For example, the members in level 1 are determined by the ADV message in level 0. Define a set $C_k$ composed of all the members in level $k$. The $m^{th}$ member in level $k$ is $p_k^m$, and $N(p_k^m)$ are all the non-level sensor nodes regarding $p_k^m$. Therefore, $p_0^1$ denotes the leader. Assuming the current level is level $i$, all the members of level $i$ will send an ADV message consisting of an ID and a level number to members not of that level. Based on the strength of the received ADV message, members then determine if they belong to $N(p_k^m)$.

To show the details of how the transmission path is determined, define $p_i$ as the $i^{th}$ non-level node. The distance between nodes $p_i$ and $p_j$ is $d(p_i, p_j)$. The non-level nodes merely receive the ADV message to be sent by $p_k^m$.

As long as either of the criteria, which are $d(p_j, p_k^m) < d(p_i, p_k^m)$ or $d(p_j, p_k^m) < d(p_i, p_j)$, are satisfied, these nodes will become members of the next level.

For example, assume all members of level $k$ have been found. Next, confirm a member from all the level $k+1$ members, which are known as pertaining to members in level $k$. Given two nodes $p_i$ and $p_j$, both $p_i$ and $p_j$ belong to $N(p_k^m)$. Based on the

above criteria, one of the conditions is established, then $p_j$ is identified as a member in level $k+1$. It is clear that node $p_j$ is closer to $p_k^m$ than the other node.

As shown in Fig. 3, $p_k^m$ represents the $m^{th}$ member in level $k$, and is intended to locate the members belonging to level $k+1$ from among nodes $p_a$, $p_b$, $p_c$, $p_d$ and $p_e$. Since all these non-level nodes merely receive the ADV message sent by $p_k^m$, nodes $p_a$, $p_b$, $p_c$, $p_d$ and $p_e$ all pertain to $N(p_k^m)$, that is, $N(p_k^m)=\{p_a, p_b, p_c, p_d, p_e\}$. According to the time interval between the transmission and reception of the ADV message, it is discovered that $d(p_c, p_k^m)$, as well as $d(p_d, p_k^m) < d(p_e, p_k^m)$, and then $d(p_e, p_c)$, as well as $d(p_e, p_d) > d(p_e, p_k^m)$. For the first criterion, $d(p_d, p_k^m) < d(p_e, p_k^m)$ is satisfied; then node $p_d$ is identified as a member of level $k+1$ in the same manner. For the second criterion, $d(p_e, p_d) > d(p_e, p_k^m)$ is satisfied, and then $p_e$ is identified as a member of level $k+1$ in the same manner.

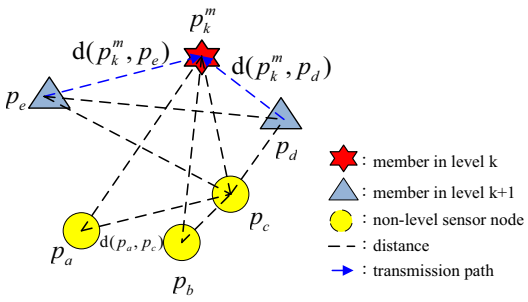



**Fig. 3.** Determination of data transmission path between nodes

Applying such criteria to node $p_a$, we find that $d(p_c, p_k^m)$, $d(p_d, p_k^m)$, $d(p_e, p_d) < d(p_a, p_k^m)$ and then $d(p_a, p_c) < d(p_a, p_k^m)$. It is concluded that $p_a$ does not pertain to level $k+1$, for neither of the two criteria are satisfied. Similarly, $p_b$ and $p_c$ are both identified as non-members of level $k+1$.

Fig. 4 illustrates the node distribution in which $p_b$ is the leader, the only member in level 0, while nodes $p_a$, $p_c$, $p_e$ and $p_h$ are members, located by $p_b = p_0^1$, in level 1. Likewise, nodes ($p_f$, $p_d$, $p_g$, $p_i$, and $p_l$), ($p_j$, $p_k$, $p_m$, $p_o$, and $p_n$), ($p_r$, $p_q$, and $p_p$) and $p_s$ are members in levels 2, 3, 4 and 5, respectively. The absence of a linkage message in response to the ADV message, sent by level members, represents the completion of the network configuration.

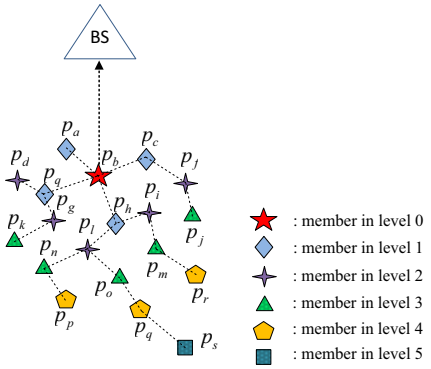



**Fig. 4.** The novel hybrid architecture

### 4.2. The Hybrid Architecture with IBCA

In this section, we propose the application of the IBCA [22], [23] to our hybrid architecture. The complete IBCA as given by:

```
program IBCA algorithm
  Parameter definition:
    S as the set of the sensor nodes entered into sleep
      mode;
    A as the set of the sensor nodes in active mode;
    L as the set of live sensor nodes;
    Node as the number of live sensor nodes;
    N(p_i) as the set of neighbor nodes of the target
      sensor node;
    D(p_i) as the set of the intersections of p_i and N(p_i)
      where D(p_i) must be overlapped by other sensor node
      at least once;
  begin
    All the live nodes belong to A;
    repeat
      Locate N(p_i), where p_i ∈ L, N(p_i) = {p_j | d(p_i, p_j) ≦ 2R_s},
      p_i ≠ p_j, p_j ∈ L, and p_j ∈ A;
      Compute the intersections of p_i and neighbor nodes,
      where p_j ∈ N(p_i);
      if p_i is permitted to enter sleep mode.
        D(p_i) is an empty set;
        p_i is a redundant node permitted to sleep;
        p_i ∈ S;
```

```
        else
           p_i remain active;
           p_i ∈ A ;
        end if
        i++;
      until i = Node
end.
```

First, we use the IBCA to find the redundant sensor nodes. Next, these redundant sensor nodes enter sleep mode in order to reduce the energy consumption of the WSN. We then set up the hybrid architecture with the active sensor nodes using IBCA. In our algorithm, all the sensor nodes can be divided into two modes: active mode or sleep mode. Therefore, the fewer nodes that are active, the lower the energy consumption for each round will be.

In our hybrid architecture, each sensor node has the functions of data detection, collection and transmission. A sensor node will send its own data to its neighbor node closest to the BS. The neighbor nodes will then perform data collection on the received data and their own detected data. After fusing the data, they will transmit the fused data to the neighbor node closest to the BS. This process will repeat until the data reaches a node with no sensor node closer to the BS than itself, and complete data will then be transmitted to the BS.

In order to extend the network lifetime, the IBCA for the hybrid architecture can be divided into three steps: active node selection, leader node selection and hybrid architecture.

**Active node selection.** The system will select sensor nodes to be active nodes using the IBCA, and instruct redundant sensor nodes to enter sleep mode.

**Leader node selection.** The system will compare the residual energy of each sensor node in the hybrid architecture and select the active sensor node with the maximum remaining energy as the leader.

**Implementing hybrid architecture.** The members of levels should send an ADV message to non-level nodes, and then a linkage message is sent to the remainder of those non-level nodes. The ADV messages of the linkage contain their identification, the members pertaining to the next level and the reception time. There are two criteria for judging if the condition is satisfied for them to be members of the next level. This ADV message is a small message which includes the node's ID and a header, defined as an announcement message. This determines which non-cluster head node must join which cluster in this round, according to the signal strength of the ADV message from each CH.

Once members of the level have sent the ADV message until it does not contain any linkage message, the WSN configuration is complete. Fig. 5 shows the flow chart of the hybrid architecture with IBCA.

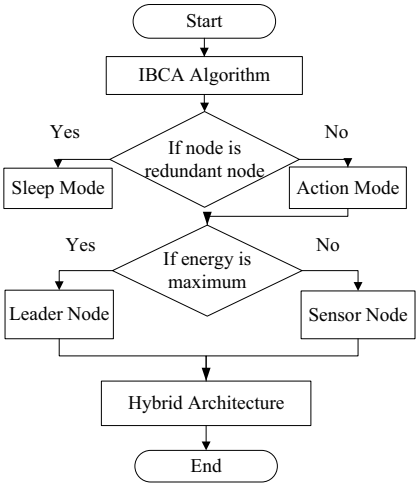



**Fig. 5.** The flow chart of our hybrid architecture with IBCA

## 5. Simulation Results

We propose LEACH with IBCA, and hybrid architecture with IBCA in this paper. In this section, we use C programming language to simulate and compare the system efficiencies, which are LEACH architecture [10], LEACH architecture with PBCA [21], LEACH architecture with IBCA and the hybrid architecture with IBCA. We compare the system efficiencies, which are based on total residual energy, number of deceased nodes and total sensing area. The simulation assumed 100 sensor nodes randomly distributed over a 100 meter by 100 meter field, a sensing radius $R_S$ of 15 meters and BS located at (50, 150), with the parameters specified in Table 1. From [10], [25], the optimal cluster number $k_{opt}$ is employed in the LEACH architecture. The transmission for energy consumption is given by:

$$E_{Tx}(l,d) = \begin{cases} l \times (E_{elec} + \varepsilon_{fs} \times d^2), & d < d_0 \\ l \times (E_{elec} + \varepsilon_{mp} \times d^4), & d \geq d_0 \end{cases} \tag{4}$$

Where $E_{Tx}(l,d)$ is the transmission of energy that packet size is $l$ and the distance is $d$ ; $E_{elec}$ is the electronics energy; $d_0$ is distance threshold; $\varepsilon_{fs}$ and $\varepsilon_{mp}$ represent the amplifier energy of free space and the amplifier energy of multi-path, respectively. The receiving for energy consumption $E_{Rx}(l)$ is defined as:

$$E_{Rx}(l) = l \times E_{elec} \tag{5}$$

Figure 6 shows the total residual energy each round for four architectures. Our novel hybrid architecture with IBCA and LEACH architecture with IBCA obtained a large

total residual energy in each round. The original LEACH architecture has the lowest residual energy, owing to the absence of IBCA. In addition, the LEACH architecture with IBCA has greater residual energy than does LEACH architecture with PBCA because IBCA can select more sensor nodes to enter sleep mode than PBCA can. The main reason for this is that the number of sensor nodes that can be used for judgment will become greater. Finally, the reason for the largest total residual energy is that our hybrid architecture ensures a minimum transmission distance between nodes.

**Table 1.** Simulation Parameter

| Variables | Value |
|---|---|
| Initial energy | $E_{init} = 0.25$ J |
| BS location | (50,150) |
| Number of package | $l = 4000$ bits |
| Electronic energy | $E_{elec} = 50$ nJ/bit |
| Energy consumed in data fusion | $E_{DA} = 5$ nJ/bit/signal |
| Amplifier energy of free space | $\varepsilon_{fs} = 10$ pJ/bit/m$^2$ |
| Amplifier energy of multi-path | $\varepsilon_{mp} = 0.0013$ pJ/bit/m$^4$ |

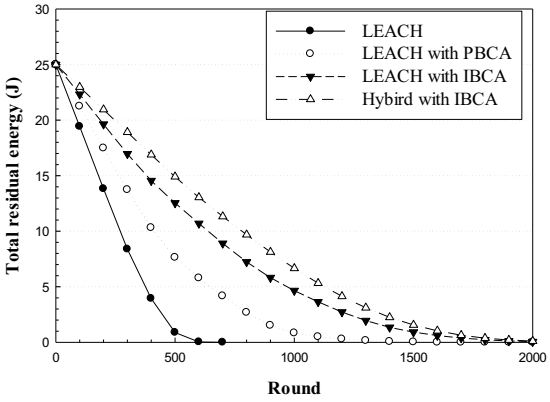



**Fig. 6.** Comparison of total residual energy each round

The number of deceased nodes each round for four architectures are shown in Fig. 7. In the original LEACH architecture, the first node dies at about round 290, and the 20th node dies at about round 354. In the LEACH architecture with PBCA, the first node dies at about round 301, and the 20th node dies at about round 397, while in LEACH architecture with IBCA, nodes die at rounds 315 and 409, respectively. However, the first node perishes at about round 372, and the 20th at round 537 in the hybrid architecture with IBCA. Therefore, it is demonstrated that our hybrid architecture with

IBCA has the lowest sensor node mortality rate, while the original LEACH architecture has the highest.

Total sensing area is shown in Fig. 8. Prior to round 300, the four architectures cover the same sensing area. After round 500, the hybrid architecture with IBCA maintains the largest sensing area, and the original LEACH the smallest. Thus, it is shown that for sensing area each round, the hybrid architecture with IBCA covers the maximum sensing area, while the original LEACH covers the minimum. The main reason for this is that the hybrid architecture with IBCA has an extended system lifetime.

In addition, we changed the sensing radius to, and compared the performances of the three WSN architectures in Figs. 9-11.

As shown in Fig. 9, the total residual energy of the original LEACH was depleted at round 700. The residual energy of LEACH with PBCA was 6.6 J at round 700, and the residual energy of LEACH with IBCA was 11 J at round 700. However, the residual energy of the hybrid architecture with IBCA was 13 J at round 700. Our novel scheme, the hybrid architecture with IBCA, and LEACH with IBCA obtain greater total residual energy than the other architectures.

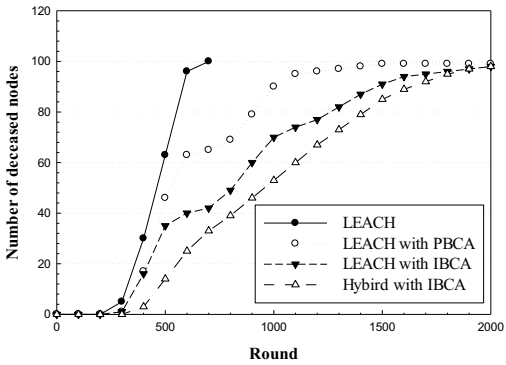



**Fig. 7.** Comparison of the number of deceased nodes each round

As can be seen in Fig. 10, the first node dies at about round 292, and the 20th node dies at about round 365 in the original LEACH architecture. In the LEACH with PBCA architecture, the first node dies at about round 311, and the 20th node dies at about round 406. In the LEACH with IBCA architecture, this does not occur until approximately rounds 320, and 418, respectively. However, in our novel scheme, the hybrid architecture with IBCA, the first node dies at about round 370, and the 20th at round 587. Therefore, it is demonstrated that our hybrid architecture with IBCA extends node lifespan more than other architectures do.

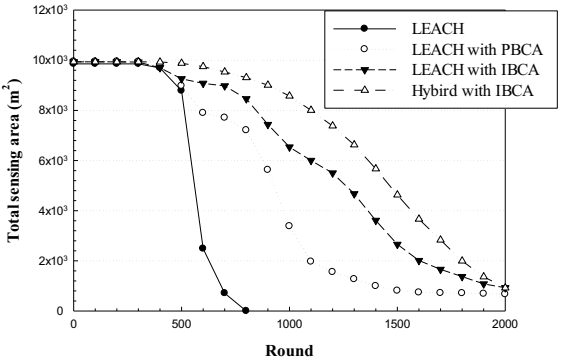



**Fig. 8.** Comparison of total sensing area each round.

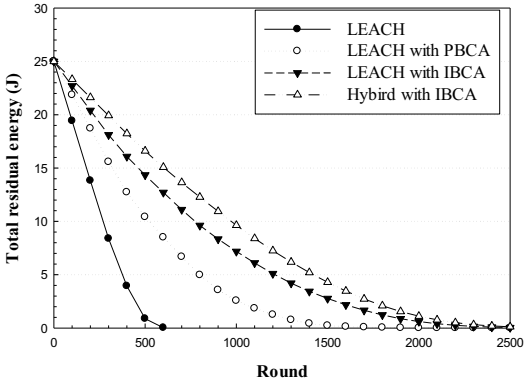



**Fig. 9.** Total residual energy for $R_s = 18$ m

As demonstrated in Fig. 11, subsequent to round 580, our hybrid architecture with IBCA maintains the largest sensing area, while the original LEACH maintains the smallest. Prior to round 410, the hybrid architecture with IBCA covers the same sensing area as do LEACH with PBCA and LEACH with IBCA, which is larger than that covered by the LEACH architecture alone.

Furthermore, we assume that each node has a sensing radius of $R_s = 15$ m , and number of bits $l = 6000$ bits . We compare the WSN performance of the three architectures in Figs. 12-14.

As shown in Fig. 12, the original LEACH runs out of energy at round 500, the LEACH with PBCA has 3.4 J at round 500, and the LEACH with IBCA has 8.1 J at round 500. The hybrid architecture with IBCA has 10.5 J at round 500. Our novel scheme, the hybrid architecture with IBCA, obtains the maximum total residual energy.

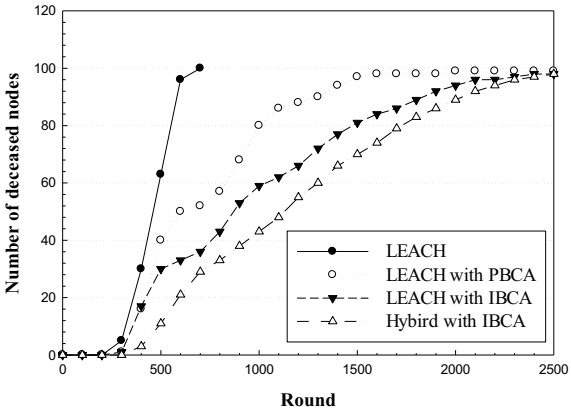



**Fig. 10.** Number of deceased nodes for $R_s = 18$ m

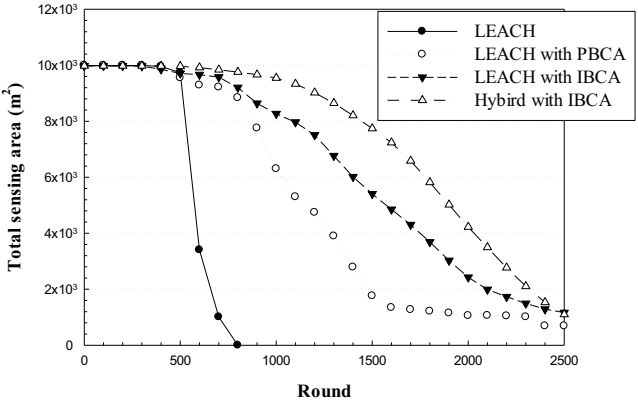



**Fig. 11.** Total sensing area for $R_s = 18$ m

As can be seen from Fig. 13, the first node dies at about round 188, and the 20th node dies at about round 218 for the original LEACH, while in the LEACH with PBCA, the first and 20th nodes die at about rounds 200, 254, respectively. In the LEACH with IBCA architecture, nodes die at about rounds 203, 277, respectively. However, in the hybrid architecture with IBCA, the first node dies at about round 226, and the 20th at about round 371. It is demonstrated that the hybrid architecture with IBCA has extended the nodes' lifetime more than the original LEACH, the LEACH with PBCA, and the LEACH architecture with IBCA have.

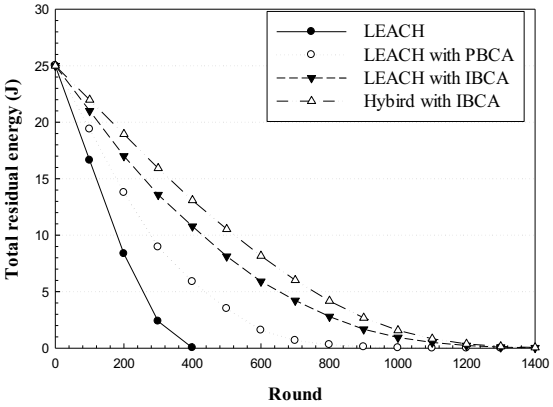



**Fig. 12.** Total residual energy for $R_s = 15$ m and $l = 6000$ bits .

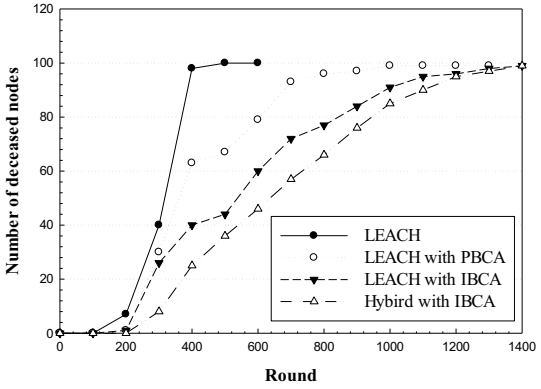



**Fig. 13.** Number of deceased nodes for $R_s = 15$ m and $l = 6000$ bits .

Fig. 14 shows that the hybrid architecture with IBCA maintains the largest sensing area subsequent to round 320, while the original LEACH architecture maintains the smallest.

As shown in Figs. 6-14, we ran simulations for different packets and different node sensing ranges. Notwithstanding these different input parameters, we obtained the same results for every simulation. Therefore, our proposed algorithms consistently outperform the original LEACH architecture and the LEACH architecture with PBCA, in terms of energy consumption, number of live nodes and sensing area.

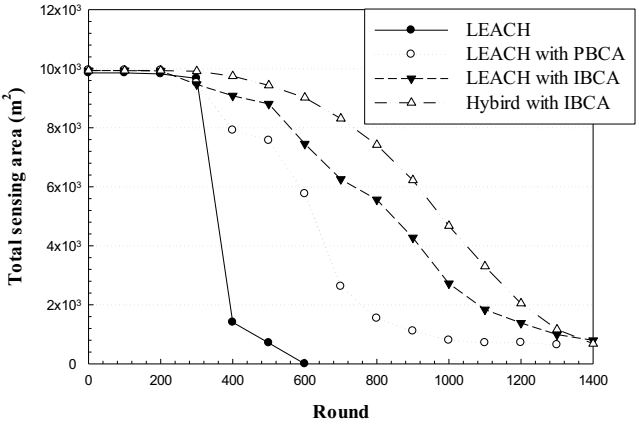



**Fig. 14.** Total sensing area for $R_s = 15$ m and $l = 6000$ bits

## 6. Conclusions

In this paper, we proposed a novel hybrid architecture with IBCA, and LEACH architecture with IBCA. Our novel hybrid architecture is based on both PEGASIS and LEACH architecture, with the four features that (a) it requires that only one leader perform data aggregation and transmit data to the BS for each round, (b) each sensor node must serve as the leader once each round, (c) the minimum transmission distance between nodes is guaranteed, and (d) it requires less aggregated data packets than does PEGASIS. In the hybrid architecture, the data aggregation task is assigned to a plurality of nodes, decreasing the energy required during the data transmission. Our proposed architectures use IBCA that the WSN's sensor nodes are classified into two types, i.e., active nodes, which responsible for detecting data, and the sleep mode nodes, which remain idle. Therefore, the entire system requires less live sensor nodes to cover a sensor field. The nodes to enter sleep mode are determined using IBCA, and do not perform any functions, which reduces energy consumption. Finally, the system is constructed using only active nodes, further reducing the energy consumption of the WSN. On the other hand, IBCA selected a greater number of redundant nodes than did PBCA, the main reason being that, with IBCA, a greater number of sensor nodes can be used for judgment. For this reason, the application of IBCA to the hybrid architecture improves the system lifetime.

Simulation results show that our proposed hybrid architecture with IBCA, and IBCA combined with LEACH demonstrate excellent performance compared with both the original LEACH architecture and LEACH architecture combined with PBCA, in terms of total residual energy, death rate of sensor nodes and total sensing area.

**Acknowledgments.** This work was supported in part by the National Science Council (NSC) of Republic of China under grant No. NSC102-2221-E-025-001 and NSC103-2221-E-025-010.

## References


1. Al-Karaki, J. N., Ul-Mustafa, R., Kamal, A. E.: Data Aggregation and Routing in Wireless Sensor Networks: Optimal and Heuristic Algorithms. Computer Networks, Vol. 53, No. 7, 945-960. (2009)
2. Burrell, J., Brooke, T., Beckwith, R.: Vineyard Computing: Sensor Networks in Agricultural Production. IEEE Pervasive Computing, Vol. 3, No. 1, 38-45. (2004)
3. Estrin, D., Girod, L., Pottie, G., Srivastava, M.: Instrumenting the World with Wireless Sensor Networks. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing. Salt Lake City, USA, 2033-2036. (2001)
4. Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: A Survey on Sensor Networks. IEEE Communications Magazine, Vol. 40, No. 8, 102-114. (2002)
5. Kumarawadu, P., Dechene, D. J., Luccini, M., Sauer, A.: Algorithms for Node Clustering in Wireless Sensor Networks: A Survey. In Proceedings of the IEEE 4th International Conference on Information and Automation for Sustainability. Colombo, Sri Lanka, 295-300. (2008)
6. Bonnet, P., Gehrke, J., Seshadri, P.: Querying the Physical World. IEEE Personal Communications, Vol. 7, No. 5, 10-15. (2000)
7. Islam, M. J., Islam, M. M., Islam, M. N.: A-sLEACH: An Advanced Solar Aware LEACH Protocol for Energy Efficient Routing in Wireless Sensor Networks. In Proceedings of the IEEE 6th International Conference on Networking. Sainte-Luce, France, 1-4. (2007)
8. Jin, Y., Liu, R., He, X., Huang, Y.: A distributed power management design based on MOST networks. Computer Science and Information Systems, Vol. 8, No. 4, 1097-1115. (2011)
9. Qian1, Q., Shen, X., Chen, H.: An Improved Node Localization Algorithm Based on DV-Hop for Wireless Sensor Networks. Computer Science and Information Systems, Vol. 8, No. 4, 953-972. (2011)
10. Heinzelman, W. B., Chandrakasan, A. P., Balakrishnan, H.: An Application-Specific Protocol Architecture for Wireless Microsensor Metworks. IEEE Transactions on Wireless Communications, Vol. 1, No. 4, 660-670. (2002)
11. Yigitel, M. A., Incel, O. D., Ersoy, C.: Design and Implementation of A QoS-Aware MAC Protocol for Wireless Multimedia Sensor Networks. Computer Communications, Vol. 34, No. 16, 1991-2001. (2011)
12. Guo, P., Jiang, T., Zhang, K.: Novel 2-Hop Coloring Algorithm for Time-Slot Assignment of Newly Deployed Sensor Nodes without ID in Wireless Sensor and Robot Networks. Computer Communications, Vol. 35, No. 9, 1125-1131. (2012)
13. Karkvandi, H. R., Pecht, E., Yadid-Pecht, O.: Effective Lifetime-Aware Routing in Wireless Sensor Networks. IEEE Sensors Journal, Vol. 11, No. 12, 3359-3367. (2011)
14. Huang, B., Hao, F., Zhu, H., Tanabe, Y., Baba, T.: Low-Energy Static Clustering Scheme for Wireless Sensor Network. In Proceedings of the IEEE International Conference on Wireless Communications, Networking and Mobile Computing. Wuhan, China, 1-4. (2006)
15. Wei, B., Hu, H. Y., Fu, W.: An Improved LEACH Protocol for Data Gathering and Aggregation in Wireless Sensor Networks. In Proceedings of the IEEE International Conference on Computer and Electrical Engineering. Phuket, Thailand, 398-401. (2008)
16. Lindsey, S., Raghavendra, C. S.: PEGASIS: Power-Efficient Gathering in Sensor Information Systems. In Proceedings of the IEEE Aerospace Conference. Big Sky, Montana, 1125-1130. (2002)
17. Lindsey, S., Raghavendra, C., Sivalingam, K. M.: Data Gathering Algorithms in Sensor Networks Using Energy Metrics. IEEE Transactions on Parallel and Distributed Systems, Vol. 13, No. 9, 924-935. (2002)
18. Chen, Y. L., Lin, J. S.: Energy Efficiency Analysis of a Chain-Based Scheme via Intra-Grid for Wireless Sensor Networks. Computer Communications, Vol. 35, No. 4, 507-516. (2012)

19. Huang, C. F., Tseng, Y. C.: The Coverage Problem in a Wireless Sensor Network. In Proceedings of the ACM 2nd International Conference on Wireless Sensor Networks and Applications. San Diego, USA, 115-121. (2003)
20. Quang, V. T., Takumi, M.: An Algorithm for Sensing Coverage Problem in Wireless Sensor Networks. In Proceedings of the IEEE Sarnoff Symposium. Princeton, New Jersey, 1-5. (2008)
21. Quang, V. T., Takumi, M.: A Novel Gossip-Based Sensing Coverage Algorithm for Dense Wireless Sensor Networks. Computer Networks, Vol. 53, No. 13, 2275-2287. (2009)
22. Wang, X., Xing, G., Zhang, Y., Lu, C., Pless, R., Gill, C.: Integrated Coverage and Connectivity Configuration in Wireless Sensor Networks. In Proceedings of the 1st International Conference on Embedded Networked Sensor Systems. Los Angeles, USA, 28-39. (2003)
23. Wang, X., Xing, G., Zhang, Y., Lu, C., Pless, R., Gill, C.: Integrated Coverage and Connectivity Configuration for Energy Conservation in Sensor Networks. ACM Transactions on Sensor Networks, Vol. 1, No. 1, 36-72. (2005)
24. Chen, Y. L., Cheung, F. K., Chang, Y. C.: A Low-Energy Adaptive Clustering Hierarchy Architecture with an Intersection-Based Coverage Algorithm in Wireless Sensor Networks. In Proceedings of the Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, Taichung, Taiwan. (2013)
25. Chen, Y. L., Wang, N. C., Shih, Y. N., Lin, J. S.: Improving Low-Energy Adaptive Clustering Hierarchy Architectures with Sleep Mode for Wireless Sensor Networks. Wireless Personal Communications, Vol. 75, No. 1, 349-368. (2014)

**Young-Long Chen** (SM'03-M'05) received the Ph.D. degree in electrical engineering from National Chung Cheng University, Chia-Yi, Taiwan, in 2007. From 1995 to 1999, he worked for Formosa Petrochemical Corporation as a Design Engineer. From 1999 to 2007, he was a Lecturer with the Department of Electrical Engineering, Chienkuo Technology University, Taiwan. From 2007 to 2009, he was an Associate Professor with the Department of Electrical Engineering, Chienkuo Technology University, Taiwan. Since 2009, he has been with the Department of Computer Science and Information Engineering, National Taichung University of Science and Technology, Taiwan, where he is currently a Professor. His research interests include wireless and mobile communications and networks, wireless sensor networks, information security, digital signal processing, fuzzy neural networks and embedded systems. He is a member of the IEEE.

**Mu-Yen Chen** is an Associate Professor of Information Management at National Taichung University of Science and Technology, Taiwan. He received his PhD from Information Management from National Chiao-Tung University in Taiwan. His current research interests include artificial intelligent, soft computing, bio-inspired computing, context-awareness, RFID, multi-sensor networks management, financial engineering, and data mining. Dr. Chen's research is published or is forthcoming in Information Sciences, Journal of Educational Technology & Society, Journal of Information Science, The Electronic Library, Computers and Mathematics with Applications, Neural Network World, Quantitative Finance, Expert Systems with Applications, Applied Soft Computing, Soft Computing, and a number of national and international conference proceedings.

**Fu-Kai Cheung** received the M.S. degree in the Graduate School of Communication Engineering from the National Yunlin University of Science and Technology, Yunlin, Taiwan in 2010. His research interests are in wireless sensor networks, mobile communications, and communication protocols.

**Yung-Chi Chang** received the B.S. degree in Department of Computer Science and Information Engineering from National Taichung Institute of Technology, Taichung, Taiwan, in 2011. The M. S. degree is received in department of computer science and information engineering from National Taichung University of Science and Technology, Taichung, Taiwan, in 2013. His research interests in wireless sensor networks, mobile communications, and embedded systems.

*Received: September 26, 2013; Accepted: January 24, 2014.*

Computer Science and Information Systems 11(3):1037–1054 DOI: 10.2298/CSIS130920063L

# The Efficient Implementation of Distributed Indexing with Hadoop for Digital Investigations on Big Data

Taerim Lee[1], Hyejoo Lee[2], Kyung-Hyune Rhee[1], and Sang Uk Shin[1]

[1] Pukyong National University,
Busan, Republic of Korea
{taeri, khrhee, shinsu}@pknu.ac.kr
[2] Kongju National University,
Gongju, Republic of Korea
hyejoo2010@gmail.com

**Abstract.** Big Data brings new challenges to the field of e-Discovery or digital forensics and these challenges are mostly connected to the various methods for data processing. Considering that the most important factors are time and cost in determining success or failure of digital investigation, the development of a valid indexing method for efficient search should come first to more quickly and accurately find relevant evidence from Big Data. This paper, therefore, introduces a Distributed Text Processing System based on Hadoop called DTPS and explains about the distinctions between DTPS and other related researches to emphasize the necessity of it. In addition, this paper describes various experimental results in order to find the best implementation strategy in using Hadoop MapReduce for the distributed indexing and to analyze the worth for practical use of DTPS by comparative evaluation of its performance with similar tools. To be short, the ultimate purpose of this research is the development of useful search engine specially aimed at Big Data indexing as a major part for the future e-Discovery cloud service.

**Keywords:** Electronic Discovery, e-Discovery, Digital Forensics, Evidence Search, Indexing Performance, Hadoop MapReduce, Distributed Indexing.

## 1. Introduction

Recently, the number of lawsuits is rapidly increasing among business corporations due to the conflict of their interest. These types of incidents can be found easily around us, especially international disputes occur with frequency in the field of patent. But no matter who wins or loses the case, the most important thing is all companies involved in litigation are damaged economically with billions of dollars and they don't even know when the fight is going to be over. The only way to stop this remorseless dispute is that one litigant party secure crucial evidence closely related to the litigation issues and applies it to prove their legitimacy in trial. So, many business owners, professional executives and legal experts start to realize how serious the situation is and they are growing more interested in the skills of e-Discovery or digital forensics.

Electronic discovery (or e-discovery, eDiscovery) refers to discovery in civil litigation which deals with the exchange of information in electronic format called ESI (Electronically Stored Information). This legal system was the subject of amendments to the Federal Rules of Civil Procedure (FRCP), effective December 1, 2006, as amended to December 1, 2010 [14]. In addition, the growing number of legal cases for civil or criminal trials where critical evidence are stored in digital storages has been submitted as the digital forms of information with a high preference. So, state law now frequently addresses issues relating to electronic discovery. These changes make every litigant have a responsibility to produce their own evidence for themselves, so the use of digital forensic or e-Discovery tools becomes a necessary. Furthermore, the appearance of various IT compliances [15] pushes many global companies to reconstruct their business process and adopt a professional e-Discovery service or solution because traditional ERP (Enterprise Resource Planning) solutions are not satisfied enough to cope with fast-growing requirements of IT compliance effectively. To solve this problem, vendors who are related to the field of e-Discovery and forensics have competitively developed and released their own service systems applying the state-of-the-art technologies, but many drawbacks and challenges are still remain. Among them, the most serious problem is about the cost for doing an e-Discovery work.

In general, e-Discovery work is performed by jurists and IT experts who are collaborating with each other. When the litigation is filed, attorneys or a legal team hired by the litigant analyze the contents of petition and figure out major issues at first. And then, they produce a keyword list about evidence which must be secured on the basis of analyzed issues and deliver it to the IT experts. By using this keyword list, IT experts or a team composed of specialists for information retrieval or people of experience at forensic tools search the relevant data as potential evidence and visualize them for review. After this procedure, attorneys can review or analyze again the extracted data from various points of view such as suitability, sensitivity and confidentiality [11] in legal hold[1] situation [16]. According to this explanation, although doing an e-Discovery work sounds so easy, actually this is very complicated work and there are many cases which this procedure is not going well because of several unexpected variables such as system errors, data loss, or technical failures. In particular, electronic information is considered different from paper information because of its intangible form, volume, transience and persistence. Also, electronic information is usually accompanied by metadata that is not found in paper documents and that can play an important part as evidence (for example, the date and time a document was written could be useful in a copyright case). The preservation of metadata from electronic documents creates special challenges to prevent spoliation. For these reasons, developing an almighty system that can solve all kinds of problems is realistically impossible, so vendors focus on the development of functions for helping people deal with various e-Discovery tasks.

Fig. 1 shows the Electronic Discovery Reference Model usually called EDRM [7] and it was designed as part of an effort to provide essential requirements and guidelines of e-Discovery work to the people concerned, especially it also provide reference technologies to vendors. But, this model is too comprehensive and fundamental, so it is not suitable for reflecting the requirements caused by the latest practical problems.

[1] The legal hold is a process to preserve all forms of relevant information from malicious or inadvertent falsification.

Particularly in order to propose a solution for a specific problem, identification of key tasks about this matter and analysis of experimental result from various and concrete cases should be accompanied. As we know, according to some facts about the cost problem, the biggest cost-consuming part is the procedure for the attorney's review and it is bound up with time - the time spent considering whether each document is responsive or not [5]. For sticking to e-Discovery's basic principle that each document must be reviewed by a law expert before it is chosen as evidence and cutting down review expenses at the same time, the effectiveness of information retrieval to decide documents as review target should be improved. In the end, development of effective search method for finding relevant evidence more quickly and accurately is the key to the solution of cost problem. But in recent days, the biggest difficulty for doing this is the problem of Big Data.

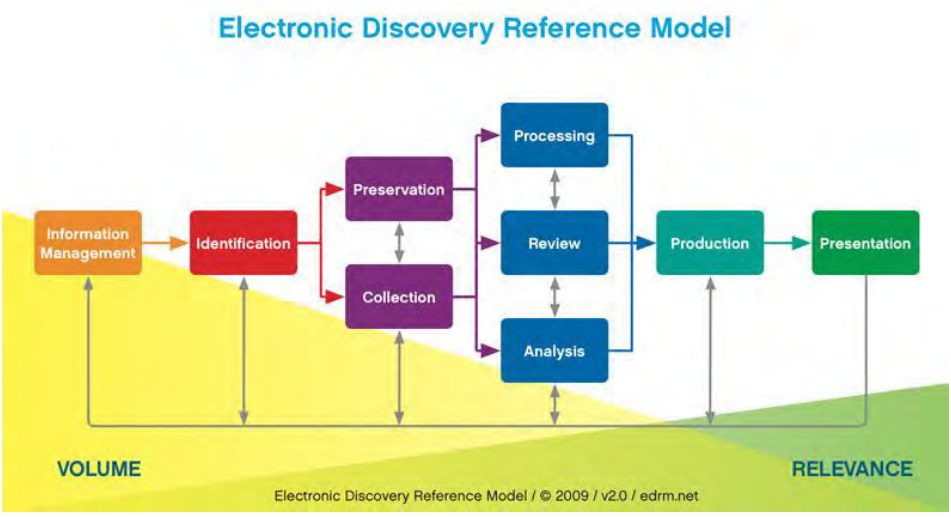



**Fig. 1.** Electronic Discovery Reference Model [7]

Big Data [17] is a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. That means the special techniques must be needed to process the data within a tolerable elapsed time, not excepting the digital investigation case. In other words, Big Data brings new challenges to the field of e-Discovery or forensics as like it does in other research areas and those are mostly connected to data processing methods for capture, curation, storage, search, sharing, transfer, analysis, and visualization. Suppose the litigant has a Big Data as an investigation target and he did not consistently manage them by using a special tool for search, the situation may be getting worse. At a time like this, all tasks for evidence search should be performed extemporarily from start to finish and more unfortunate thing is that such case occurs quite frequently. General text retrieval tools used in the digital forensics at present perform retrieval at an average speed of approximately 20 MB/s with respect to one query. It means 14 hours or more are required to retrieve a query in data of 1 TB [10] and time or cost cannot be expected exactly in case of Big Data. Especially, the most

time-consuming job for search is indexing because word vectors of every document have to be created. Of course, there are many ways of search without index, but these methods usually have well-known limitations such as metadata analysis or processing speed. To solve these kinds of problems caused by Big Data, new forms of document processing method for search needs to be established. In 2004, Google published a paper on a process called MapReduce [6]. Its framework provides a parallel programming model and associated implementation to process huge amount of data. This framework is very suitable for handling this kind of job.

This paper, therefore, introduces a Distributed Text Processing System based on Hadoop called DTPS which was belonged in our previous work [12] and explains about the distinctions between DTPS and other similar researches to emphasize the necessity of it. The role of DTPS is basically to make a searchable index files including metadata and it aims to handle this kind of job a lot faster than any other thing. Accordingly, this paper describes a series of experimental results by using different prototypes of DTPS. Each experiment was designed to evaluate the performance of specific prototype version which depends on its design for implementation. Final goal of these experiments is to find the best architecture and implementation strategy of DTPS using Hadoop as a major part for evidence search in the future e-Discovery cloud service. At last, the conclusion of this paper and our future work will be described.

## 2. Related Works

### 2.1. Distributed Lucene

Distributed Lucene [4] was the HP's project to create distributed free text index with Hadoop and Lucene in 2008. Given the origins of Hadoop, it is very natural it should be used as the basis of web search engines, a representative case of Big Data processing. It is currently used in Apache Nutch [8], an open source web crawler that creates the data set for a search engine. Nutch is often used with Apache Lucene, which provides a free text index. Despite the link between Hadoop and Lucene [8], at the time of developing there is no easy, off the shelf way to use Hadoop to implement a parallel search engine with a similar architecture to the Google search engine. So, after Doug Cutting came up with an initial design for creating a distributed index using Hadoop and Lucene, HP Labs published the technical report to describe their work for implementing a Distributed Lucene based on this design. For this reason, this project was necessarily undertaken in order to better understand the architectural style used in Hadoop. It means Distributed Lucene had a few limitations in respect of the function because developers only focused on the smooth combination of two other external open source projects, to the exclusion of its performance.

One notable point is Distributed Lucene does not use HDFS directly although the code for it is heavily influence by HDFS and was written by examining the HDFS code. The important reason for doing so it is not possible for multiple clients (or slaves, working nodes) to write the same index in HDFS. Therefore, each client in Distributed

Lucene must create own index about their quotas at first, and then they have to wait their turn for additional updating the same index. But in order to parallelize index creation, it is desirable for multiple clients to be able to access the same index and it is quite feasible through the reconstitution of operation methods for creating index. The other point is about Lucene. Lucene indexes generally contain a number of different files, some of which may be smaller than the 64MB block size for HDFS, so storing them in HDFS may not be efficient. Also, a few limited search techniques provided by Lucene at that time could be used because this project was suspended at the development quality of alpha.

### 2.2. Katta

Katta [1] is a scalable, failure tolerant, distributed, data storage for real time access. Katta serves large, replicated, indices as shards to serve high loads and very large data sets. These indices can be of different type. Currently implementations are available for Lucene and Hadoop map files. In other words, Katta is recent project adopted latest version of Lucene which provides various functions for using advanced search techniques like machine learning. But, Katta's structure and workflow may does not fit for meeting the growing demands of Big Data in digital investigation.

Fig. 2 shows how Katta works. As you can see, Katta uses an independent cluster or single server for creating index and it uses HDFS or shared file system only for storing the result of indexing. Strictly speaking, Katta is an application for distributed Lucene running on many commodity hardware servers very similar to Hadoop MapReduce, Hadoop DFS, HBase, Bigtable or Hypertable and this means a series of tasks for creating index are not processed dispersively. As a result, the efficiency of indexing and searching is totally influenced by the performance of each slave or cluster which deals with this kind of job. Well, this could be a good way of handling the simultaneous requests from multiple users, but it is not helpful when large amounts of data should be processed at a time. Therefore, Katta suggests a special recommendation and a sample code called SequenceFileCreator to manage this situation. While different and large data sources will probably exist, if user want to leverage the power of Hadoop while indexing SequenceFileCreator could be a good idea to get the data to the Hadoop DFS as raw text or as a sequence file [1]. However, it is just a sample code that will not make the available index, and this means enabling a fully distributed indexing based on Katta is the responsibility of developers using it.

### 2.3. Forensic Indexed Search System: HFSS

HFSS [10] is the forensic indexed search system which has been developed at Electronics and Telecommunications Research Institute (ETRI) of Korea. It provides a distributed forensic indexing and a web-based Forensic search service using Hadoop because it gives scalability as well as performance improvement. The most remarkable feature of HFSS is how to use the HBase to overcome the limitations about HDFS block size and MapReduce task. In MapReduce, book-keeping information is saved to keep

track of the task's status and process during a job initialization. Since a map task takes a block at a time when default input format is used, handling a lot of small files which needs a lot of map tasks causes book-keeping overhead. For this reason, HFSS uses a NAS to store original documents as target of indexing and loads extracted data into HBase table. The way of text processing DTPS try to seek is exactly same with HFSS, but using the HBase to store index information may cause the overhead time when the query is requested for search compared to the general way of using index files directly. Also, users who want to apply the advanced search techniques or who familiar with the traditional ways of search may feel HFSS is too difficult for them because available search method must be kept within bounds of HBase in NoSQL.

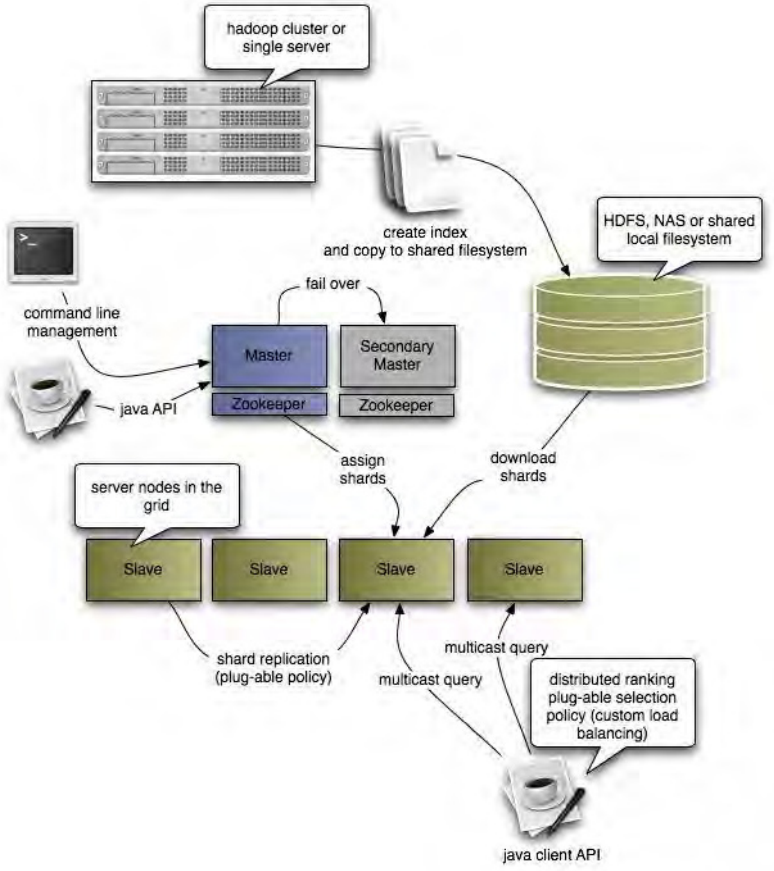



**Fig. 2.** About Katta: How Katta works [1]

## 3. Implementation Method for Efficient DTPS with Hadoop

In this section, the implementation scopes of DTPS prototype will be described. This includes system architecture for overcoming problems mentioned in section 2, basic requirements for text processing, constraints of each function and implementation strategies for the differentiation of performance test.

### 3.1. Prototype Design of DTPS

Fig. 3 shows the initial design of DTPS. This architecture was designed by considering the actual service type for digital investigation and combination with e-Discovery cloud service in the future. According to the service process, a simple use scenario and workflow of DTPS are as follows. Naturally, additional functions for doing this were added.

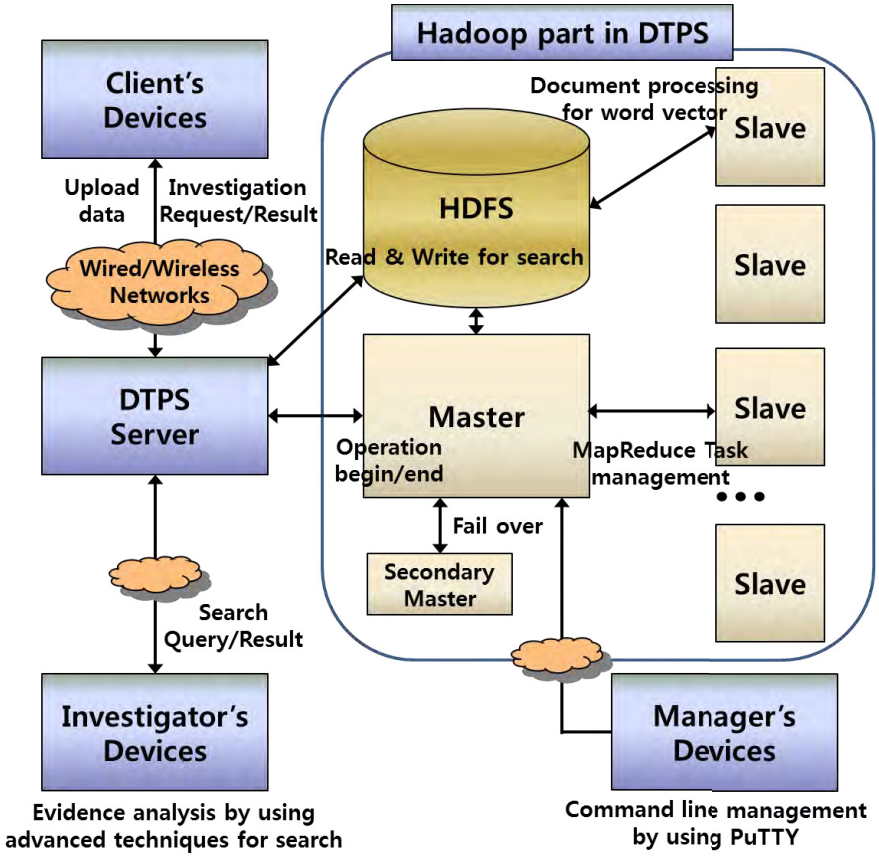



**Fig. 3.** Initial Design and Workflow of DTPS

When the clients call for an investigation at the beginning, they start to upload a target data of investigation to the server's local file system. After successful completion

of uploading, making input files in HDFS and extracting metadata runs in parallel, and DTPS server gives the order to the master for starting the operation of text processing. At the word of command, the master and slaves carry out their own tasks as programmed by MapReduce for creating word vector representations, and work results are stored in HDFS again. DTPS manager can monitor this whole process and perform various tasks for the management of Hadoop configurations through the command line tool like PuTTY or DTPS server application. Meanwhile in order to find evidence, investigators create a series of queries with advanced search techniques and send them to the DTPS server. And then, server returns the results of search to the investigator for analyzing and selecting some crucial evidence. Finally, DTPS server returns the result of investigation to the clients for the future trial.

### 3.2. Basic Requirements

The role of Hadoop MapReduce in DTPS prototype is restricted only for making a vector representation of document, and it is the most important and time-consuming task for creating index. These indices can be used for various search operations from the simple Boolean method to the advanced methods (for example, ranked search and document classification using machine learning algorithms). Also, DTPS uses HDFS directly to save its input documents and outputs. Outputs consist of six files, and descriptions about these files are shown in Table 1.

In Table 1, there are two files which are optionally created for supporting the advanced search techniques of DTPS. The first is XML file for archiving the metadata information of documents. In some cases, the metadata may be more important than the contents of document as evidence, and patent infringement case could be a good example like this. Therefore, considering that metadata search is the one of the helpful functions for e-Discovery solution, the implementation of it on DTPS takes priority above everything else. The metadata information extracted by DTPS is as follows and this is the simple design for gathering common elements from various document formats.

```
<DTPS_Metadata>
  <document>
   <docid>E:\ComSIS_EDaaS_cam_rdy.doc</docid>
   <metadata>
    <Author>Taerim Lee</Author>
    <creator>Taerim Lee</creator>
    <Creation-Date>2013-04-03T00:52:00Z</Creation-Date>
    <Last-Save-Date>2013-04-03T01:56:00Z</Last-Save-Date>
    <Last-Modified>2013-04-03T01:56:00Z</Last-Modified>
   </metadata>
  </document>
  <document> ... more documents </document>
</DTPS_Metadata>
```

**Table 1.** Summary of DTPS Index File Extensions

| Name | Extension | Brief Description |
|---|---|---|
| Term Dictionary | .tdf | The term dictionary, stores term list (Index number of each term, all terms in document collection). |
| Document List | .dlf | The document list, stores document information (Index number of each document, all docIDs in collection, docID is the original path of each file before it was collected from its local file system). |
| Posting List | .plf | The posting list, stores term information(Same index number of term in .tdf, Total Term Frequency in collection, Document Frequency, the list of value pairs {document index number : Local Term Frequency}, Local Term Frequency is the frequency number of specific term in one document). |
| Document Vector | .dvf | The vector representations of all documents in collection (Same index number of document in .dlf, the list of value pairs {term index number : weighted term vector}, weighted term vector is calculated by TF-iDF method). |
| Metadata | .xml | The common metadata information of all documents in collection (docID, author, creator, creation-date, last-save-date, last-modified). |
| Training Set | .dat | The training examples for document classification using machine learning method in specific category. This file is created by modifying the .dvf file, additionally stores indicator values of relevance. |

This is the example of extracted metadata from MS Word file and the XML tags apply to the original names of metadata as it is. The reason why we use the XML format is to consider the flexible extension of metadata structure caused from the additional information requirements. Apart from the indexing process, metadata extraction will be performed in the beginning process of DTPS in order to upload indexing target documents to the HDFS. So, this file is the first outcome produced by DTPS. The second is DAT file for training set and this file will be created by modifying the .DVF (Document Vector File). DTPS uses SVM$^{light}$ [9] for document classification as an advanced search method based on machine learning. Therefore, DAT file is exactly same with input format of SVM$^{light}$ so modifying the .DVF means adding target values of relevance(+1 as the target value marks a positive example, -1 a negative example respectively) to the document vector representation. Fig. 4 shows the example of training set partly extracted from the second file.

Above this, essential functions of DTPS programmed in Hadoop MapReduce for creating index are as follows.

– Document Loader: The step for loading individual documents from storage given by the source name of document information. Loading refers to the operation of opening a stream to fit for specific file format of each document. But, this prototype takes care of two types of file stream for local file system and HDFS.

- Feature Extraction [13]: The step for converting the document into clear word format called term dictionary. So, tokenization and removing stop words is performed.
- Feature Selection [13]: The step for selecting a subset of relevant features for use in model construction. This model is a vector space written by statistical characteristics of language. For prototype, best known scoring scheme of TF-iDF was applied with the log frequency weight of term and cosine normalization.

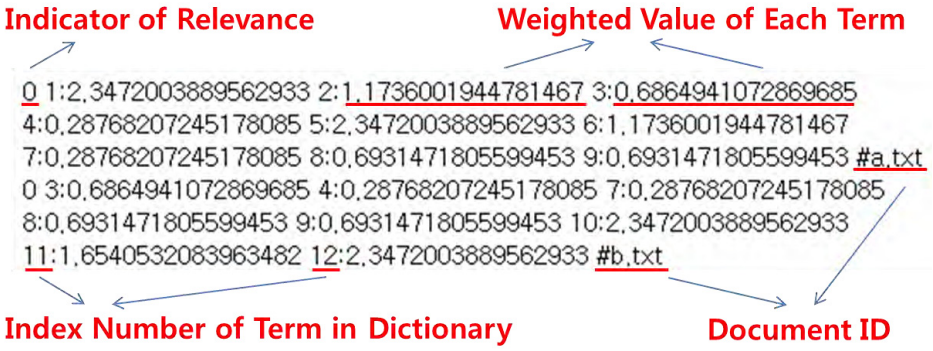



**Fig. 4.** The Example of DTPS Output

### 3.3. Implementation Strategy for Differentiated Experimental Design

Basic tasks in DTPS with MapReduce are as follows.

1. Preparation: Get input files ready in HDFS and extract metadata from the original documents (.xml)
2. Job Configuration: Set input paths of HDFS to decide the processing unit and the number of MapReduce tasks
3. Map: Reading the contents of each document - Tokenization - Stopword Filtering - Creating {Key, Value} with {docID, Term}. The docID is a path and name of the document in local file system, and term is a tokenized word in a currently processed document. The docID can be basically extracted by using the configure function in org.apache.hadoop.mapred.MapReduceBase.
4. Reduce: Comparing {docID, Term} - Counting a term frequency in same document and a document frequency in same collection - Writing the term dictionary (.tdf), the document list (.dlf), and the posting list file (.plf)
5. Completion: Calculating a TF-iDF weight - Creating a vector representation (.dvf) file of all documents in collection, and SVM input file (.dat) with the additional option

In order to confirm what the best implementation method for DTPS is, five test cases were established and differentiated factors in implementation for each case were applied. Details about experiments are summarized in Table 2.

**Table 2.** The summary of each experiment (For the description of differentiated factors, each number for basic tasks was simply used on the behalf of its name. - 1. Preparation 2. Job Configuration 3. Map 4. Reduce 5. Completion)

| No | Purpose | Differentiated factors of each case in implementation |
|---|---|---|
| Case 1 | Basic Test (Comparison Group) | 1. Just copy input files from server (local file system) to HDFS.<br>2. Processing Unit: Each file in document collection. The number of map tasks is same with the number of files.<br>3. Reading one line in each document and processing it.<br>4 & 5. No different from the basic tasks |
| Case 2 | Memory Test | 1, 3, 4 & 5. Same with Case 1<br>2. In order to avoid a possible failure caused by the book-keeping overhead in Case 1, add some codes to define a new text input format for processing multiple files like the MultiFileWordCount example in Hadoop. The number of map tasks is two. Additional factors for execution are :<br>- Running DTPS in Hadoop with increasing the heap size of Java.<br>- Running DTPS in Hadoop with adding physical RAM to the master. |
| Case 3 | Compression Test | 1 & 5. Same with Case 1<br>2. Same with Case 2 fundamentally, but add some codes to the Hadoop Job configuration for setting a compression library, Snappy. The number of task is same with Case 2.<br>3. Reading one line in each document and creating {docID, Term} with tokenization. After that, compress the outputs of mapper by using a Snappy before the reduce step.<br>4. Decompressing the mapper's outputs, and then start reduce task. |
| Case 4 | Merge Test | 1. To make one file for MapReduce input in HDFS, the content of each document was extracted, and merged to the input.xml. It is the same way as the metadata extraction.<br>2. Processing unit for inputs: created input file. The number of map tasks is the quotient (total size of created input file divided by Hadoop block size) + 1, same with the total number of used blocks for saving this input file in HDFS.<br>3. There are two primary XML tags, docID and content. After getting the docID at first, and then use the content to make pairs with tokenized terms. Actually, each value of the content tag is the full text of one document.<br>4 & 5. Same with Case 1 |
| Case 5 | Compression with Merge Test | This is the combination case between 3 and 4.<br>1. Same with Case 4<br>2. Same with Case 3, but the number of task is only one like Case 4.<br>3, 4 & 5. Same with Case 3 |

Case 1 is a basic test with no additional conditions as a comparison group. So, it will be performed by using a default configuration of Hadoop and document collection as it is. On the other hand, Case 2 is for comparing the performance according to the Hadoop configurations about memory and it has special meaning to suggest an alternative when experiment for Case 1 is failed because of memory overflow in Big Data processing. To do this, codes for processing multiple files were additionally implemented, and it makes only two map tasks will be required. Well, Case 3 was established to know the different performance of Hadoop when a specific library for compression was applied. Various libraries can be used for this test such as LZO, gzip or b2zip, but Snappy was only considered in our experiment because of its well-known advantages [3]. In Case 4, in order to extract the information of accurate docID from the merged input file, extraordinary measures are required because all files in HDFS are stored separately based on its block size and merging let the file lose its own metadata. Therefore, implemented code for merging writes one XML file with special tag for keeping this kind of information every time the new document is added. In our prototype, only two tags are used for input.xml, docID and content. It is because other metadata will be extracted and merged separately through the preparation task.

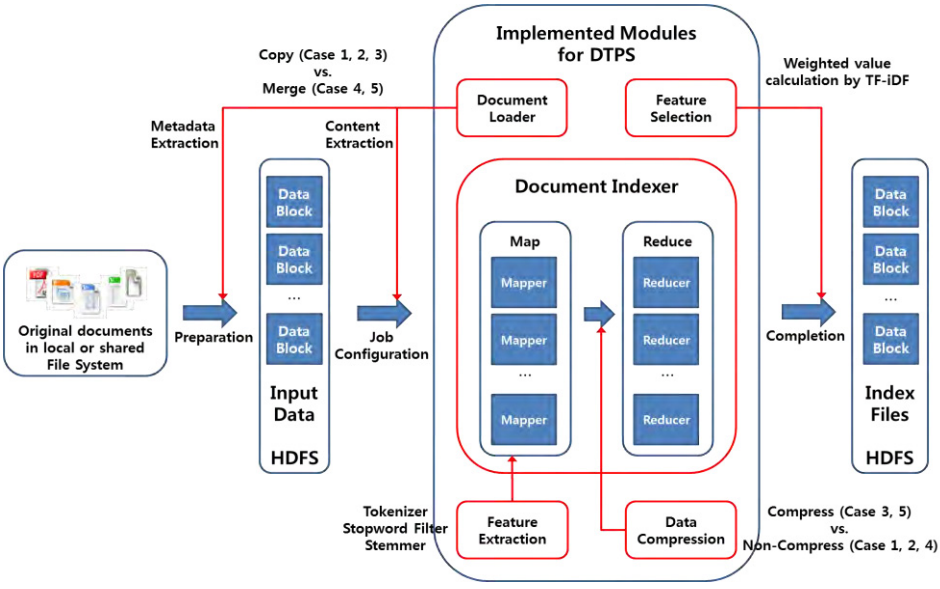



**Fig. 5.** The implemented modules of DTPS and control flow for each different experiment

Fig. 5 shows the implemented modules of DTPS based on the basic requirements for creating index and control flow for applying the features of each experiment in Table 2. Document Loader and Data Compression can be divided into separate parts, which can perform several different roles in this control flow. For example, Loader has a document parser for data extraction, and it is used for Preparation stage in the case of merge test, but it will be used for Job Configuration and Map stage in the case of copy test. So, they would be the biggest influence to the experimental results.

# 4. Evaluation Results and Analysis

Document collection used in our experiments is the EDRM Enron Email Data Set v2 [2]. This collection consists of 685,979 .txt files in 159 directories and the total size is about 4GB (3,991,162,863 bytes). Each .txt file was made by data extraction from email and its attachment files. For wide use of it, there are more types being provided by EDRM, such as PST or XML version.

## 4.1. Configurations of DTPS

Environment for developing DTPS and configurations of test-bed are as follows.

- System Hardware
  1) Hadoop Master: Two different masters for the memory test in Case 2. One of the masters has a 4GB RAM with default size of java heap which is generally used in PC, and another master has a 6GB RAM with 2GB maximum size of heap.
  2) Hadoop Slave: Two slaves with same hardware devices. Each slave has a Core 2 Duo CPU and 4GB size of RAM.
- OS: Ubuntu 12.04 LTS 64bit for the availability of extended RAM size
- IDE: Eclipse Standard, Kelper Release
- Programming Language: Java-6-oracle 1.6.0_45 version
- Library: Apache Hadoop 1.0.4 , Apache Tika 1.4 , Snappy 1.1.0 , SVM$^{light}$

Apache Tika was used for detecting and extracting metadata and structured text content from various documents using existing parser. Compared to the time of preparation task (Case 4 merging vs. Case 2 non-merging), the performance of Tika could affect our experiments. But, SVM$^{light}$ was applied for practical use of DTPS in order to give an example as an advanced search based on machine learning. It, therefore, has nothing to do with our experiments at all.

## 4.2. Test Results and Analysis

Table 3 shows the result of each experiment. The performance of DTPS was compared by the time of job completion for text processing and each test was repeatedly performed at least three times to find out the effects of initial job and different conditions of slaves such as communication status, unexpected errors or Hadoop fail over. It is generally recognized that the first job of Hadoop after it has started takes more time for warming up than the second or further executions although it is a same job.

Interestingly, while there was no real effect of Hadoop initial job, master's fail over for unexpected slaves' error would be the biggest influence to the completion time. In Case 1, all experiments were failed because of java errors associated with memory overflow. As we mentioned earlier in section 2 that too many tasks in MapReduce cause the book-keeping overhead, so failing job is a natural outcome in this case and additional actions are required to handle this problem like the rest of experiments. On

the contrary, text processing job succeeded without any problems in Case 2, but there is no great difference according to the RAM size. It means the memory size in Hadoop is more important factor for guessing the success or failure of a specific MapReduce job than its performance. Consequentially, considering a much bigger collection than used in our experiment, blindly increasing the size of RAM is a leap in the dark and will not help. In Case 3, there seem to be no advantages of Snappy in processing speed. But, Table 4 shows it was effective enough to improve the MapReduce performance on the other side.

**Table 3.** The time of job completion in each experiment

| Number of tries / Test Cases | Case 1 | Case 2 | Case 3 | Case 4 | Case 5 |
|---|---|---|---|---|---|
| The first try | Failed | 4h,38m,41s | 4h,30m,19s | 8m,54s | 8m,43s |
| The second try | Failed | 4h,25m,25s | 4h,26m,55s | 8m,41s | 8m,29s |
| The third try | Failed | 5h,2m,29s | 4h,25m,7s | 8m,45s | 8m,55s |
| Average time except the first try | N/A | 4h,43m,57s | 4h,26m,1s | 8m,43s | 8m,42s |
| Average time to prepare input | 4 hours | | | 2 hours | |
| Total time of completion | N/A | 8h,43m,57s | 8h,26m,1s | 2h,8m,43s | 2h,8m,42s |

**Table 4.** The Part of Hadoop Log Information : Non-Using vs. Using Snappy

| | Counters | Non-use | Use |
|---|---|---|---|
| FileSystem Counters | FILE_BYTES_READ | 7,148,098,440 | 2,579,056,504 |
| | FILE_BYTES_WRITTEN | 8,122,898,403 | 3,086,829,404 |
| Map-Reduce Framework | Map output bytes | 31,193,534,663 | 31,193,534,663 |
| | Reduce shuffle bytes | 969,587,271 | 502,559,596 |

As you see in Table 4, Snappy optimizes the distribution of Map outputs for decreasing the number of times being read and written by system I/O. It means Snappy enables MapReduce tasks to be processed more smoothly because the probability of system fail over is relatively lessened. We guessed the reason why the job completion time actually makes no difference with Snappy non-use case is the use of gigabit LAN for fast communication between master and slaves in DTPS test-bed. But, in other situations as Mapper makes a tremendous amount of output by using a bigger collection than our experiment or Hadoop has a poor network environment, the compression library would be very necessary. Before everything else, the deadly problem is the wasted space of storage in all preceding cases because general document is much smaller than HDFS block size. Naturally, Case 4 solved this problem and produced the most notable result in all experiments. Considering the possible overhead time caused by merging to make one input file with tagging information, there have been substantial improvements in the processing speed. In fact, the merging takes less time to prepare input file compared to uploading a full document collection to HDFS, and that means there is no overhead. Finally, from the experimental result of Case 5, we can confirm the

best strategy of implementation for DTPS is the combination of making an integrated input file and compressing intermediate data processed in MapReduce task.

And now let's take the final evaluation result from the comparison between three different simulations, Lucene indexing on a single machine, Katta indexing with the SequenceFileCreator (same conditions with the hardware configurations of DTPS), and DTPS indexing. Actually, we made a new code for SequenceFileCreator by modifying a sample code provided by Katta. Because this code only made several random records in sequence file format from one text file in order to recommend one way of using Hadoop for distributed indexing. That means it cannot process multiple files as general indexing tools did and does not fit for our test. On the contrary, our modified SequenceFileCreator can convert all texts in document collection to sequence file records, but search result produced by using these records is not available because it does not know which documents are including specific contents that match the user's information request. So, the goal of this trial is just to compare the work time required for preparing input file between Katta and DTPS. Additionally, for the rapid progression of this assessment, the master was replaced to the better system (Quad-Core i7 CPU, 12 GB RAM) based on the time result in our previous experiments and the same document collection was used.

**Table 5.** The performance comparison on the indexing speed between Lucene, Katta and DTPS

| Average Time \ Cases | Single Lucene | Katta | DTPS |
|---|---|---|---|
| Preparing input | N/A | 18m | 40m |
| Indexing | 1h, 31m | 20m | 7m |
| Additional time | N/A | 4m | N/A |
| Total | 1h, 31m | 42m | 47m |

In Table 5, only Katta needs additional time for deploying and adding the Lucene analyzer to its index and total indicates the overall time demanded by each system for being ready to search. Single Lucene means the basically same case with the Katta indexing, non-using the SequenceFileCreator and just copy indices from the independent content server to the HDFS for search. So, it takes the longest time in comparison with the others even if the time for copying index will be excluded. Meanwhile, DTPS takes a little more time than Katta, but it does not matter considering the index of Katta is not available in this experiment as above mentioned. As a result, if the Katta's SequenceFileCreator writes more information with modified data structure like the input of DTPS to make available index, we can expect the time required by both systems will be nearly the same. Also, we can be sure that merging all texts from multiple documents into one input file is a most useful way of improving the performance of distributed indexing in Hadoop. What was interesting about Katta, it uses only a Map task because Lucene provides the way of incremental indexing for merging distributed index files. Although additional verification of which method is better to make a final index, merging outputs by Reduce task in DTPS or incremental indexing by Lucene in Katta, may be needed, but at least DTPS is the winner in this experiment.

## 5. Conclusions

This paper described the research for implementing a Distributed Text Processing System using Hadoop MapReduce. Considering the latest requirements of e-Discovery caused by Big Data problems, major object of DTPS was the development of indexing method for search in order to find relevant evidence more quickly and accurately from large-scale data. To do this, five experiments were performed manipulating the code of MapReduce, the memory size of java heap and the type of input. As a result, we confirmed that the best strategy of implementation for DTPS is the combination of making an integrated input file and compressing data processed in MapReduce task. Also, in order to compare the performance of DTPS with similar tools, additional experiment was conducted and the result showed DTPS is useful enough to carry out a series of work for indexing effectively. On the guess that three different projects introduced in the section of related works may be undesirable for processing the Big Data according to circumstances in digital investigation, we hope this paper can clearly give the direction on developing the advanced e-Discovery or digital forensic service. From now on, considering the additional requirements of DTPS for using as the e-Discovery cloud service, complete realization and research on the accuracy improvement of search will be our future work.

**Acknowledgement**. This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (No.2011-0029927).

## References

1. Katta. 101tec, Inc. (2010), [Online]. Available: http://katta.sourceforge.net/ (current August 2013)
2. Trec legal track: Identification and download helpers for edrm enron v2 dataset. Text REtrieval Conference (2010), [Online]. Available: http://trec-legal.umiacs.umd.edu/corpora/trec/legal10/ (current August 2013)
3. Snappy, a fast compressor/decompressor. Google Project Hosting (2013), [Online]. Available: https://code.google.com/p/snappy/ (current August 2013)
4. Butler, M.H., Rutherford, J.: Distributed lucene: A distributed free text index for hadoop. Hewlett-Packard Development Company, L.P. (2008), [Online]. Available: https://www.hpl.hp.com/techreports/2008/HPL-2008-64.pdf (current August 2013)
5. Cohen, A.I., Kalbaugh, E.G.: ESI Handbook: Sources, Technology and Process. Aspen Publishers, New York City, USA (2008)
6. Dean, J., Ghemawat, S.: Mapreduce: Simplified data processing on large clusters. In: Proceedings of the Sixth Symposium on Operating System Design and Implementation. pp. 1–14. Google Research Publications, San Francisco, USA (2004)
7. EDRM: Edrm framework guides. EDRM LLC (2006), [Online]. Available: http://www.edrm.net/resources/guides/edrm-framework-guides (current August 2013)
8. Hatcher, E., Gospodneti, O.: Lucene in Action. Manning Publications Co., Greenwich, Connecticut, USA (2004)
9. Joachims, T.: SVM$^{light}$, support vector machine. Cornell University Department of Computer Science (2008), [Online]. Available: http://svmlight.joachims.org/ (current August 2013)

10. Lee, J., Un, S.: Digital forensics as a service: A case study of forensic indexed search. In: Proceedings of the ICT Convergence (ICTC), 2012 International Conference on. pp. 499–503. IEEE, Jeju Island, Republic of Korea (2012)
11. Lee, T., Kim, H., Rhee, K.H., Shin, S.U.: Design and implementation of e-discovery as a service based on cloud computing. Computer Science and Information Systems, 10(2), 703–724 (2013)
12. Lee, T., Kim, H., Rhee, K.H., Shin, S.U.: Implementation and performance of distributed text processing system using hadoop for e-discovery cloud service. Journal of Internet Services and Information Security, Vol.4, No.1, pp. 12-24 (2013)
13. Manning, C.D., Raghavan, P., Schtze, H.: Introduction to Information Retrieval. Cambridge University Press, Cambridge, England (2008)
14. Smith, L., et al.: Federal rules of civil procedure. U.S. GOVERNMENT PRINTING OFFICE (2012), [Online]. Available: http://www.uscourts.gov/uscourts/rules/civil-procedure.pdf (current August 2013)
15. Tarantino, A.: Compliance Handbook (Technology, Finance, Environmental, and International Guidance and Best Practices). John Wiley & Sons Inc., New York City, USA (2007)
16. Volonino, L., Redpath, I.J.: e-Discovery For Dummies. John Wiley & Sons Inc., New York City, USA (2009)
17. White, T.: Hadoop: The Definitive Guide. O'Reilly Media, California, USA (2012)

**Taerim Lee** received his Bachelor and Master of Engineering degrees from Pukyong National University, Busan Korea in 2008 and 2010, respectively. He is currently doing a Ph.D. program in Department of Information Security, Graduate School, Pukyong National University. His research interests include digital forensics, e-Discovery, cloud computing, and machine learning.

**Hyejoo Lee** received her M.S. and Ph.D degrees from PuKyong National University, Busan, Korea in 1997 and 2000, respectively. She worked as a senior researcher in Electronics and Telecommunications Research Institute, Daejeon, Korea from 2001 to 2005. She is currently working as Post Doctor in Department of Applied Mathematics at Kongju National University, Gongju, Korea. Her research interests include digital rights management, digital watermarking, multimedia protection and image processing.

**Kyung-Hyune Rhee** received his M.S. and Ph.D. degrees from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea in 1985 and 1992, respectively. He worked as a senior researcher in Electronic and Telecommunications Research Institute (ETRI), Daejeon, Korea from 1985 to 1993. He also worked as a visiting scholar in the University of Adelaide in Australia, the University of Tokyo in Japan, the University of California at Irvine in USA, and Kyushu University in Japan. He has served as a Chairman of Division of Information and Communication Technology, Colombo Plan Staff College for Technician Education in Manila, the Philippines. He is currently a professor in the Department of IT Convergence and Application Engineering, Pukyong National University, Busan, Korea. His research interests center on multimedia security and analysis, key management protocols and mobile ad-hoc and VANET communication security.

**Sang Uk Shin** (Corresponding author) received his M.S. and Ph.D. degrees from Pukyong National University, Busan, Korea in 1997 and 2000, respectively. He worked as a senior researcher in Electronics and Telecommunications Research Institute, Daejeon Korea from 2000 to 2003. He is currently an associate professor in Department of IT Convergence and Application Engineering, Pukyong National University. His research interests include digital forensics, e-Discovery, cryptographic protocol, mobile and wireless network security and multimedia content security.

*Received: September 20, 2013; Accepted: January 17, 2014.*

Computer Science and Information Systems 11(3):1055–1069 DOI: 10.2298/CSIS130918064H

# A New Detection Scheme of Software Copyright Infringement using Software Birthmark on Windows Systems

Yongman Han[1], Jongcheon Choi[1], Seong-je Cho[1], Haeyoung Yoo[2], Jinwoon Woo[2], Yunmook Nah[3], and Minkyu Park[4]

[1] Dept. of Computer Science, Dankook University
Yongin, Korea, 448-701
{grid_ym, godofslp, sjcho}@dankook.ac.kr

[2] Dept. of Software Science, Dankook University
Yongin, Korea, 448-701
{yoohy, jwwoo}@dankook.ac.kr

[3] Dept. of Applied Computer Engineering, Dankook University
Yongin, Korea, 448-701
ymnah@dankook.ac.kr

[4] Dept. of Computer Engineering, Konkuk University
Chungju, Korea, 380-701
minkyup@kku.ac.kr

**Abstract.** As software is getting more valuable, unauthorized users or malicious programmers illegally copies and distributes copyrighted software over online service provider (OSP) and P2P networks. To detect, block, and remove pirated software (illegal programs) on OSP and P2P networks, this paper proposes a new filtering approach using software birthmark, which is unique characteristics of program and can be used to identify each program. Software birthmark typically includes constant values, library information, sequence of function calls, and call graphs, etc. We target Microsoft Windows applications and utilize the numbers and names of DLLs and APIs stored in a Windows executable file. Using that information and each cryptographic hash value of the API sequence of programs, we construct software birthmark database. Whenever a program is uploaded or downloaded on OSP and P2P networks, we can identify the program by comparing software birthmark of the program with birthmarks in the database. It is possible to grasp to some extent whether software is an illegally copied one. The experiments show that the proposed software birthmark can effectively identify Windows applications. That is, our proposed technique can be employed to efficiently detect and block pirated programs on OSP and P2P networks.

**Keywords:** Software birthmark, Import Address Table (IAT), Software piracy, Software identification, Dynamic-Link Library (DLL), Application Programming Interface (API), Windows PE

# 1. Introduction

Though recent anti-piracy measures monitor Internet for detecting illegal upload or download of music and movies, copyrighted software has been still illegally distributed over Online Service Provider (OSP) and P2P networks. Software piracy is a growing concern in today's competitive world of software. Indeed, many incidents have been reported, and many software developers and copyright holders have been victimized by software theft. The Business Software Alliance (BSA) publishes the yearly study about copyright infringement of software. The Ninth Annual BSA 2011 Piracy Study reported that 57 percent of the world's personal computer users admit to pirating software [2]. The commercial value of all these pirated software rose from $58.8 billion in 2010 to $63.4 billion in 2011. Undoubtedly, software piracy causes severe damages to software industries, stifling not only IT innovation but also job creation across all sectors of the economy. In addition, a recent report of the BSA, "Competitive Advantage: The Economic Impact of Properly Licensed Software", reported that if you use genuine software globally 1% more, there are economic benefits of about $ 73 billion, whereas if you use infringe copyright 1% more, there are economic benefits of about $ 20 billion [3].

To protect the intellectual property for software developers [7], many software protection techniques have been proposed. Among them, software birthmark is a prominent technique. A software birthmark is a unique characteristic, or set of characteristics, that a program inherently has and can be used to identify that program. Existing birthmark schemes have some limitations, though. For example, a static source code-based birthmark [17] requires source code, and is not applicable to binary executable programs. This source code-based birthmark and other birthmarks, such as static executable code-based birthmark [13], dynamic whole program path (WPP)-based birthmark [12], and dynamic API-based birthmark [18], are not resilient to semantics-preserving obfuscation attacks, such as outlining and ordering transformation [8]. Also none of the existing static birthmarks has been evaluated on large-scale programs.

We propose a new software birthmark based on the number and names of *Dynamic Link Libraries* (DLLs) and *Application Programming Interfaces* (APIs) used in Windows applications.

This birthmark can be used to detect the obfuscated Microsoft Windows applications, including large-scale programs, and consequently to detect illegal distribution of copyrighted software over OSP and P2P networks. Windows executable programs have *Portable Executable* (PE) format, and their DLL and API information is stored in a section of PE, *Import Address Table* (IAT). For each application program, the number and names of DLLs and APIs, API call sequence, and a hash value for API call sequence can be inherent to each program and can be used as a unique birthmark. According to the characteristics of the number and names of DLLs and APIs, application programs can be grouped into several categories: Ftp client, Text editor, Media player, Image viewer, Compression tool, Messenger, Cd tool, p2p, etc. A categorization system speeds up search or identification process.

In this paper, we have first construct a birthmark database (DB) which contains the number and names of DLLs and APIs, category information, each hash value of API sequence of a program, and the information indicating that a corresponding program is commercial software or not. Whenever a program, $p_i$ is uploaded or downloaded on

OSP or P2P networks, the identification process of the program consists of four steps: (1) Classifying the $p_i$ into a category, (2) Inspecting the names of DLLs and number of APIs of the $p_i$ targeting only programs classified in the same category, (3) Computing a hash value using the sequence of API calls of the $p_i$ and comparing it with hash values of programs within the identified category, and (4) In case that the categorization in step (1) is failed and then the identification in step (3) is failed too, comparing the hash value of the $p_i$ with the hash values of all programs in the entire DB. If the identified program is commercial, upload or download is not permitted.

The rest of the paper is organized as follows. Section 2 outlines the background and related work. Section 3 describes the proposed software birthmark. In Section 4, we present typical scenario and detailed steps for identifying and filtering copyright infringement software. Section 5 presents the experiment results, and finally we summarize our conclusions and describe future work.

## 2. Background and Related Work

In this section, we give an overview on Import Address Table (IAT) of the Portable Executable (PE) on Microsoft Windows. The PE is the format of an executable binary on Windows OS. We also explain MD5 hash algorithm and various software birthmark schemes.

### 2.1. Import Address Table

Microsoft Windows operating systems use the PE format for executable files, object code, and DLLs [11]. The PE format contains dynamic library references for linking, API export and import tables, resource management data and thread-local storage (TLS) data. A PE file consists of a few headers and sections that tell the dynamic linker how to map the file into memory.

When a program is loaded, the Windows loader loads all the DLLs the application uses and maps them into the process address space. A DLL is simply a file that contains one or more pre-compiled functions. That is, each DLL contains pre-compiled implementation code for API functions. The executable file lists all the functions it requires from each DLL. This loading and joining is accomplished by using the IAT. The IAT is a table of function pointers filled in by the Windows loader as the DLLs are loaded.

The IAT is a lookup table when the application is calling a function from a different module. It can be in the form of both import by ordinal and import by name [11]. The IAT of a PE file is used to store virtual addresses of functions that are imported from external PE files. From the IAT, we can obtain the feature information of the program, such as the number of DLLs, the names of DLLs, and the names of API functions in each DLL.

## 2.2. MD5 (Message-Digest algorithm5) Hash Function

MD5 hash function receives a message of arbitrary length as input and output a 128 bit value. This function is widely used to check the integrity of an original executable file. It also can be used to identify specific software. However, a hashing function generates a completely different value from one bit change (Fig. 1).

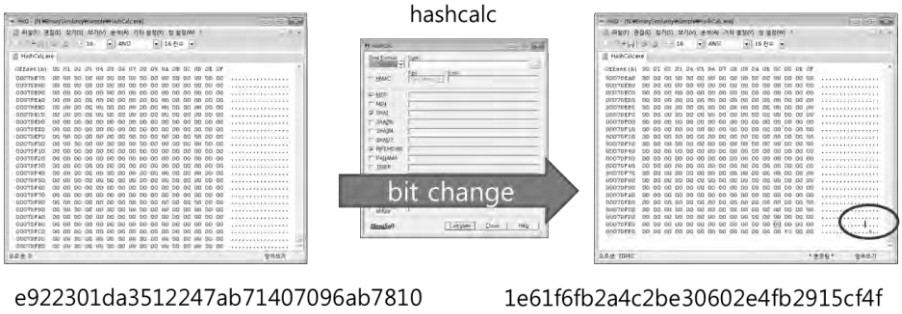



**Fig. 1.** A one bit change can generate an entirely different hash value

## 2.3. Related Work

A source code-based birthmark uses names of variables and functions [4]. This birthmark, however, no longer exists after compilation without special handling. Given only an executable file, we cannot use this birthmark for its original purpose.

Because of this limitation, many researchers are studying on API-based or system call-based birthmarks. These birthmarks are intact through compilation and can be used for detecting software theft and computer forensics.

Existing birthmarks can be classified into two categories. Static birthmarks extract statically available information in the program source code or executable files [4,9,13,17,20], for example, the types or initial values of the fields. Dynamic birthmarks, in contrast, rely on information gathered from the execution of a program [1,10,12,18].

Tamada et al. [17] proposed four types of static birthmark: constant values in field variables, sequence of method calls, inheritance structure, and used classes. All the four types are vulnerable to obfuscation techniques, such as code removal or splitting of variables [12]. In addition, their technique needs to access the source code and only works for an object-oriented programming language, such as Java.

Myles and Collberg proposed K-Gram-based birthmark, a static technique, which uniquely identifies a program through instruction sequences [13]. Instruction (opcode) sequences of length k are extracted from a program, and k-gram techniques, which were used to detect the similarity of documents [15], are used for the opcode sequence. The k-gram static birthmark is still fragile to some obfuscation methods, such as statement reordering, invalid instruction insertion, and compiler optimization.

Myles and Collberg presented another dynamic birthmark called a whole program path (WPP) and evaluated its performance on a Java program [12,14]. WPP is originally used to represent the dynamic control flow graphs (DCFGs) of a program. It collects all the compact DCFGs and regards them as a program's birthmarks. However, a WPP may not work for large-scale programs because of the overwhelming volume of WPP traces. Also, it is vulnerable to program optimization, such as loop transformations and inline functions.

Tamada et al. [18] introduced two types of dynamic birthmark for Windows applications: sequence of API function calls and frequency of API function calls. The sequence and frequency of Windows API calls are recorded during the execution of a program. Shuler and Dallmeier [16] presented a dynamic birthmark based on the extraction of API call sequence sets during program execution. API birthmarks are more robust to obfuscation than WPP birthmarks [19]. However, dynamic birthmarks need program executions which are dependent on user interactions, inputs, and environments.

Wang et al. [19] proposed two types of system call birthmark: system call short sequence birthmark and input-dependent system call subsequence birthmark. System call-based birthmarks can be platform-independent and are more robust to counter-attacks than API-based ones. They also need a program execution. Moreover, there are no easy ways to record system call traces of each application during program execution on Microsoft Windows systems.

Choi et al. [6] suggested a static API birthmark for Windows. Their birthmark is a set of possible API calls which are statically extracted by analyzing disassembled code. They did not use DLL information, which can be easily obtained from the IAT.

In our previous work [5], we have proposed the similar software birthmark to one proposed in this paper in order to identify each program. However, the previous software birthmark did not consider the sequence of API calls and its hash value, thus had some limitation to efficiently identify some programs of different versions. In addition, our previous work did not use software classification, and then had to compare the birthmark of a given program with all birthmarks in a birthmark database through the four steps. In this paper, we introduce (1) classification scheme to group some similar programs into a same category, and (2) API call sequence of a program and its hash value.

Current birthmarks are limited in their capabilities: some solutions are not strong enough to adequately prevent software theft, some cause significant performance degradation for large-scale programs, and some need program execution or work only for Java programs.

## 3. The Proposed Software Birthmark

The proposed software birthmark includes the following features (Fig. 2):

- number of DLLs and their names
- number of APIs and their names
- sequence of API calls

We extract the first two pieces of information from the IAT of the executable file.

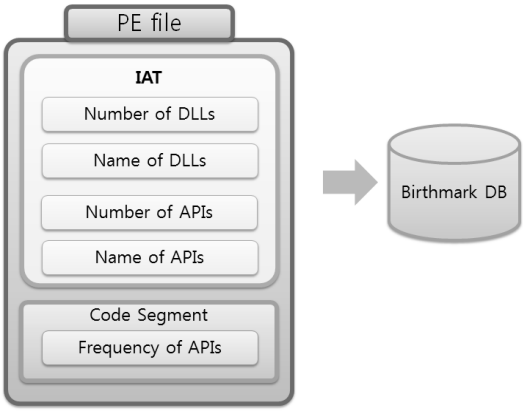



**Fig. 2.** The proposed software birthmark of Windows PE format file

Sequence of API calls can be obtained from the code segment of the executable file. The executable file is disassembled and sequence is extracted from it. We, then, calculate MD5 hash value on it (Fig. 3).

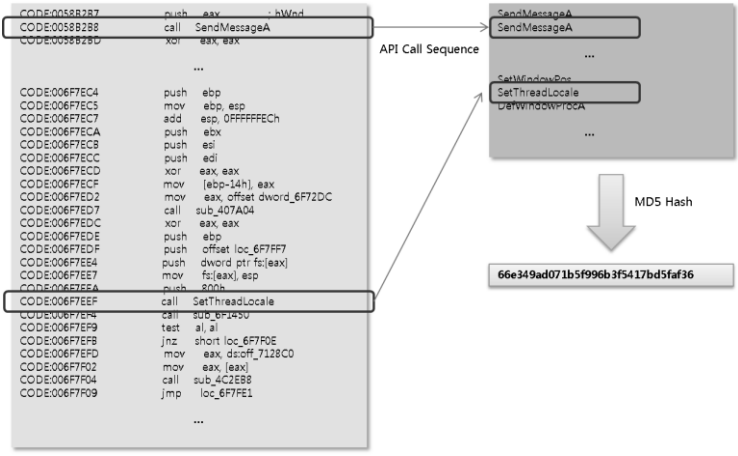



**Fig. 3.** How to caculate MD5 hash value from sequence of API calls

We store all this information to birthmark DB. The Schema of Feature Database is shown in Fig. 4. This database is a relational database and has several tables for DLL names, API names, and hash values. These tables are File information table, DLL information table, and API information table. The tables can be accessed using a file name and a DLL name.

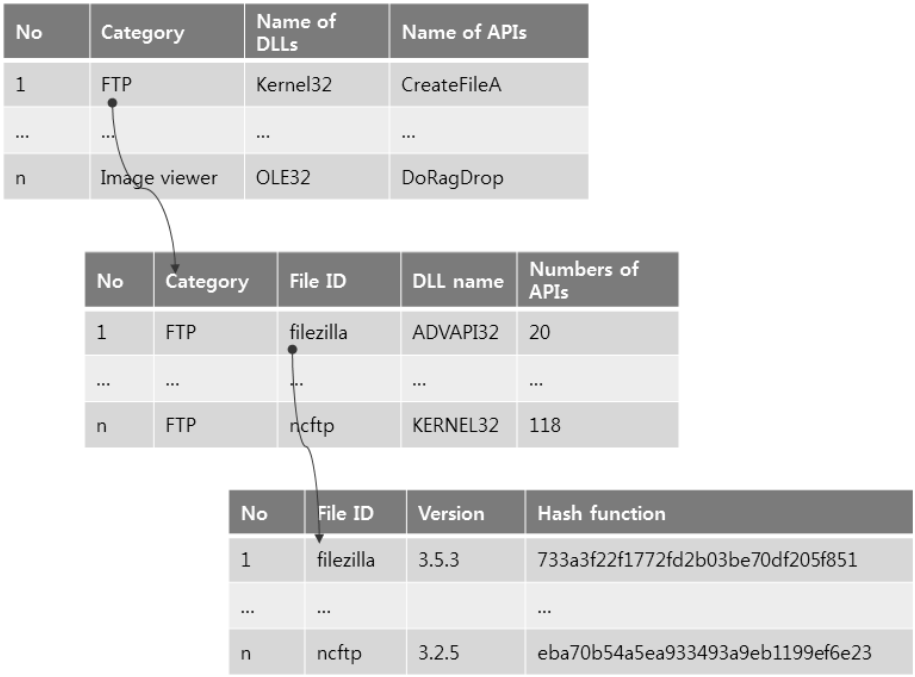



**Fig. 4.** The schema of the birthmark DB

You can see more details about this approach in our preliminary version of this paper [5].

## 4. Software Filtering using the Software Birthmark

### 4.1. Identifying and Filtering Overview

When a user tried to upload an application to an OSP, the OSP stores it at the temporary folder and asks the checking module that implements our proposed detection scheme whether it is commercial software distributed illegally. The checking module first extracts the software birthmark from the executable files of the application. The module, then, compares DLL and API information of the birthmark with category information in the birthmark DB to categorize it. After categorization, the module compare with all applications in the identified category using number of DLLs, their names, number of APIs and their names. If the module cannot identify the application, the module compares the hash value with the hash values of all applications in the same category. If the applications are not identified, the module compares the hash value with all hash values in the birthmark DB.

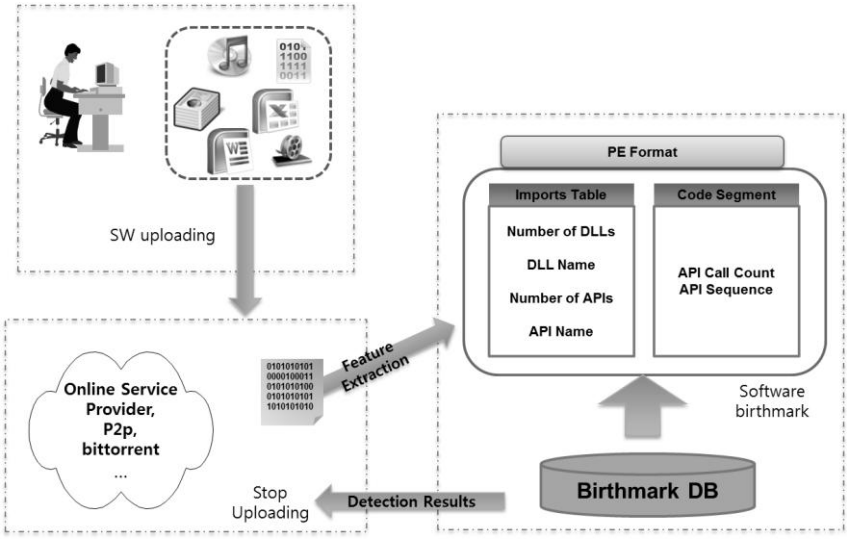



**Fig. 5.** The software identification and filtering process

If the application is illegally distributed commercial one, then the OSP stops the uploading procedure and delete the application. If the module cannot identify the application, the OSP inserts its software birthmark into the birthmark DB.

### 4.2. Detailed Steps

We describe the detailed identifying procedure. We denote the application being uploaded as $p_i$.

**Step 1:** Classifying the pi into a category. Using extracted software birthmark, we tries to identify a general kind of application. For example, if an application has software birthmark that appears in text editor, we can conclude the application may be some kind of text editor. We select 8 categories to identify, such as FTP client, Media player, Image viewer, etc. We think those categories include most representative application distributed via the Internet. This categorization helps to reduce the number of applications in the birthmark DB to compare. Software categorization, thus, can decrease comparison time when the size of the database is very large. If software cannot be identified, go to Step 3.

**Step 2:** Inspecting the names of DLLs and number of APIs of the pi targeting only programs classified in the same category. After the previous categorization, we compare names of DLL and the number of API functions used in the whole executable to the application in the same category, respectively. If the programs are not identified, go to Step 3.

**Step 3:** Computing a hash value using the sequence of API calls of the pi and comparing it with hash values of programs within the identified category. We extract the sequence of API calls from the code segment of the executable and input to the MD5 hash function. MD5 generate the 128 bit hash value. This hash value is compared to the hash values of the application belonging to the previously identified software category. We think this sequence may not change even against semantic preserving transformation attack. If the programs are not identified, then go to Step 4.

**Step 4:** comparing the hash value of the pi with the hash values of all programs in the entire DB. If an application is not identified yet, there may be some problems with categorization in Step 1. Therefore, we compare the hash value to the hash values stored in entire birthmark DB.

## 5. Experiments and Evaluation

### 5.1. Target Applications

To evaluate the effectiveness of our birthmark, we conduct an experiment using sample programs listed in Table 1. Sample programs are chosen in various categories like FTP clients, text editors, media players, etc.

**Table 1.** Sample applications

| Group | No. | Program | Version | Size (Kb) |
|---|---|---|---|---|
| FTP Client | (1) | Alftp | 5.3.2 | 4,109 |
| | (2) | Ncftp | 3.2.5 | 300 |
| | (3)-a | Filezilla | 3.5.3 | 7,994 |
| | (3)-b | | 3.5.2 | 7,993 |
| | (3)-c | | 3.4.0 | 7,463 |
| | (4)-a | WinSCP | 4.3.9 | 6,329 |
| | (4)-b | | 4.3.8 | 6,325 |
| | (4)-c | | 4.0.4 | 4,878 |
| Text Editor | (5) | Editplus | 3.20 | 1,787 |
| | (6) | Eclipse | 1.4.9 | 52 |
| | (7) | EXPAD | 0.4 | 845 |
| | (8)-a | AkelPad | 4.7.7 | 357 |
| | (8)-b | | 4.7.6 | 357 |
| | (8)-c | | 4.5.6 | 321 |
| | (9)-a | Notepad++ | 6.1.5 | 1,584 |
| | (9)-b | | 6.1.4 | 1,584 |
| | (9)-c | | 5.8.0 | 1,308 |
| Media Player | (10) | Alshow | 2.02 | 117 |
| | (11) | Coolplayer | 2.19 | 3,817 |

| | | | | |
|---|---|---|---|---|
| | (12) | GOM Player | 2.1.43 | 3,948 |
| | (13) | KMPlayer | 3.3.0 | 7,521 |
| | (14) | Loongplayer | 1.01 | 920 |
| | (15) | Mplayerc | 6.4.9.1 | 4,308 |
| | (16) | Potplayer | 1.51 | 180 |
| | (17) | Winamp | 5.6.3 | 2,156 |
| | (18)-a | Mixxx | 1.10.1 | 3,058 |
| | (18)-b | | 1.10.0 | 3,028 |
| | (18)-c | | 1.07.2 | 2,132 |
| Image viewer | (19) | Alsee | 6.8 | 6,960 |
| | (20) | Imagine | 1.0.8 | 17 |
| | (21) | Xnview | 1.99 | 4,624 |
| Compress Tools | (22) | Alzip | 8.53 | 2,855 |
| | (23) | Backzip | 5.03 | 1,920 |
| | (24) | Peazip | 4.6.1 | 4,023 |
| | (25) | TUGZip | 3.5 | 3,361 |
| | (26)-a | 7zFM | 9.22 | 411 |
| | (26)-b | | 9.20 | 412 |
| | (26)-c | | 9.04 | 383 |
| messenger | (27) | Pidgin | 2.10.6 | 49 |
| | (28) | Psi | 0.15 | 6,869 |
| | (29) | RetroShare | 0.54 | 14,340 |
| Cd tool | (30) | CDspace7 lite | 1.02 | 2,191 |
| | (31) | Dtlite | 4.41 | 4,796 |
| p2p | (32) | Emul | 5.0 | 5,624 |
| | (33) | Youdonkey | 2.35 | 240 |

## 5.2. Identifying the Target Applications

The overall comparison and identification results are shown in Table 2.

**Table 2.** Application identification results

| Group | No. | Step 1 | Step 2 | Step 3 | Step 4 |
|---|---|---|---|---|---|
| FTP Client | (1) | FTP/Media | Identified | | |
| | (2) | FTP | Identified | | |
| | (3)-a | FTP/Text | 3/17 | Identified | |
| | (3)-b | FTP/Text | 3/17 | Identified | |
| | (3)-c | FTP/Text | 3/17 | Identified | |
| | (4)-a | FTP/Media | 3/19 | Identified | |
| | (4)-b | FTP/Media | 3/19 | Identified | |
| | (4)-c | FTP/Media | 3/19 | Identified | |
| Text Editor | (5) | Text/Media | Identified | | |
| | (6) | Text/Zip/p2p | Identified | | |
| | (7) | Text/Zip/Msg | Identified | | |
| | (8)-a | Text | 2/9 | Identified | |

| | | | | | |
|---|---|---|---|---|---|
| | (8)-b | Text | 2/9 | Identified | |
| | (8)-c | Text | Identified | | |
| | (9)-a | Text | 3/9 | Identified | |
| | (9)-b | Text | 3/9 | Identified | |
| | (9)-c | Text | 3/9 | Identified | |
| Media Player | (10) | Media | Identified | | |
| | (11) | Media | Identified | | |
| | (12) | Text/Media | Identified | | |
| | (13) | Media | Identified | | |
| | (14) | Media/Zip/Msg | Identified | | |
| | (15) | Media | Identified | | |
| | (16) | Media | Identified | | |
| | (17) | Media/Zip | Identified | | |
| | (18)-a | Media/Zip | 2/11 | Identified | |
| | (18)-b | Media/Zip | 2/11 | Identified | |
| | (18)-c | Media/Zip | Identified | | |
| Image viewer | (19) | Text/Media/ Image | Identified | | |
| | (20) | Media/Image | Identified | | |
| | (21) | Image | Identified | | |
| Compress Tools (zip) | (22) | Text/Media/Zip | Identified | | |
| | (23) | Zip | Identified | | |
| | (24) | Text/Media/ Image | 0/23 | 0/23 | Identified |
| | (25) | FTP/Text | 0/17 | 0/17 | Identified |
| | (26)-a | Zip | 2/2 | Identified | |
| | (26)-b | Zip | 2/2 | Identified | |
| | (26)-c | Zip | Identified | | |
| messenger | (27) | Msg | Identified | | |
| | (28) | Msg | Identified | | |
| | (29) | Zip/Msg | Identified | | |
| Cd tool | (30) | Cd tool | Identified | | |
| | (31) | Cd tool | Identified | | |
| p2p | (32) | p2p | Identified | | |
| | (33) | p2p | Identified | | |

The Step 1 column of Table 1 represents the identified category after step 1 completes. Categorization is based on the assumption that programs in the same category use common DLLs and APIs. If an application is not clearly determined and seems to belong to two or more categories simultaneously, we compare it to all applications in both categories.

In Step 2, we try to identify only one application and uses DLL names and the number of APIs used. In Step 3, we extract the sequence of API calls from the disassembled code and generate MD5 hash value on it (Fig. 5). This hash value is compared to hash values of applications in the same category identified in Step 1.

If an application is not identified after Setp3, there are two cases we can think of.

**Case 1:** A new application. In this case, there is no information of the application considered in birthmark DB.

**Case 2:** Categorization failure. Step 1 fails to categorize an application. In our experiment, Peazip and Tugzip are such a case. In this case we compare the hash value of an application to all hash values of applications in birthmark DB.

After Step 2, we can identify most applications, but cannot identify applications with small difference. After Step 3, those applications can be identified and so MD5 hash function is effective for applications with small difference.

### 5.3. Measuring the Time to Identify an Application

We experimented with applications described in section 5.2 and obtained a detection accuracy of 95.56%. Since the difference between measured times was about 50ms, we repeated 3 times for one program and calculated the average time for each program. We calculate the average time for all programs by summing up all the average times calculated above and dividing the sum by the number of all programs. Fig.6 shows minimum, average, and maximum detection time. The minimum and maximum time equals to the smallest and largest average time, respectively. Longplayer is identified in the shortest time, 1981ms, because the number of API functions and DLL files used was small. The Peazip, on the other hand, was detected after the longest time has passed, 3345ms. In the case of Peazip, we need to complete step4 to identify. The average time is 2545ms, and most programs were discernible after Step 2.

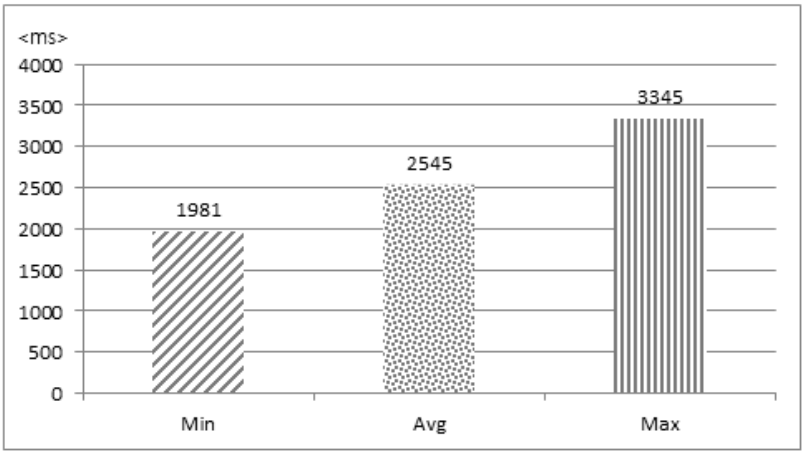



**Fig. 6.** The time required for identifying an application

## 6. Conclusion and Future Work

To detect software theft or piracy, a birthmark relies on the inherent characteristics of an application, which can be used show that one program is a copy of another. In this paper, we have proposed a new static birthmark scheme using the notion of Import Address Table, which can be used to identify Windows executable files, and MD5 hash values from sequence of API calls. Our birthmark is obtained by analyzing a Windows PE executable file and disassembling the PE file. We store this birthmark into a birthmark database and use it to compare the features of programs in concern.

We also use MD5 hash function on a sequence of API calls of an application. The sequence is extracted from the disassembled code of the application. This sequence is strong against the semantic preserving transformation attack.

We are working on ways to improve the efficiency of detecting illegal software and to elaborate comparisons with frequently used DLLs.

**Acknowledgements.** This research project was supported by Ministry of Culture, Sports and Tourism (MCST) and from Korea Copyright Commission in 2013, and by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (NIPA-2014-H0301-14-1023) supervised by the NIPA (National IT Industry Promotion Agency).

## References

1. Bai, Y., Sun, X., Sun, G., Deng, X., Zhou, X.: Dynamic K-gram based Software Birthmark. In Proceedings of 19th Australian Conference on Software Engineering. IEEE Computer Society, 644-649. (2008)
2. BSA: Shadow Market: 2011 BSA Global Software Piracy Study. Business Software Alliance, (2012)
3. BSA: Competitive Advantage: The Economic Impact of Properly Licensed Software. Business Software Alliance, (2013)
4. Burrows, S., Tahaghoghi, S., Zobel, J.: Efficient plagiarism detection for large code repositories. Software-Practice and Experience, Vol. 37, No. 2, 151-175. (2007)
5. Choi, J., Han, Y., Cho, S., Yoo, H., Woo, J., Park, M.: A Static Birthmark for MS Windows Applications Using Import Address Table. In Proceedings of the 7th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing 2013. IEEE, 129-134. (2013)
6. Choi, S., Park, H., Lim, H., Han, T.: A Static Birthmark of Binary Executables Based on API Call Structure. In Proceedings of 12th Asian Computing Science Conference. Springer, 2-16. (2007)
7. Collberg, C., Thomborson, C.: Software Watermarking: Models and Dynamic Embeddings. In Proceedings of the 26th ACM SIGPLAN-SIGACT symposium on Principles of Programming Languages. ACM, 311-324. (1999)
8. Kim, H., Khoo, W. M., Lio, P.: Polymorphic Attacks against Sequence-based Software Birthmarks. In Proceedings of 2nd Software Security and Protection Workshop. ACM. (2012)
9. Lim, H., Park, H., Choi, S., Han, T.: A Static Java Birthmark Based on Control Flow Edges. In Proceedings of the 33rd Annual IEEE International Computer Software and Applications Conference. IEEE Computer Society, 413-420. (2009)

10. Lu, B., Liu, F., Ge, X., Liu, B., Luo, X.: A Software Birthmark Based on Dynamic Opcode n-gram. In Proceedings of the First IEEE International Conference on Semantic Computing. IEEE Computer Society, 37-44. (2007)
11. Microsoft.: Microsoft Portable Executable and Common Object File Format Specification. Revision 8.2. (2010)
12. Myles, G., Collberg, C.: Detecting Software Theft via Whole Program Path Birthmarks. In Proceedings of 7th International Information Security Conference. Springer, 404-415. (2004)
13. Myles, G., Collberg, C.: K-gram based software birthmarks. in Proceedings of the 2005 ACM Symposium on Applied Computing. ACM, 314-318. (2005)
14. Myles, G.: Software Theft Detection Through Program Identification. PhD thesis. Department of Computer Science. The University of Arizona. (2006)
15. Schleimer, S., Wilkerson, D., Aiken, A.: Winnowing: Local Algorithms for Document Fingerprinting. In Proceedings of the 2003 ACM SIGMOD international conference on Management of data. ACM, 76-85. (2003)
16. Schuler, D., Dallmeier, V.: Detecting Software Theft with API Call Sequence Sets. In Proceedings of the 8th Workshop on Software Reengineering. ACM German Chapter, 56-57. (2006)
17. Tamada, H., Nakamura, M., Monden, A., Matsumoto, K.: Design and Evaluation of Birthmarks for Detecting Theft of Java Programs. In Proceedings of IASTED International Conference on Software Engineering. ACTA Press, 569-575. (2004)
18. Tamada, H., Okamoto, K., Nakamura, M., Monden, A., Matsumoto. K.: Dynamic Software Birthmarks to Detect the Theft of Windows Applications. In Proceedings of International Symposium on Future Software Technology. Software Engineers Association, 280-285. (2004)
19. Wang, X., Jhi, Y., Zhu, S., Liu, P.: Detecting Software Theft via System Call Based Birthmarks. In Proceedings of 25th Annual Computer Security Applications Conference. IEEE Computer Society, 149-158. (2009)
20. Xie, X., Liu, F., Lu, B., Chen, L.: A Software Birthmark Based on Weighted K-gram. In Proceedings of IEEE International Conference on Intelligent Computing and Intelligent Systems. IEEE, 400-405. (2010)

**Yongman Han** is doing a Ph.D. in Computer Science from University of Dankook, Korea in 2012. He has a master's degree from Dankook University. His research interests include software similarity, software theft, software engineering, software quality. He has authored and co-authored several journals and conference papers.

**Jongcheon Choi** is doing a Ph.D. in Computer Science from University of Dankook, Korea in 2005. He has a master's degree from Dankook University. His research interests include computer security, software theft, system software, software Protection. He has authored and co-authored several journals and conference papers.

**Seong-je Cho** received the B.E., the M.E. and the Ph.D. in Computer Engineering from Seoul National University in 1989, 1991 and 1996 respectively. He was a visiting scholar at Department of EECS, University of California, Irvine, USA in 2001, and at Department of Electrical and Computer Engineering, University of Cincinnati, USA in 2009 respectively. He is a Professor in Department of Computer Science, Dankook University, Korea from 1997. His current research interests include computer security, mobile security, operating systems, and software protection.

**Haeyoung Yoo** received the Ph.D. degree in Department of Computer Engineering, Ajou University in 1994. He is now a Professor in Department of Software Science, Dankook University, Korea. And he is now a vice-chairman of Korea Copyright Commission. His research interests include software development methodology, content technology policy and development, software testing and web engineering. He has authored and co-authored several journals and conference papers and software engineering textbook.

**Jinwoon Woo** received the Ph.D. degree in Department of Computer science, University of Minnesota, USA in 1990. He is now a Professor in Department of Software Science, Dankook University, Korea. His research interests include algorithm, information security, software assurance. He has authored and co-authored several journals and conference papers.

**Yunmook Nah** received the Ph.D. degree in Department of Applied Computer Engineering, Seoul National University in 1993. He is now a Professor in Department of Computer Engineering, Dankook University, Korea. His research interests include database, data modeling, large distributed database. He has authored and co-authored several journals and conference papers and database textbook.

**Minkyu Park** is the corresponding author of this paper. He received the B.E. and M.E. degree in Computer Engineering from Seoul National University in 1991 and 1993, respectively. He received Ph.D. degree in Computer Engineering from Seoul National University in 2005. He is now an Associate Professor in Konkuk University, Korea. His research interests include operating systems, real-time scheduling, embedded software, computer system security, and HCI.

*Received: September 18, 2013; Accepted: January 21, 2014.*

Computer Science and Information Systems 11(3):1071–1090 DOI: 10.2298/CSIS130922065D

# Pairwise and Group Key Setup Mechanism for Secure Machine-to-Machine Communication

Inshil Doh[1], Jiyoung Lim[2], Shi Li[1], and Kijoon Chae[1]

[1] Dept. of Computer and Science and Engineering,
Ewha Womans University, Seoul, Korea
isdoh1@ewha.ac.kr, lishi1116@gmail.com, kjchae@ewha.ac.kr
[2] Dept. of Computer Software,
Korean Bible University, Seoul, Korea,
jylim@bible.ac.kr

**Abstract.** In the ubiquitous environment, more and more devices are deployed in our daily life, and need to communicate with one another. M2M (Machine-to-Machine) communication is considered to be one of the major issues in future networks. M2M is expected to bring various benefits in wireless communications when it is interconnected with cellular networks. Considering the characteristics of cellular M2M networks, traditional security solutions are not proper to be applied to cellular M2M networks because the M2M network itself is vulnerable to various attacks. We consider security aspects for cellular M2M communications and propose a key management mechanism including the pairwise key and group key establishment. Our proposal could provide reliability and efficiency for the cellular M2M communication network in the secure manner.

**Keywords:** M2M, cellular M2M communication, security, pairwise key, group key

## 1. Introduction

A machine can communicate with another machine directly in wireless manners. The Machine-to-machine (M2M) has attracted a lot of people and industries for its ability to increase efficiency and improve productivity while reducing operating costs. It has great application areas and it can be connected with other infrastructure and brings much more powerful and efficient results. M2M devices or M2M Equipments (M2MEs) will ultimately connect to core network services through a variety of means, from direct broadband or capillary wireless networks, to wired networks.

Connectivity to these wireless and wired networks is an essential part of M2M communication networks. There is a need to be able to integrate a variety of application-specific technologies into a complete end-to-end solution to be offered by service providers [1]. The leap in technology would not be possible without the support of the wide area wireless communication infrastructure in particular cellular data networks. It is estimated that there are already tens of millions of such smart devices connected to

cellular networks worldwide and within the next 3-5 years this number will grow to hundreds of millions [2, 3]. Table 1 shows various application areas in the M2M communication. Among the application areas, the cellular M2M provides the ability to connect diverse devices and applications by enabling fixed assets, such as electric meters, or mobile assets, such as fleet vehicles. The cellular M2M is the best option to connect assets over great distances using already established, robust, and proven networks. The cellular technology is effective across widely varied industries because it is easy to integrate and cost-effective to deploy [4].

**Table 1.** M2M Application Area [5]

| Market | Description | Applications |
|---|---|---|
| Security | Abnormal situation detection | Suveilance |
| | Homeland/industiry security | Alert |
| Energy | Remote collect data on flow rate, pressure, temperature | AMR (automatic meter reading) |
| Transport | Tracking | Fleet Management |
| | Telematics services | Toll payment |
| | ITS | Emergency alerts |
| Commerce | Monetics | E-payment |
| | | Virtual wallet solution |
| Automotive | Adapted insurance rate | "Pay as you drive" |
| | Telematics services | Remote diagnotic |
| Home Automation | Remote monitoring, Managing | Surveillance |
| | | Energy management |
| Healthcare | Patients monitoring, Curing | Blood pressure check |

The cellular M2M market benefited from increasing numbers of mobile network operators launching M2M service offerings as their core service market grows increasingly mature and saturated. ABI Research expects cumulative cellular M2M connections to rise to 364.5 million globally by 2016. This report discusses the market and technical trends impacting the cellular M2M connectivity services market, analyzes cellular M2M connectivity service provider strategic responses, and forecasts cellular M2M connections and revenue growth for the period from 2007 through 2016, segmented by regions, applications, and air interface standards [6, 7].

There are lots of advantages of the cellular M2M. While Ethernet or Wi-Fi only provides the local coverage, cellular networks provide the ubiquitous coverage and the global connectivity. Users are already familiar with cellular networks, and they could use M2M applications easily on proprietary platforms [8].

A cellular M2M has great applications including telematics, asset management, U-healthcare, security and so on. Its application area will be drastically expanded. The more an organization relies on information technology and the more mobile it is, the greater the risks of security breaches. The success and the expansion of M2M depend on protecting security issues such as confidentiality, integrity, and availability of the data. As in Fig.1, basically, a Machine Type Communication (MTC) device, or an M2ME can be managed by the MTC server to be used by MTC users. They can be interconnected with each other through a Mobility Management Entity (MME), a Packet Gateway (P-

GW), and a Serving Gateway (S-GW). In some cases, they can communicate directly with each other. An M2ME is easy to be lost and hard to detect the malfunction. When integrity is not guaranteed, the equipments are excluded for services. In addition, M2MEs from one server or from one M2M user should be authenticated as one group, and they need to provide the individual communication at the same time.

Our contribution is that we have newly suggested the pairwise key establishment mechanism to make the direct communication more flexible and more secure for devices in mobile environments. It is especially important because the devices are mobile and can be located in the communication range of the others. If pairwise keys need to be distributed by the eNB every time they are required, the keys could be captured by the attackers. In addition to that, pairwise keys need to be generated in more structured way for better management. We propose a pairwise key generation mechanism using the key related information which is computed and delivered by the eNB efficiently in the energy and time consumption. Our proposal includes followings.

- Key establishment between eNB and mobile M2ME
- Key establishment between a pair of M2MEs for direct communication
- Group key establishment among M2MEs for group communication: Depending on the group key generation mechanism in our previous work [9], we defined functional group keys and regional group keys based on the location or their functions.
- We further simulated our proposal to compare the energy consumption for each type of devices. We also analyzed the computation, communication, and security aspects.

The remainder of this paper is organized as follows. Section 2 describes the system architecture for our proposal. Keys required for cellular M2M communication and the pairwise key agreement and group key agreement mechanisms are explained in Section 3. Section 4 evaluates the performance and security. Finally, we conclude our paper in Section 5.

## 2. Related Works

The 3GPP SA3 studied in TR 33.812, “Feasibility study on remote management of USIM application on M2M equipment”. Its goal is to make it possible that the network can provision the remote management of USIM and ISIM application at M2M equipments in a secure way in a 3GPP system [10]. One of the main issues in TR 33.812 is to investigate candidate security solutions and signaling procedures for provisioning and the remote management of USIM/ISIM applications at M2M equipments in a secure manner. When an M2M is connected with cellular networks, its vulnerabilities to various attacks are increased. Security vulnerabilities get more serious when M2M is adopted on the top of cellular communication technologies. As a result, the growth of cellular M2M services would be limited without providing the service security.

We have proposed the key establishment and management mechanisms in our previous work [11]. Especially, for the direct communication between two devices, we suggested key distribution by an eNB (evolved Node B). When a pair of devices need a direct communication, they request pairwise keys to the eNB, and under the cooperation of eNBs, pairwise keys are generated and distributed by the eNBs for communication

between two devices. In this work, we enhanced the pairwise key establishment mechanism to provide the security for the service.

In general, group key management mechanisms can be classified into three categories. In centralized key management schemes, a group manager generates group keys and distributes the key to authenticated group members and manages the key material and lists. Blundo, C. et al. proposed a mechanism in which a server chooses a *t*-degree polynomial randomly and distributes them to neighbor nodes and the member nodes substitute the polynomial with their IDs; hence, all the nodes share one group key [12]. Wang, Y. and Ramamurthy, B. proposed four safe group communication methods [13]. Information for group key rekeying is unicasted to each node. This creates a heavy overload when group size grows. Broadcasting is proposed to solve the overhead problem. The broadcasting mechanism requires heavier overhead when groups are generated; however, rekeying cost is relatively low. Overlapping is also proposed to prevent a flooding attack. Finally, the group information pre-distribution minimizes the group generation time. A lot of researches have been done for centralized group key management. However, in mobile communication environment, parent-child relationship changes constantly because of devices movements. Even if centralized management is very stable and secure, it is not proper for adopting in mobile networks.

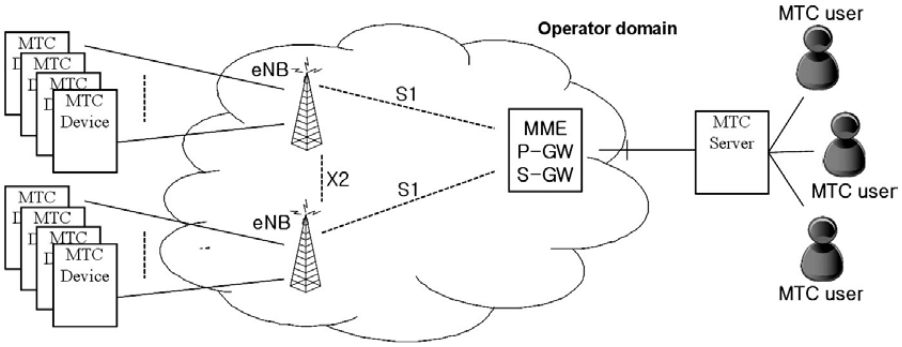
In distributed key management, multiple key managers generate group keys and distribute them to authentic members. Zhang, W. and Cao, G. proposed a mechanism (PCGR) that pre-distributes key related information and generates group keys [14]. When the group key rekeying is required, nodes cooperate and a new group key is computed. This scheme is applied in our proposal and will be more described in subsection 3.3. Huang, J. H. et al. proposed a level key infrastructure for multicast and group communication that uses level keys to provide an infrastructure that lowers the cost of nodes joining and leaving [15]. This scheme has a drawback in that process delay increases even when many nodes are changed. Zhu, S. et al. proposed a key management protocol for sensor network designed to support in-network processing, while at the same time restricting the security impact of a compromised node [16]. This mechanism is safer, because it uses four different kinds of keys. However, key update consumes much overhead. Adusumilli, P., Zou, X. and Ramamurthy, B. proposed a Distributed Group Key Distribution (DGKD) protocol which does not require existence of central trusted entities such as group controller or subgroup controllers [17]. Aparna, R. and Amberker, B.B. proposed a key management scheme for managing multiple groups. They uses a combination of key-based and secret share-based approach for managing the keys and showed that it is possible for members belonging to two or more groups to derive the group keys with less storage [18]. Kim, Y., Perrig, A, and Tsudik, G. investigated a novel group key agreement approach which blends key trees with Diffie--Hellman key exchange [19]. It yielded a secure protocol suite called Tree-based Group Diffie-Hellman (TGDH) that is both simple and fault-tolerant.

Contributed management mechanisms rekey the group keys through nodes' cooperation without specific key managers. Yu, Z. and Guan, Y. propose a group key management mechanism [20] in which basic matrix G and secret matrices A,B are assigned to each sensor node; each matrix is used to generate group keys among nodes in the same groups and different groups, respectively. The advantage of this mechanism is that the probability of generating group keys is high. However, when the grid size is

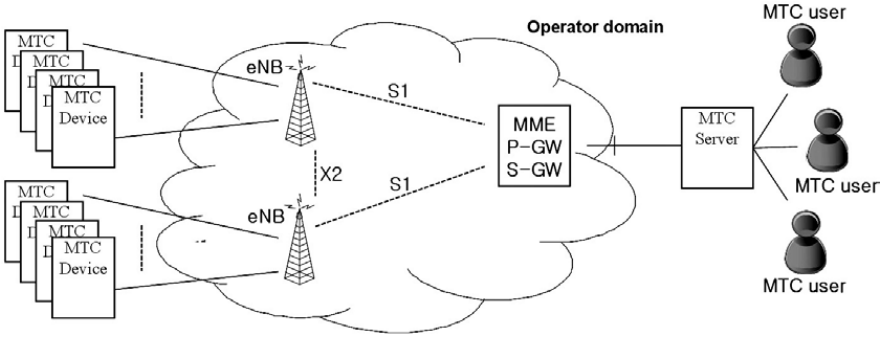
large, much energy is wasted and when the grid size is small, group keys may not be generated.

## 3. System Architecture

Fig. 1 shows the M2M service infrastructure under the cellular communication environment. As in the figure, devices can communicate with one another with the help of eNBs which play the role of intervention. In this work, we basically assume that the devices can communicate directly when they are located closely enough while they move.




(a) Traditional M2M Communication Service

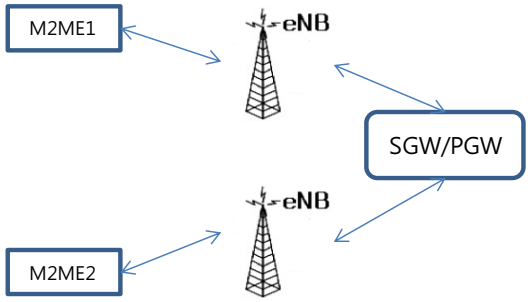



eNB: evolved Node B
P-GW: Packet Gateway
S-GW: Serving Gateway
(b) Cellular M2M Communication Service

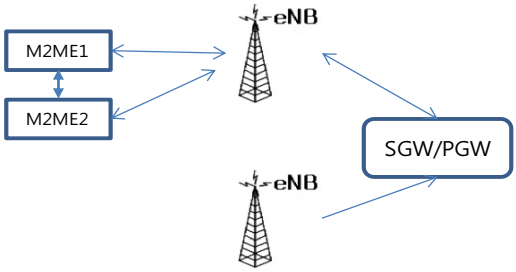
**Fig. 1.** M2M Service Infrastructure

M2M devices could have high mobility. They need to communicate with other various devices while they are on the move. As in Fig. 2, a pair of M2M devices can communicate with each other when they meet and recognize that they are located in each

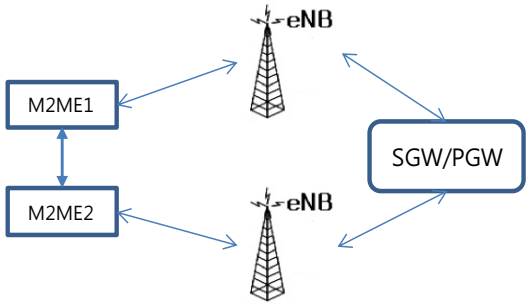
other's communication range. When they are enough close, they do not need the help of eNBs, but talk to each other as in (b) and (c) of Fig. 2. And for the direct communication, they need to be distributed pairwise keys for the secure connection. The pairwise keys are very important because they are the base for various security services such as confidentiality, integrity, authentication, and so on. Especially, these keys should be generated and deleted often without the breach of security in the mobile ubiquitous environment. The proper pairwise key generation mechanism should be provided.




(a) Default data cellular M2M communication




(b) Locally routed M2M communication




(c) Direct mode M2M communication

**Fig. 2.** Cellular M2M device communication

# 4. Key Establishment for M2M Communication

Various keys are required for the secure cellular M2M communication. They need to be established for data encryption, authentication, integrity, and so on. The keys required for the M2ME communication are as follows. We assume that each pair of eNBs shares pairwise keys for the secure communication among them. This assumption is reasonable because they are connected with one another in the wired infrastructure and considered relatively safe.

**Pairwise Keys between an eNB and an M2ME.** For the default data communication in the cellular M2M communication, an eNB and an M2ME need to share a pairwise key. In the initial stage, each M2ME belongs to specific eNB. However, they have the mobility and can meet the other devices while they are on the move. Even the devices move and communicate with other devices, the security should be provided in the reliable manner.

**Pairwise Keys between M2MEs.** As described in Section 2, M2MEs can communicate with each other with the help of an eNB, or they can communicate directly when they are in each other vicinity as in (b) and (c) of Fig. 2. There are a lot of devices and many instant direct connections are established and abolished. We need to support the situation with proper pairwise keys.

**Functional Group Keys for M2MEs.** Some M2MEs need the group communication. When it is for the functional group communication, they can share a group key. The group keys need to be managed by the Mobility Management Entity (MME) in Fig. 1, because the M2MEs are still group members even if they move from one cell to another.

**Regional Group keys for M2MEs.** The regional group can be formed in some region of the network field. When an M2ME moves in the region, they need to be provided the group key while they stay in the region and want to receive the data traffic of the group (Fig. 2 (b)). When they leave the region, the key is not valid anymore and the old group key needs to be rekeyed depending on the membership policy.

## 4.1. Key Establishment between eNB and Mobile M2ME

In our previous work [9], we have proposed the key establishment and authentication mechanism based on the USIM card for the ubiquitous healthcare system. For the cellular M2M communication, we basically assume that the USIM card and A3 and A8 algorithms are deployed in each M2ME. Based on the assumption, we can apply the initialization, key establishment, and authentication mechanism in our previous work to the cellular M2ME communication.

When an M2ME device is registered to an eNB, the ID of M2M device and a hashed key, $H(K_i)$ for the key generation are transferred to the eNB and M2ME through a secure channel. After registering the IDs of the device, the device and the eNB need to

generate key chains with hash functions and A8 algorithm. The authentication processes for mobile devices to eNB is described in Fig. 3.

After getting the $ID_{M2ME}$ and $H(K_i)$ of the M2M device, the eNB generates a nonce and encrypts it with $H(K_i)$ for the device to process A3 algorithm for authentication. After receiving and decrypting $Enc_{H(Ki)}$(nonce), the device computes A3 to generate $RES_{M2ME}$ and sends this value back to the eNB. The eNB also computes $RES_{eNB}$ with $H(K_i)$, nonce, and A3, and compares two values. If the eNB verifies the results are the same, authentication is completed. Then, two parties generate a hash chain and exchange the commitment values for the pairwise key generation. In this way, two parties prepare the keys for the future communication. Each computes the session key by computing A8 algorithm with the seed value from the key chain.

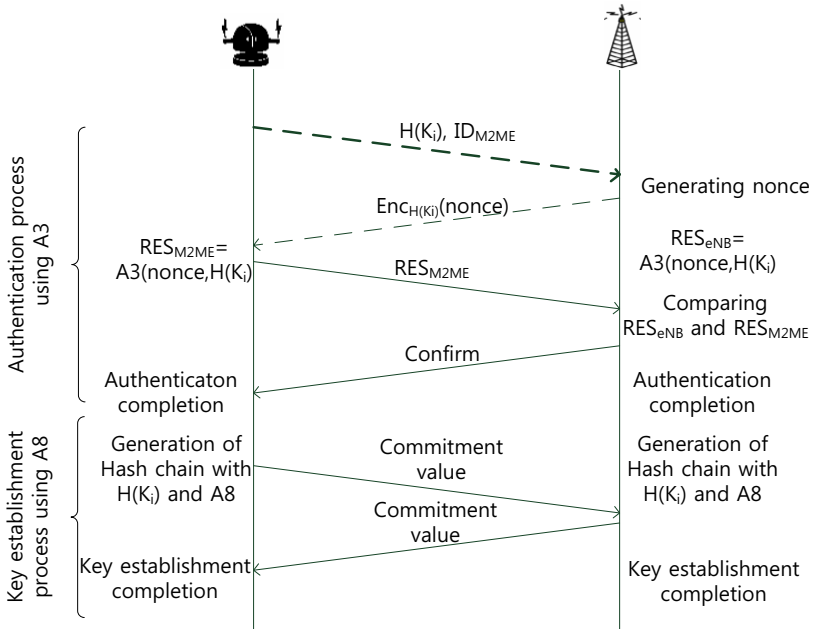



**Fig. 3.** M2ME authentication and key generation in the cellular M2M communication system

When an M2ME moves in the cell, the eNB notifies it to an MME and receive the security information from the eNB where the M2ME has left. The information is renewed periodically for the security purpose.

### 4.2. Key Establishment between a Pair of M2MEs for Direct Communication

When M2MEs are communicating directly with each other, there are many advantages. Time and frequency resources can be reused and the latency can be reduced. For direct communication, pairwise keys are required for security. The pairwise key establishment processes follows on.

The eNB randomly generates an n×n grid with a set of $2^n$ bivariate polynomials $\Phi=\{F_i(x,y), G_i(x,y)\}i=1,2,\ldots,n$ as shown in Fig. 4. Each row i in the grid is associated with a polynomial $F_i(x,y)$, and each column i is associated with a polynomial $G_i(x,y)$. Each M2ME located in the eNB communication range will be randomly assigned to a unique intersection in the grid. For the M2ME at the coordinate (i, j) in the grid, (i, j) is considered as the ID of M2ME, and eNB distributes the polynomial shares of ($F_i(x,y)$, $G_j(x,y)$) to the M2ME. In an example in Fig. 4, an M2ME (h, k) is assigned to the polynomial shares of ($F_h(x,y)$, $G_k(x,y)$), and an M2ME (p, q) is assigned to ($F_p(x,y)$, $G_q(x,y)$) similarly. And the polynomial shares belonging to an M2ME have two intersections with the polynomial shares belonging to the other one in the grid which are marked by stars in Fig. 4. The intersection polynomial shares are ($F_h(x,y)$, $G_q(x,y)$) and ($F_p(x,y)$, $G_k(x,y)$) respectively.

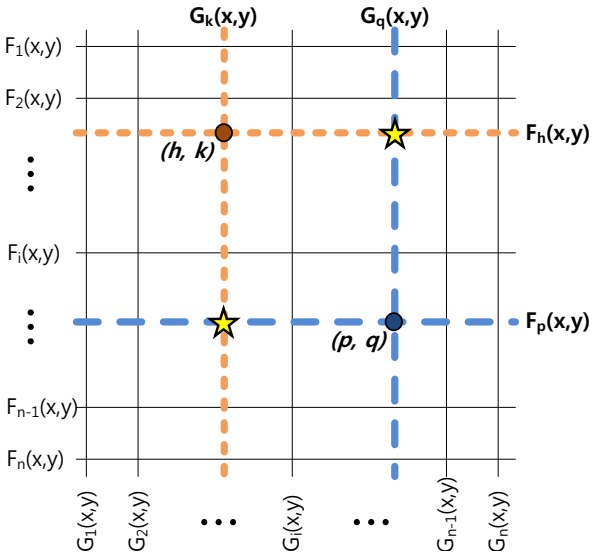



**Fig. 4.** Grid based key information distribution to M2MEs by eNB

When two M2MEs located in the same communication range of an eNB want to transmit secret messages to each other directly, they should encrypt the message by using pairwise keys between them. The pairwise key generation process is as follows:

- If there are two M2MEs want to communicate with each other directly, as mentioned above, the eNB will generate two points (e.g., (h,k) and (p,q)) as the ID of each M2ME in the n×n polynomial grid and distribute the IDs and the polynomial shares at the intersection of corresponding point in the grid (e.g., $\{F_h(x,y), G_k(x,y)\}$ and $\{F_p(x,y), G_q(x,y)\}$) to them respectively. As a result, the first M2ME obtains its ID(h,k) and polynomial share $\{F_h(x,y), G_k(x,y)\}$, and the second M2ME also receives its ID(p,q) and the corresponding polynomial share $\{F_p(x,y), G_q(x,y)\}$ as shown in Fig. 4.
- According to the above theory, an eNB can also find another two intersection polynomial shares in the grid as the star points shown in Fig. 4 and one polynomial

share of them will be selected as the common secret information of the pairwise key. Assume that the intersection star point at the top-right corner ($F_h(x,y)$, $G_q(x,y)$) is selected here. And an eNB will inform two M2MEs of the selected polynomial part from each of them. Here, the first part $F_h(x,y)$ comes from the polynomial share of an M2ME (h,k) and the second part $G_q(x,y)$ is from an M2ME (p,q).

– By utilizing the average coordinate of two M2MEs in the grid and another bivariate polynomial e(x,y) which is pre-distributed in all components of the system, two M2MEs can generate the pairwise key between them at each side. The process can be seen in Fig. 5.

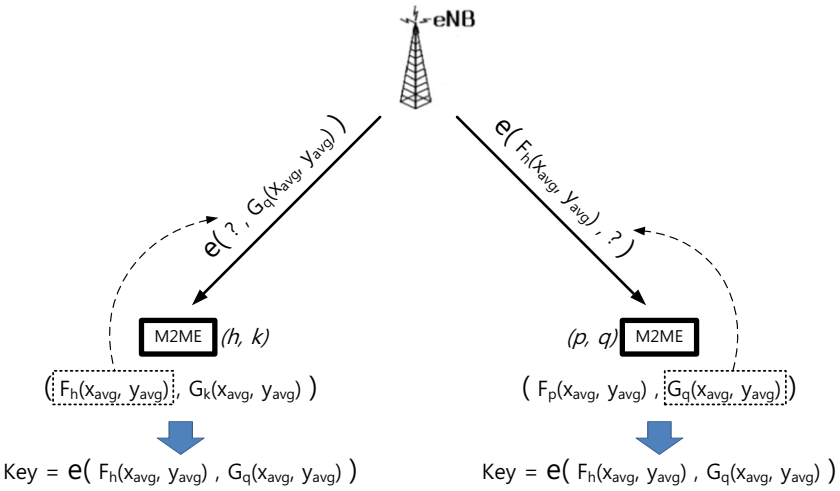



**Fig. 5.** Pairwise key generation process between M2MEs

An eNB replaces the variable y in the bivariate polynomial e(x, y) by the value of $G_q(x_{avg}, y_{avg})$ and transmits the result polynomial with only one unknown x to M2ME (h, k), where Gq is the polynomial selected from polynomial share of an M2ME (p, q) as mentioned above and $(x_{avg}, y_{avg})$ stands for the average coordinate of M2MEs (h, k) and (p, q), the calculation is as follows:

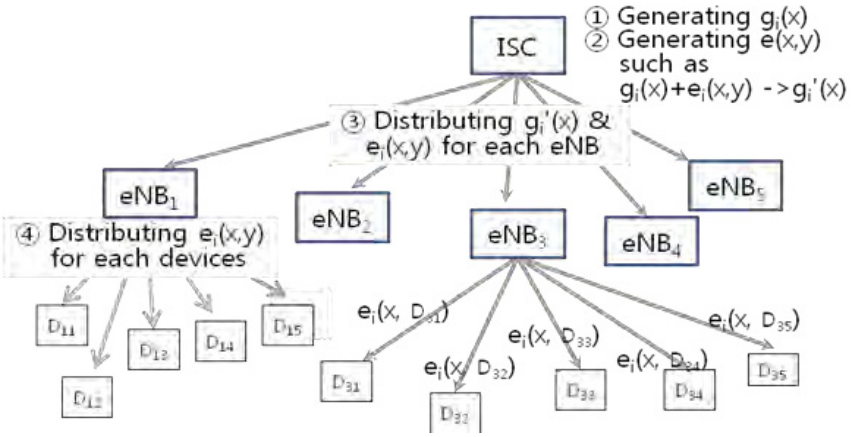
$$(x_{avg} , y_{avg}) = ((h+p)/2, (k+q)/2) \quad (1)$$

After receiving the polynomial $e(x, G_q(x_{avg}, y_{avg}))$ with unknown x, an M2ME (h, k) replaces the variable x by the value of $F_h(x_{avg}, y_{avg})$ which is calculated by its own polynomial $F_h$ selected in step 2 and the average coordinate of two M2MEs, then an M2ME (h, k) can obtain the pairwise key, $e(F_h(x_{avg}, y_{avg}), G_q(x_{avg}, y_{avg}))$ shared with an M2ME (p, q). For an M2ME (p, q), it can also calculate the pairwise key by operating the similar process. According to the calculation above, M2MEs (h, k) and (p, q) can generate their pairwise key as:

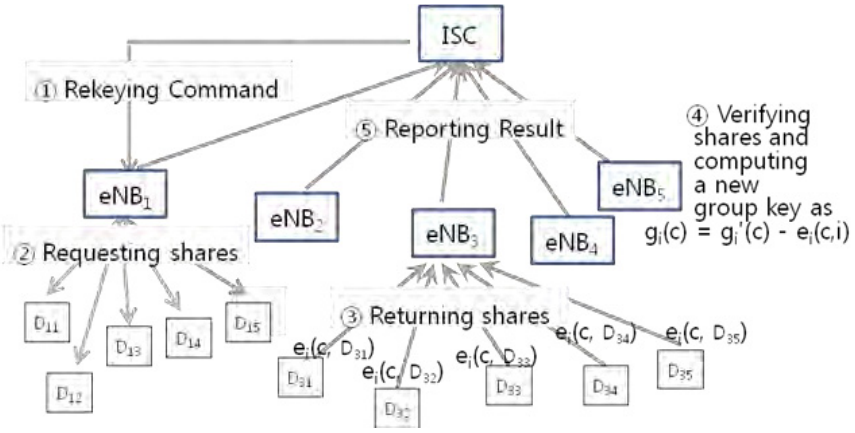
$$\text{Key} = e( F_h(x_{avg}, y_{avg}) , G_q(x_{avg}, y_{avg}) \quad (2)$$

### 4.3. Group Key Establishment among M2MEs for Group Communication

The group based policing and addressing are required in the cellular M2M communication. The network shall enable the broadcast to a specific group of devices. In our previous work, we proposed an energy-efficient and secure channel group key establishment and rekeying management scheme for mobile IPTV services [9, 22].




(a) Group key initialization flow among ICS, eNB, and Devices




(b) Group key rekeying for all devices

**Fig. 6.** Group key management based on PCGR

It adopted Pre-distribution and local Collaboration-based Group Rekeying (PCGR), a group key management scheme for sensor networks [14]. We basically considered the cellular network environment where many mobile devices are provided IPTV services through eNBs and an ISP (Internet Service Provider). Because the mechanism is to generate group keys for the group communication and to rekey the group keys is can be efficiently adopted for the cellular M2M group communication. Its process is shown in Fig. 6 and here is the brief description.

- ISC generates the channel key polynomials, g(x)s for each channel and encryption polynomials e(x,y)s. ISC then distributes encryption polynomials e(x,y)s and encrypted polynomials g'(x)s to each eNB under the channel service.
- After receiving the polynomial information, the eNB distributes the shares of the encryption polynomials of its own to its member nodes and deletes the original polynomial information.
- On rekeying time, the eNB gathers computed shares from its devices to compute the new keys. When the verification of the shares from the devices is successful, the eNB computes the new group key with the shares and distributes the new group key to its members.

The details are omitted here. One more important advantage of our proposal is that we can reuse the polynomials distributed in the pairwise key setup phase to compute group keys. In the case, not only the communication overhead, but also the computation and storage overhead can be decreased.

Group keys can be classified into functional group keys and regional group keys based on the situation where they are used. They do not have any difference in the establishing or rekeying process. When nodes need any functional group communication even if they are located in physically different regions, they generate keys and share them among functional group members. The difference between functional group keys and regional group keys are shown in Fig. 7.

**Functional Group Key**. When the group communication is required for specific functions among M2MEs, group keys are to be established among M2MEs which accomplish the functions. In this case, M2MEs can be scattered in many cells. For example, some of the M2MEs need to provide specific data or need to play the role of relaying for other M2MEs. In this case, several designated M2MEs require group keys and even if they are mobile, memberships are not often changed. Even if the M2ME moves to another cell, an MME can manage their locations if only the M2ME maintains the group membership. When new M2MEs move in to the specific cell, the eNB of the cell notifies it to the MME and the location information is managed by the MME while the group membership and the group key are not changed. It is because the M2ME is still the functional group member even if its location is changed.

**Regional Group Key.** The regional group membership is related to the specific region of the network. In the regional group, the group membership could be changed depending on the policy and the mobility of the M2MEs. The ratio of M2MEs to eNB in the regional group is higher than that of functional group membership. The overhead for managing the group keys can be decreased when eNBs provide the secret share and the

new group keys generated are just distributed to M2MEs by the eNBs. After generating the new regional group key, eNB distributes the new key encrypted with the old group key to each group member M2ME. When the M2ME moves out of the regional group, it is notified to the MME, and the group keys can be rekeyed according to rekeying policy. The process is as in Fig. 7.

- When an M2ME moves in the new cell and the M2ME is still in the regional group area, the eNB asks if the M2ME wants to get the group service.
- If the answer is yes, the eNB notifies this to the MME and the MME just modifies the location information without rekeying the group key because the M2ME is still the group member.
- Otherwise, eNB notifies the answer to the MME, and the MME decides if the group key should be rekeyed or not.
- If rekeying is required, the MME requests the key share to the eNBs.
- The eNB replies with the key shares.
- The MME computes the new group key, sends it back to eNBs. Finally, eNBs distribute the new group key to each member devices in each cell.

When the devices move into another cell in the regional group keying or if there is any change in their subscription, group keys need to be rekeyed right away for the security protection while a pairwise key between a pair of M2MEs keeps for certain period because they can communicate with each another again in short time difference.

## 5. Performance Analysis

When an M2M is coupled with the cellular communication, the cellular network security mechanisms can be basically applied. To the best of our knowledge, there is no key agreement mechanism proposal for cellular M2M group communication. Even through traditional security mechanisms in cellular network can be applied, direct M2M communication has different characteristics and different security mechanisms are required. It is not possible that we compare our key agreement proposal with other mechanisms because there is no proper one to be compared. We would like to analyze our proposal to show how it is efficient. We also consider the communication, computation overhead, and security aspects for cellular M2M communications related to proposed key establishment.

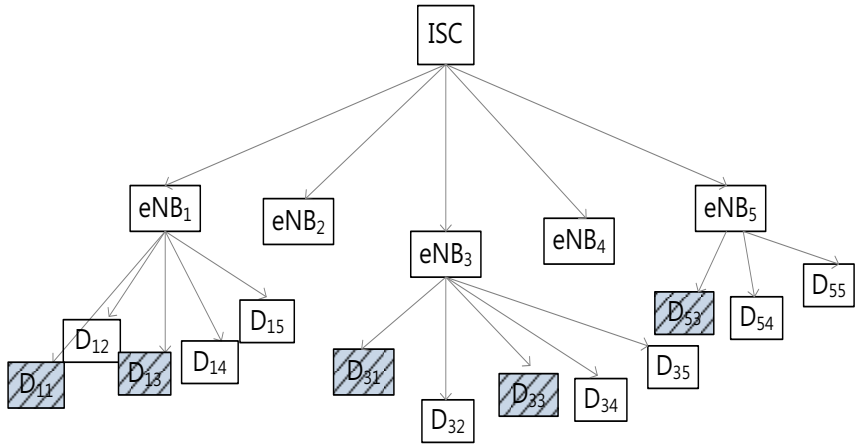
### 5.1. Simulation Result

In Fig. 8, we can see the communication time between a pair of M2MEs. When they are located in its own communication range, they can talk to each other in direct mode. Communication time in direct mode is much shorter than the case in which they communicate passing through the eNB.

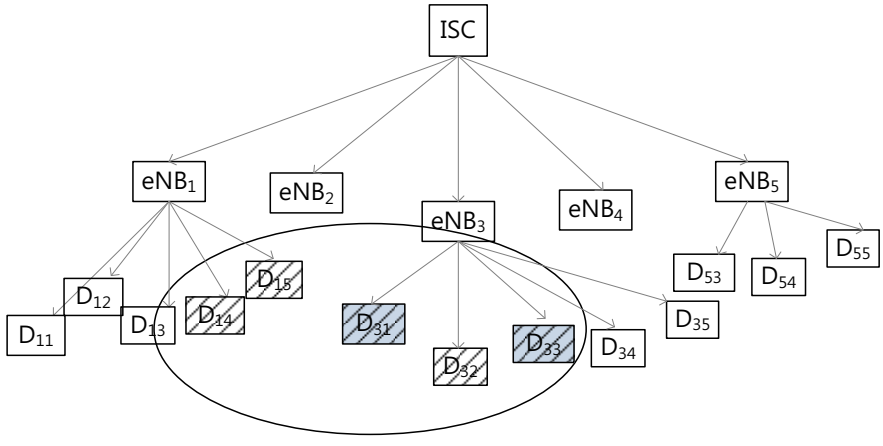
Fig. 9 shows the energy usage of M2MEs and the eNB in M2ME direct communication. Sending M2ME consumes more energy than the receiving M2ME, and of course eNB consumes basic energy for its own function as the base station, while

eNB in indirect mode consumes more energy than the M2MEs because it needs to relays the data in between as in Fig. 10.

In Fig.11, we can see that key information is sent by the eNB to each M2MEs, and the receiving energy of the M2MEs increases a little, while energy consumption of eNB for sending data increases. It shows that M2MEs do not consume much energy for key information distribution.




(a) Functional Group




(b) Regional Group

**Fig. 7.** Two different kinds of groups for the efficient group key management

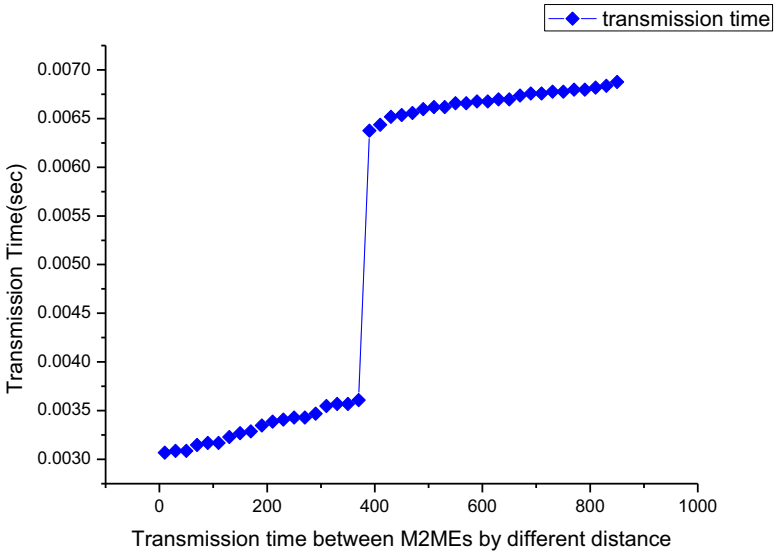



**Fig. 8.** Transmission time between M2MEs as a function of distance

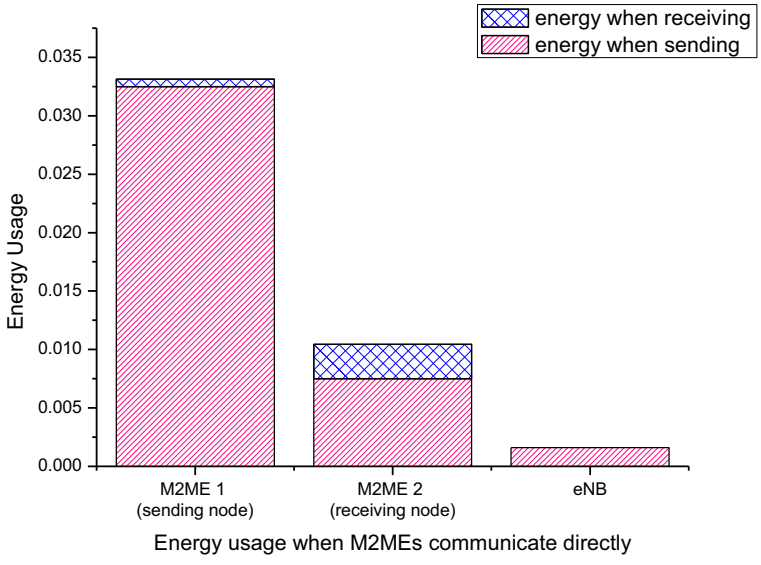



**Fig. 9.** Energy consumption in direct mode M2ME communication

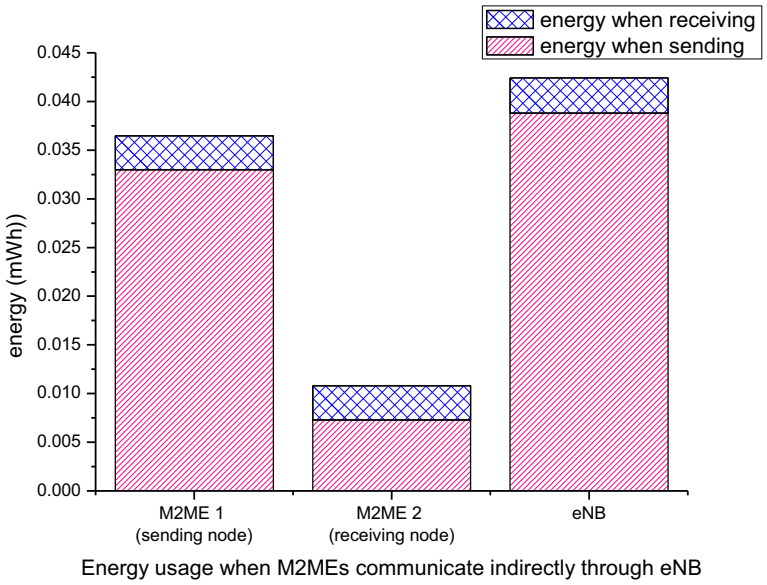



**Fig. 10.** Energy consumption in indirect mode M2ME communication

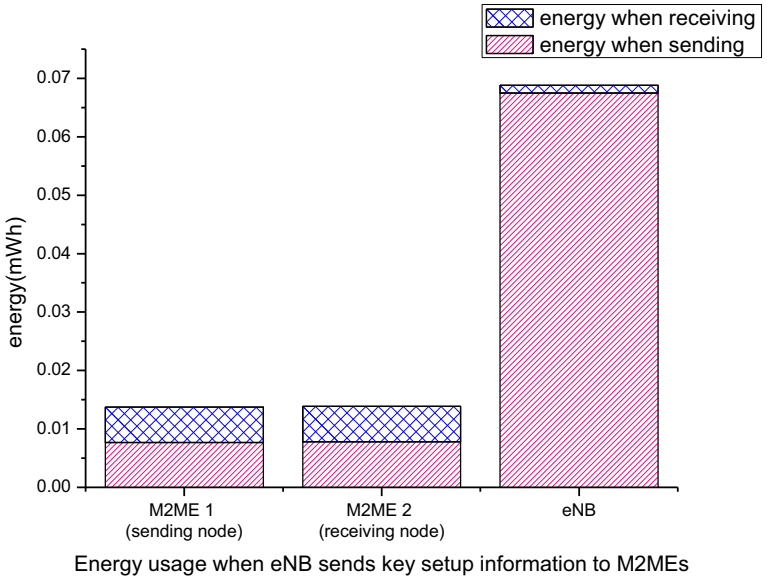



**Fig. 11.** Energy consumption of eNB and M2MEs for key information delivery

### 5.2. Communication and Computation Analysis

In our previous work [11], we have proposed key establishment and authentication mechanism based on USIM card for ubiquitous healthcare system. In the mechanism, we proposed that when direct M2M communication is required, they request key information to their own eNBs, and the eNBs generate a pairwise key between two devices through cooperation. The problem is that when more and more pairs of devices request direct communication, key generation overhead is getting heavy on the eNBs. In addition, key management is getting difficult and complicated.

In our newly proposed mechanism, all the M2MEs are pre-distributed the e(x,y) in the setup stage, and when they want to communicate with one another, they are additionally distributed two more polynomials to compute pairwise keys. With the polynomials, they can make pairwise keys with another M2ME no matter how many communication partners they may have. Because M2MEs can compute pairwise keys and communicate with one another, the overload of the eNBs is getting lighter, and the management is also very simple.

For communication overhead, the M2MEs request key shares to eNBs, and once they get two more polynomials for direct communication, they don't need to request keys to eNBs but can calculate their own keys. When there are a lot of pairs wanting direct communication, our proposal decreases the communication overhead in great amount.

For computation and storage overhead, each M2MEs need to store three polynomials for setting up the pairwise keys. However, only coefficients are delivered and the storage required is not big. In addition, the computation is very simple, and it does not cost much for mobile M2MEs. In addition, using the polynomials distributed, group keys can be computed and the overall computation and storage overhead can be lowered further.

### 5.3. Security Analysis

In this subsection, we consider the security aspect of our proposal. As described in 3.2, in our proposal, there are two intersection points, and one of the points is chosen to setup the pairwise key. This increases the security level because even if some security information revealed, the attackers have 50% chance to compute the pairwise keys. Especially, periodic redistribution of polynomials makes the security level high.

**Confidentiality.** In cellular M2M communication, personal information such as location, account data, the content of the data can be revealed if the data are not encrypted. For encrypting the data, traffic encryption keys are used. In our work, we have proposed the pairwise key agreement between M2MEs and eNBs or between M2ME communications. We also proposed the group key establishment process for the secure group communication. Even the attackers would eavesdrop on the data using the keys properly, the confidentiality could be achieved.

**Authentication.** Basically, a machine needs to authenticate the other entities before their communication. In many cases, they need to mutually authenticate each other. In

our proposal, by adopting the algorithms in the USIM card, the device and an eNB can mutually authenticate each other. For the communication between the devices, additional authentication process is required.

**Access Control.** For the devices to get the access to the network, they need a process for getting the admission. The process is out of the scope or our work. However, through the admission step in cellular network, access can be controlled by the eNBs, and basic key related information can be acquired for further security functions.

**Integrity.** Integrity is required for keeping data from being forged or modified by the attackers. The keys from our proposal can be used for encrypting the data and the data can be decrypted only by the receiver. If pairwise keys could be delivered by the eNB, and the eNB could be not compromised, integrity could be obtained.

**Privacy.** In many cases, M2MEs are deployed closely to human beings. The data can contain very personal information which is not supposed to be disclosed. These days, privacy is one of the major security issues to be protected. Privacy protection is one of our future works.

## 6. Conclusions

More and more M2MEs are connected to traditional infrastructures in wired or wireless environments. Especially, connection between cellular network and M2M equipment is expected to bring great impacts and the market share in the future network. When M2MEs communicate with one another in the cellular infrastructure, the possibility of security breaches is getting higher while the great deal of application services are provided. In this work, we proposed key establishment mechanisms for secure communication among entities in the cellular M2M network. The mechanism includes pairwise keys for the M2M communication and the group communication among the M2MEs. Our key agreement proposal can provide security and reliability for the cellular M2M communication.

**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 (2012R1A1A3019459).

## References

1. Cha, I., Shah, Y., Schmidt, A. U., Leicher, A., and Meyerstein, M.: Trust in M2M communication. IEEE Vehicular Technology Magazine, Vol.4, Issue 3, pp. 69-75. (2009)
2. 3G machine-to-machine (M2M) communications: Cellular 3G, WiMAX, and municipal Wi-Fi for M2M applications. Technical report, ABI Research (2007)
3. Ryberg T.: The global wireless M2M market. Technical report, Berg Insight (2009)

4. Fledderjohn, D.: Learn Cellular M2M Basics. Field Tchnologies Online
5. M2M Technology and Services of KT, KNOM Tutorial (2011)
6. Cellular M2M Connectivity Services - Research Report by ABI Research (2012)
7. Shafiq, M. Z., Ji, L., Liu, A. X., Pang, J., and Wang J.: A First Look at Cellular Machine-to-Machine Traffic – Large Scale Measurement and characterization. SIGMETRICS'12, June pp. 11-15 (2012)
8. Dohler, M., Watteyne, T., Alonso-Zárate, J.: Machine-to-Machine: An Emerging Communication Paradigm. Mobilight 2010, MONAMI 2010, PIMRC 2010, Globecom 2010 (2010)
9. Doh, I., Lim, J., and Chung, M.: Group Key Management for Secure Mobile IPTV Service. In Proceedings of Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, pp. 352-357 (2012)
10. 3GPP TR 33.812, [Online]. Available: http://www.3gpp.org/DynaReport/33812.htm (current Jun 2014)
11. Doh, I., Lim, J., and Chae, K.: Key establishment and management for Secure Cellular Machine-to-Machine Communication. In Proceedings of the Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, pp. 579-584 (2013)
12. Blundo, C., Santis, A. D., Herzbeerg, A., Kutten, S., Vaccaro, U., and Yung, M.: Perfectly-Secure Key Distribution for Dynamic Conference. Information and Computation, Vol. 146, Issue 1, pp. 1-23 (1998)
13. Wang, Y., Ramamurthy, B., and Xue, Y.: Group Rekeying Schemes for Secure Group Communication in Wireless Sensor Networks. Proceedings of the IEEE International Conference on Communications, pp. 3419-3424 (2007)
14. Zhang, W., and Cao, G.: Group Rekeying for Filtering False Data in Sensor Networks: A Predistribution and Local Collaboration Based Approach. IEEE Infocom 2005, Vol. 1, pp.503-514 (2005)
15. Huang, J. H., Buckingham, J., and Han, R.: A Level Key Infrastructure for Secure and Efficient Group Communication in Wireless Sensor Networks. Proceedings of The International Conference on Security and Privacy for Energing Areas in Communications Networks, pp. 249-260 (2005)
16. Zhu, S., Setia, S., and Jahodia, S.: LEAP+: Efficient Security Mechanisms for Large-Scale Disributed Sensor Networks. ACM Transactions on Sensor Networks, Vol. 2, Issue 4, pp. 500-528 (2006)
17. Adusumilli, P., Zou, X., and Ramamurthy, B.: DGKD: Distributed Group Key Distribution with Authentication Capability. Proceedings of the IEEE Workshop on Information Assurance and Security, pp. 276-293 (2005)
18. Aparna, R., and Amberker, B. B.: Key management scheme for multiple simultaneous secure group communication. Proceedings of the IEEE Internet Multimedia Services Architecture and Applications (IMSAA), pp. 1-6 (2009)
19. Kim, Y., Perrig, A., and Tsudik, G.: Tree-based group key agreement. ACM Transactions on Information and System Security (TISSEC), Vol. 7, Issue 1, pp. 60-96 (2004)
20. Yu, Z., andGuan, Y.: A Robust Group-based Key Management Scheme for Wireless Sensor Networks. Proceedings of the IEEE Communications Society 2005, Vol. 4, pp. 1915-1920 (2005)
21. Park, J., Doh, I., and Chae, K.: Security Approach for Ubiquitous Healthcare Services through Wireless Communication. In Proceedings of ACSA 2012, pp. 381-385 (2012)
22. Doh, I., Lim, J., and Chae, K.: Key Management Approach for Secure Mobile Open IPTV Service. Computer Science and Information Systems 2013, Vol. 10, pp. 843-864 (2013)

**Inshil Doh** received the B.S. and M.S. degrees in Computer Science at Ewha Womans University, Korea, in 1993 and 1995, respectively, and received the Ph.D. degree in Computer Science and Engineering from Ewha Womans University in 2007. From 1995-1998, she worked in Samsung SDS of Korea to develop a marketing system. She was a research professor of Ewha Womans University in 2009~2010 and of Sungkyunkwan University in 2011. She is currently an assistant professor of Computer Science and Engineering at Ewha Womans University, Seoul. Her research interests include wireless network, sensor network security, and M2M network security.

**Jiyoung Lim** is the corresponding author of this paper. She received the B.S. and M.S degrees in Computer Science at Ewha Womans University, Korea, in 1994 and 1996, respectively and received the Ph.D. degree in Computer Science and Engineering from Ewha Womans University in 2001. She is currently an associate professor of Computer Software at Korean Bible University, Seoul, Korea. Her research interests include wireless/sensor network security, and M2M network security.

**Shi Li** received the B.S. degree in the Department of computer science and engineering from Harbin Institute of Technology, China in 2010. She is currently a Ph.D candidate in the Department of computer science and engineering at Ewha Womans University, Seoul, Korea. Her research interests include sensor network security, smart grid security and content delivery network security.

**Kijoon Chae** received the B.S. degree in mathematics from Yonsei University in 1982, an M.S. degree in computer science from Syracuse University in 1984, and a Ph.D degree in Electrical and computer engineering from North Carolina State University in 1990. He is currently a professor of Computer Science and Engineering at Ewha Womans University, Seoul, Korea. His research interests include network security, sensor network, network protocol design and performance evaluation.

*Received*: *September 22*, *2013*; *Accepted*: *February 18*, *2014.*

Computer Science and Information Systems 11(3):1091–1112 DOI: 10.2298/CSIS130924066C

# A Secure E-Mail Protocol Using ID-based FNS Multicast Mechanism

Hsing-Chung Chen[1], Cheng-Ying Yang[2], Hui-Kai Su[3], Ching-Chuan Wei[4], and Chao-Ching Lee[1]

[1] Department of Computer Science and Information Engineering,
Asia University, Taichung 413, Taiwan
shin8409@ms6.hinet.net, cdma2000@asia.edu.tw
[2] Department of Computer Science, University of Taipei,
Taipei 100, Taiwan
cyang@uTaipei.edu.tw
[3] Department of Electrical Engineering,
National Formosa University, Yunlin 632, Taiwan
hksu@nfu.edu.tw
[4] Department of Information and Communication Engineering,
Chaoyang University of Technology, Taichung 413, Taiwan
ccwei@cyut.edu.tw

**Abstract.** Electronic mail (e-mail) has been used to transfer various types of electronic data in Internet. Usually, a user has to send an e-mail to a specific group of users with a secure delivery mechanism. In this paper, a novel and feasible e-mail delivery mechanism using the secure multicast protocol with an ID-based factorial number structure (FNS) is proposed in the multicast system. In the proposed e-mail delivery mechanism, the e-mail is required to be encrypted before sending out in order to safeguard the message via a public channel, such as wire public switching communication links and wireless communication systems. Without loss generality, the public-key system is adopted in the proposed secure multicast system for a convenient and easy key management. The proposed scheme outperforms the existing methods for more easily to construct secure e-mail system. Furthermore, the security of the proposed scheme is analyzed, including replay attack, sender impersonation attack, unknown key-share attack, forgery attack and insider attack. Finally, the computation complexities of the proposed mechanism are discussed. The result shows that the proposed scheme outperforms the CRT-based secure e-mail scheme.

**Keywords:** factorial number structure, e-mail, security, cryptography.

## 1. Introduction

People widely use electronic mails (e-mails) to communicate with each other in Internet. Delivering an e-mail in Internet, people could exchange not only normal text-based letter, but also sensitive rich electronic files. Because of the popularity, e-mail systems become an adversary's or a malicious user's targets. Among the e-mail security issues, basic and primary concerns are the confidentiality and authentication for the e-mails [1].

Some data cryptosystems [2] can satisfy those concerns. Users can utilize a specific interactive key to encrypt and to verify their e-mails. However, the e-mail system is a kind of store-and-forward system in which e-mail servers act as a proxy to accept, forward, and store users' e-mails. User does not need continuously on-line to connect with an e-mail server. When a user wants to get the emails that are received and stored in the server, he/she has to access the email server first. For example, sender $B$ intends to send an e-mail to receiver $A$. Sender $B$ firstly sends the e-mail to the mail server $S_B$, and then the mail server $S_B$ forwards the e-mail to receiver $A$ 's mail server $S_A$. Following, the mail server $S_A$ stores the e-mail in the storage. As receiver $A$ connects to the e-mail server $S_A$, receiver $A$ sends a request for new e-mails, and the mail server $S_A$ forwards the stored e-mail to receiver $A$. Obviously, e-mail users are not always on-line. However, the e-mail users could not exchange the session key in time within a secure on-line system. To solve this difficulty, there are several challenges, such as authentication and secure key distribution [2], to mail server. Public key systems could provide a solution but need much time to deal with encrypt or decrypt. The hybrid cryptosystems to prevent the high computation is also provided [3]. Another more efficient solution is provided by Pretty Good Privacy (PGP) protocol.

PGP was designed and implemented for distributed networks in 1991. It is a well-known secure e-mail protocol that provides confidential data between senders and receivers. It is available on almost any platform which aims to be used within existing e-mail systems [4], [5], [6], [7]. PGP protocol [8] utilizes the idea in the hybrid cryptosystems to securely transfer a session key to both of the corresponding sender and receiver. A sender in the PGP system is given a certificated public key. The certificated public key can be applied to a secure channel to transfer the session key within the session key is used for encrypting the emails between the sender and the receiver. A user cannot verify the validity of PGP keys for each other. However, under many circumstances, a sender needs to send a single email to each other. Hence, how to transfer a session key to multi-receivers is a challenge for securing e-mail systems. Hung-Min Sun et.al. [9] proposed two novel e-mail protocols to provide a perfect forward secrecy. The basic protection in an e-mail system is to encrypt the bulk mail using a conventional cryptosystem with a short-term key and to protect the short-term key using a public-key cryptosystem with the receiver's public key. Amna Joyia, et.al. [10] found that an attacker can easily track from email header which are normally transported in clear text. Furthermore, this information can be manipulated for malicious purposes like sending spam messages to the extracted user identities, analyzing traffic to extract the behavior of both sender and receiver. All these attacks lead to vivid threat to the user's privacy. Then, he designed and implemented a secure and privacy enhanced email system which provided the solution to ensure the privacy of e-mail users. However, in the multicasting system, it uses PGP scheme to send e-mail for lots of specific receivers. It has to send the e-mail one-by-one. For example, as a user usually needs to send an e-mail to a group of users, in the exiting e-mail protocols such as Simple Mail Transfer Protocol (SMTP), the e-mail server forwards the copies of this e-mail to the receivers. Intending to deliver an e-mail to receivers $A$, $C$ and $D$, receiver $B$ initially sends an e-mail to the mail server $S_B$. Then, the mail server $S_B$ forwards the copies of this e-mail to the mail servers $S_A$, $S_C$ and $S_D$ for

receivers, respectively. Next, the mail servers $S_A$, $S_C$ and $S_D$ wait for the request for the new e-mails from the receivers. For example, if $S_C$ receives a request sent by the receiver $C$, $S_C$ forwards the copy of the e-mail to the receiver $C$. In the repeatedly transmission, there exists a redundant computation which causes a significant delay. Hence, to send e-mails in the multicast system, it has to look for another efficient solution.

In 1985, the Identity-Based Cryptosystems and Signature Schemes were first proposed by Adi Shamir [11]. A novel type of cryptographic scheme was proposed to enable any pair of users to communicate securely. In 2005, McCullagh, N. [12] proposed another solution for secure e-mail with identity-based encryption. It could allow an arbitrary string of characters and numbers to serve as a public key. It had some effects in simplifying public-key encryption. In 2010, Anastasios Kihidis et.al. [13] presented a complete implementation of a practical Identity Based Encryption (IBE) infrastructure for secure e-mail communication. It attempted to simultaneously provide a fully functional and user-friendly IBE system. A packet construction mechanism using an ID-based factorial number structure (FNS) was proposed by Chen H.C. [14], [15], [16] for a secure system to provide a feasible solution for a secure multicast system. In 2010, Zhang M.Q. et. al. [17] presented a secure and efficient ID-based fair multi-party exchange protocol with off-line semi-trusted third party. Application of multi-receiver identity-based encryption. In 2013, Mingwu Zhang et.al. [18] proposed an efficient anonymous multi-receiver encryption scheme to achieve the security properties of confidentiality and anonymity. The anonymity of the proposed scheme could securely against outer attackers and inner attackers simultaneously, and also presented a dual-anonymous multi-receiver encryption that could support the security properties such as identity privacy of both sender and receiver.

In 2013, Chen H.C. [15] proposed a secure multicast protocol for e-mail systems. A user usually needs to send an e-mail to a group of users. The proposed secure multicast protocol [15] for e-mail systems could provide perfect forward secrecy to ensure confidentiality and authentication. The protocol [15] employs the Chinese Remainder Theorem (CRT), RSA public key cryptosystems, and one-way hash functions. The protocol can save redundant key materials used for the e-mails. However, CRT will take a very long time in the calculation to factor for a large integer. In this paper, a secure multicast key protocol is proposed a solution to the e-mail systems for distributing a session key to the specific group. Due to the concerning for the securely transferring the session key, the proposed protocol adopts ID-based FNS [14] to replace CRT [15] is proposed. The scheme is based on ID-based FNS [14] with the hybrid cryptographic algorithms of public-key and secret-key system [19-25]. In the manner, not only the e-mail construction in multicast system can be efficiently retained, but also the easy key management and fast computation [21], [22], [23], [24], [25] to process a multiple secure e-mail delivery can be proficiently achieved. The proposed protocol benefits for an excellent secure broadcast e-mail system [25]. The rest of this paper is organized as the followings, the fundamental theory of the ID-based FNS are addressed in Section 2. Two scenarios consist of corresponding schemes are proposed in Section 3. Security and complexity analyses are described in Section 4. Finally, conclusions are given in Section 5.

## 2. Fundamental Theory of the ID-based FNS

ID-based FNS [15] in a secure multicast key scheme begins with Lemma 1.

**Lemma 1.** $P^*_{(i,0)}, P^*_{(i,1)}, \ldots, P^*_{(i,j)}, \ldots, P^*_{(i,k)}, \ldots, P^*_{(i,m-1)}$ be positive integers, where $P^*_{(i,j)} \neq P^*_{(i,k)} \neq 0$, $0 \leq j$, $k \leq m-1$, $j \neq k$. The values of $P_{(i,0)}, P_{(i,1)}, \ldots, P_{(i,m-1)}$ are gotten as the followings, $P_{(i,o)} = \sum_{j=0}^{m-1} p^*_{(i,j)}$, $P_{(i,1)} = \sum_{j=1}^{m-1} p^*_{(i,j)}$, …, $P_{(i,m-1)} = \sum_{j=m-1}^{m-1} p^*_{(i,j)} = p^*_{i,m-1}$ such that the inequality relation of $P_{(i,0)} > P_{(i,1)} > \ldots > P_{(i,m-2)} > P_{(i,m-1)}$ is satisfied.

□

**Theorem 1.** Let $P^*_{(i,0)}, P^*_{(i,1)}, \ldots, P^*_{(i,m-1)}$ be positive integers, where $P^*_{(i,j)} \neq P^*_{(i,k)} \neq 0$. And, the values of $P_{(i,0)}, P_{(i,1)}, \ldots, P_{(i,m-1)}$ are obtained by Lemma 1, respectively. There exists a positive integer, $Z_i = \sum_{j=0}^{m-1} \alpha_{(i,j)} p_{(i,j)}$, such that the individual positive integer can be retrieved by the equation,

$$\begin{aligned} p^*_{(i,j)} &= p_{(i,j)} - p_{(i,j+1)} \\ &= \left\{ \left[ \frac{Z_i}{\alpha_{(i,j)}} \right] mod(T_i - x_{(i,j-1)}) \right\} - \left\{ \left[ \frac{Z_i}{\alpha_{(i,j+1)}} \right] mod(T_i - x_{(i,j)}) \right\}, \end{aligned} \tag{1}$$

where some notations are defined as followings.

$T_i = max\{P_{(i,0)}, P_{(i,1)}, \ldots, P_{(i,m-1)}\}$, $\alpha_{(i,0)} = \prod_{j=m-2}^{0} (T_i - x_{(i,j)})$, $\alpha_{(i,1)} = \prod_{j=m-2}^{1} (T_i - x_{(i,j)})$, … , and $\alpha_{(i,m-2)} = \prod_{j=m-2}^{m-2} (T_i - x_{(i,j)})$, $\alpha_{(i,m-1)} = 1$,

$I_0, I_1, \ldots, I_i, \ldots, I_{m-1}$ : The Identity numbers that are corresponding to the positive integers $P^*_{(i,0)}, P^*_{(i,1)}, \ldots, P^*_{(i,m-1)}$ respectively. The numbers of $I_i, i = 0,1,\ldots,m-1$, are pre-sorted by decreasing order as the relations: $I_0 > I_1 > \ldots > I_i > \ldots > I_{m-1}$.

$x_{(i,-1)} = -1$ , $x_{(i,0)} = \sum_{j=0}^{m-1} I_j$ , $x_{(i,1)} = \sum_{j=1}^{m-1} I_j$ , …, and $x_{(i,m-1)} = \sum_{j=m-1}^{m-1} I_j$ that satisfy $(p_{(i,0)} - x_{(i,j)}) >> p_{(i,j)}$ for $j \in \{1,2,\ldots m-1\}$.

As observed Theorem 1, the list of generated $p^*_{(i,j)}$, $j = 0,1,2,\ldots m-1$ is shown as the followings,

$$p^*_{(i,0)} = p_{(i,0)} - p_{(i,1)} = \sum_{j=0}^{m-1} p^*_{(i,j)} - \sum_{j=1}^{m-1} p^*_{(i,j)},$$

$$p^*_{(i,1)} = p_{(i,1)} - p_{(i,2)} = \sum_{j=1}^{m-1} p^*_{(i,j)} - \sum_{j=2}^{m-1} p^*_{(i,j)},$$

$$\vdots$$

$$p^*_{(i,m-2)} = p_{(i,m-2)} - p_{(i,m-1)} = \sum_{j=m-2}^{m-1} p^*_{(i,j)} - \sum_{j=m-1}^{m-1} p^*_{(i,j)},$$

$$p^*_{(i,m-1)} = p_{(i,m-1)}.$$

□

The fact of *Theorem 1* can be seen in the Appendix of Ref. [15]. Let ID-based FNS be more easily readable and, therefore, Example 1 be given as below.

**Example 1.** Assume that there exists five identical numbers which are sorted by the decreasing order, $I_0 = 31$, $I_1 = 29$, $I_2 = 23$, $I_3 = 12$ and $I_4 = 9$, where these numbers are the published identification numbers for the users, $u_0$, $u_1$, $u_2$, $u_3$ and $u_4$, respectively, in the communication group. Assume that participant $u_0$ whose identical number is $I_0$ wants to send a secure multicast key to $u_1$, $u_2$, $u_3$ and $u_4$ by a broadcast mechanism. Following, $u_0$ will choose the positive integers $p^*_0 = 98$, $p^*_1 = 123$, $p^*_2 = 65$, $p^*_3 = 72$ and $p^*_4 = 132$ corresponding to $I_0$, $I_1$, $I_2$, $I_3$ and $I_4$, respectively. Then, according to *Lemma 1*, $u_0$ computes the values *of* $P^*_0, P^*_1, ..., P^*_4$ *as the followings,* $p_0 = \sum_{i=0}^{4} p^*_i = 490$, $p_1 = \sum_{i=1}^{4} p^*_i = 392$, $p_2 = \sum_{i=2}^{4} p^*_i = 269$, $p_3 = \sum_{i=3}^{4} p^*_i = 204$, $p_4 = \sum_{i=4}^{4} p^*_i = 132$ individually, such that $p_0 > p_1 > p_2 > p_3 > p_4$ is satisfied. In order to pack the five numbers, $p_0, p_1, p_2, p_3, \text{and } p_4$, into a fixed integer $Z_0$, the accumulative operation in the decreasing order is launched. Therefore, $u_0$ obtains $x_0 = \sum_{i=0}^{4} I_i = 104$, $x_1 = \sum_{i=1}^{4} I_i = 73$, $x_2 = \sum_{i=2}^{4} I_i = 44$, $x_3 = \sum_{i=3}^{4} I_i = 21$, $x_4 = \sum_{i=4}^{4} I_i = 9$, respectively.

Moreover, $u_0$ calculates $T = max\{p_0, p_1, p_2, p_3, p_4\} = 490$. The $\alpha_i$ for $i = 0, 1, 2, 3, 4$ is defined by Equation (1) and found as the follows:

$$\alpha_0 = \prod_{i=3}^{0} (T - x_i) = 33669065388,$$

$$\alpha_1 = \prod_{i=3}^{1} (T - x_i) = 87225558,$$

$$\alpha_2 = \prod_{i=3}^{2} (T - x_i) = 209174,$$

$$\alpha_3 = \prod_{i=3}^{3} (T - x_i) = 469,$$

and $\alpha_4 = 1$.

According to the given results: $\alpha_i$'s and $p_i$'s values for $i = 0, 1, 2, 3, 4$, the fixed integer $Z = \sum_{i=0}^{4} \alpha_i p_i = 16532090822347$ is constructed. Then, sender $u_0$ sends the packet $\{16532090822347 \,||\, 490\}$ to $u_1$, $u_2$, $u_3$ and $u_4$ by using a broadcast mechanism.

Next, according to *Theorem 1*, each one of the participants calculates the values for $x_{-1}, x_1, x_2, x_3$ and $x_4$ by using the published identical numbers and attempts to directly extract the $p_i$ values from the summed $Z_0$ by using Equation (1).

$$p_0 = \left\lfloor \frac{Z}{\alpha_0} \right\rfloor mod(T - x_{-1}) = 490,$$

$$p_1 = \left\lfloor \frac{Z}{\alpha_1} \right\rfloor mod(T - x_0) = 392,$$

$$p_2 = \left\lfloor \frac{Z}{\alpha_2} \right\rfloor mod(T - x_1) = 269,$$

$$p_3 = \left\lfloor \frac{Z}{\alpha_3} \right\rfloor mod(T - x_2) = 204,$$

$$p_4 = \left\lfloor \frac{Z}{\alpha_4} \right\rfloor \mathrm{mod}(T - x_3) = 132.$$

$$p_0 - p_1 = 490\text{-}392 = 98 = p_0^*,$$

$$p_1 - p_2 = 392\text{-}269 = 123 = p_1^*,$$

$$p_2 - p_3 = 269 - 204 = 65 = p_2^*,$$

$$p_3 - p_4 = 204 - 132 = 72 = p_3^*,$$

$$p_4 = p_4^* = 132.$$

When the resulting $p_i^*$ values are recovered from the original ones.

## 3. Proposed Scheme

Some notations are defined and listed in Table 1. Two scenarios and the corresponding schemes are described and designed in Section 3.1 and 3.2, respectively. The first scenario deals with that a sender sends an e-mail to one recipient. For the multicast concerning, the second scenario scopes with that a sender sends an e-mail to specific recipients.

The following notations are used to describe the security protocol and cryptographic operations in this paper.

**Table 1.** Notations

| Notations | Descriptions |
|---|---|
| $U_i$ | The $i$ -th user in the e-mail system |
| $TC$ | The trust center |
| $S$ | The mail server |
| $M$ | A plain-text content of the e-mail |
| $K_c$ | The communication key is randomly generated by the e-mail sender |
| $PK_i$ | A user $U_i$ 's public key |
| $SK_i$ | A user $U_i$ 's secret key corresponding to the $PK_i$ |
| $ID_j$ | A uniquely identifies user $U_i$ where $ID_1 > ID_2 > \cdots > ID_n$ |
| $E_k(M)$ | A asymmetric encryption algorithm using to encrypt the e-mail message $M$ via a secret key $k$ |
| $D_k(C)$ | A asymmetric decryption algorithm using to decrypt the encrypted e-mail message $C$ via a key $k$ |
| $Sig_k(m)$ | A signature algorithm used to generate signature of the message $m$ using the secret key $k$ |
| $h(\cdot)$ | A cryptographically secure one-way hash function |
| $\|\|$ | A catenation symbol |
| $A \rightarrow B$ | A symbol indicates that the certain message sent from the entity $A$ to the entity $B$ |

### 3.1. Scenario I: A sender sends an e-mail to one recipient

A sender sends an e-mail to one recipient. In Fig. 1, it shows that an e-mail is sent from sender B to receiver A, individually. There are three parts in this scenario. The first one is Pre-computation that consists steps, S1 and S2. Another one is Sending phase that describes the steps from S3 to S13. The other one is Receiving phase that illustrates the steps from S14 to S23.

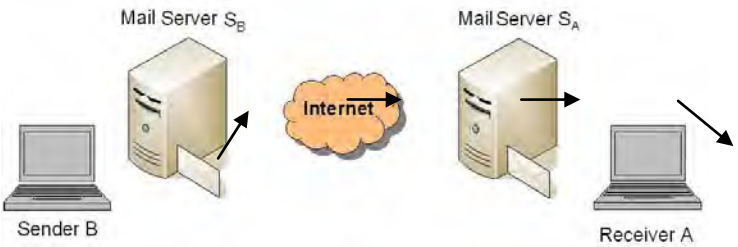



**Fig. 1.** The scenario of an e-mail sent from a sender B to a single receiver A [15]

1) Pre-computation

Step S1: $TC \rightarrow U_i$: $g$, $n$. $TC$ generates randomly a number $g$, and chooses a big prime $n$ sent to the $U_i$. The sender chooses a secret key $k_i$ and computes the public key $e_{k_i} = g^{k_i} \bmod n$. Then, the receiver also chooses a secret key $k_j$ and computes the public key $e_{k_j} = g^{k_j} \bmod n$.

Step S2: $U_i \rightarrow S$: $e_{k_i}$, $e_{k_j}$, $Sig_{SK_i}(e_{k_i})$, $Sig_{SK_i}(e_{k_j})$, $ID_j$. A user $U_i$ generates another pair of public key and secret key $(e_{k_i}, k_i)$ and $(e_{k_j}, k_j)$. This pair of public key and secret key are not related to the pair of public key $PK_i$ and secret key $SK_i$ pre-distributed by the system. The user $U_i$ sends $e_{k_i}$, $e_{k_j}$ and $Sig_{SK_i}(e_{k_i})$, $Sig_{SK_i}(e_{k_j})$ to the e-mail server. Note that this procedure is executed after the user $U_i$ finished receiving an e-mail. Then e-mail server arranges all the $ID_j$ where $ID_1 > ID_2 > \cdots > ID_n$.

2) Sending Phase

Step S3: $S \rightarrow U_1$: $e_{k_1}, e_{k_2}$, $Sig_{SK_1}(e_{k_1}), Sig_{SK_2}(e_{k_2})$, $ID_1, ID_2$.

Step S4: $U_1$ randomly generates the communication key $K_c$.

Step S5: The encrypted e-mail message $C = E_{K_c}(M)$ is encryption under the chosen key $K_c$, where $M$ is the content message of the e-mail.

Step S6: The $p^*_{i,j}$ for each receivers is computed by applying $p^*_{i,j} = E_{k_{i,j}}(K_c)$ for all $j = 1, 2$. The generation of $k_{i,j}$ will follow the rule of $k_{i,j} = (e_{k_j})^{k'_i} \bmod n = (g^{k'_j})^{k'_i} \bmod n = g^{k'_i k'_j} \bmod n$.

Step S7: Each $p^*_{i,j}$ for all $j = 1, 2$ is computing using the following equations, $p_{i,1} = \sum_{j=1}^{2} p^*_{i,j}$, $p_{i,2} = \sum_{j=2}^{2} p^*_{i,j}$, such that the decreasing order relation $P_{i,1} > P_{i,2}$ is satisfied. Moreover $T = p_1$ is set, which the maximal value of the set is of $\{P_{i,1}, P_{i,2}\}$. A polynomial $f(x)$ is then

constructed by the originator as the followings, $f(x) = T + (x - k_{i,1}) \times (x - k_{i,2})$.

Step S8: Then set $\{P_{i,1}, P_{i,2}\}$ is encrypted to be a sub-packet $\Gamma$ by the way of $\Gamma = E_T(p_{i,1} \,||\, p_{i,2})$.

Step S9: Sum up $x_{i,j} = \sum_{j}^{n} ID_j$. Define the initial value $x_{-1} = -1$. Compute $x_{i,1} = \sum_{j=1}^{2} ID_j$, $x_{i,2} = \sum_{j=2}^{2} ID_j$.

Step S10: Compute $\alpha_{i,j}$'s for $j = 1$ to $j = 2$ using the following equations, $\alpha_{i,1} = \prod_{j=2-1}^{1} (T - x_{i,j})$, and $\alpha_2 = 1$.

Step S11: Construct a basic e-mail packet lock $Z$ as the format of $Z = \sum_{j=1}^{2} \alpha_{i,j} p_{i,j}$.

Step S12: Compute a varied e-mail packet lock in bit-wise exclusive-or operation of follow: $L = Z \oplus E_T(T)$.

Step S13: $U_1 \rightarrow S$ : $C$, $L$, $f(x)$, $\Gamma$, $Y$, $t$, where $Y = Sig_{PK_1}(h(ID_1 \,||\, ID_2 \,||\, M \,||\, t))$. The parameter $t$ is a timestamp at that time.

3) Receiving Phase

Step S14: $S \rightarrow U_i$: $C$, $L$, $f(x)$, $\Gamma$, $Y$, $t$, $ID_1, ID_2$.

Step S15: Find the maximal sub-packet by computing $f(k_{i,j}) = T$, and let $\zeta = E_T(T)$.

Step S16: Decrypt the set $\{P_{i,1}, P_{i,2}\}$ by using the following equation: $D_T(\Gamma) = (P_{i,1} \,||\, P_{i,2})$.

Step S17: Find the e-mail packet lock $Z$ by computing $L \oplus \zeta = (Z \oplus E_T(T)) \oplus \zeta = Z$.

Step S18: Sum up $x_{i,j} = \sum_{j}^{n} ID_j$. Let the initial value be $x_{-1} = -1$. Compute $x_{i,1} = \sum_{j=1}^{2} ID_j$, $x_{i,2} = \sum_{j=2}^{2} ID_j$.

Step S19: Compute the $\alpha_{i,j}$'s for $j = 1$ to $j = 2$ using the following equations, $\alpha_{i,1} = \prod_{j=2-1}^{1} (T - x_{i,j})$, and $\alpha_2 = 1$.

Step S20: Compute the sub-packet $p_{i,j}^{*}$ as per following formula,

$$
\begin{aligned}
p_{i,j}^{*} &= p_{i,j} - p_{i,j+1} \\
&= \left\{ \left[ \frac{Z}{\alpha_{i,j}} \right] mod(T - x_{i,j-1}) \right\} \\
&\quad - \left\{ \left[ \frac{Z}{\alpha_{i,j+1}} \right] mod(T - x_{i,j}) \right\},
\end{aligned}
$$

where $j = 1, 2$. $p_{i,1}^{*} = p_{i,1} - p_{i,2} = \sum_{j=1}^{2} p_{i,j}^{*} - \sum_{j=2}^{2} p_{i,j}^{*}$, $p_{i,2}^{*} = p_2$.

Step S21: Decrypt communication key $K_c = D_{k_{i,j}}(p_{i,j}^{*})$.

Step S22: Recover the original content message $M = D_{K_c}(C)$.

Step S23: The $U_2$ computes the value $Y' = Sig_{PK_i}(h(ID_1 \,||\, ID_2 \,||\, M \,||\, t))$ and checks if $Y'$ equals to the value in the signature $Y$.

### 3.2. Scenario II: A sender sends an e-mail to multiple recipients

It shows a sender sends an e-mail to multiple recipients. Fig. 2 shows that the scenario that an e-mail is sent from sender B to the multiple receivers A, C, and D. Similarly, there are three parts in this scenario. The first one is Pre-computation that consists steps, M1 and M2. Another one is Sending phase that describes the steps from M3 to M13. The other one is Receiving phase that illustrates the steps from M14 to M23.

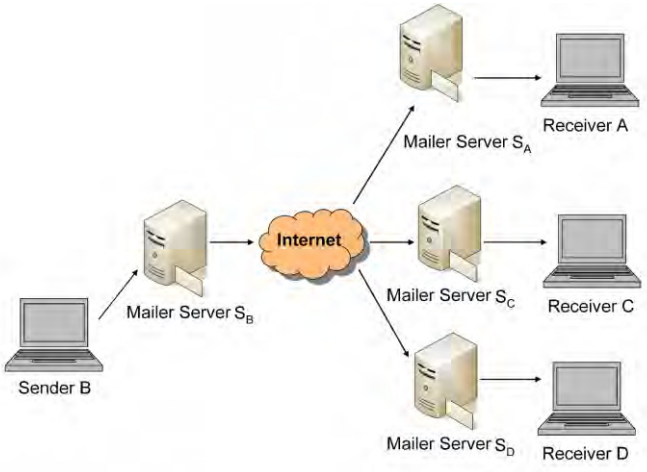



**Fig. 2.** The scenario of an e-mail sent from sender B to the multiple receivers A, C, and D [15, 16].

1) Pre-computation

Step M1: $TC \rightarrow U_i$ : $g$ , $n$ . Note that this step is similar to S1.

Step M2: $U_i \rightarrow S$ : $e_{k_i}$ , $e_{k_j}$ , $Sig_{SK_i}(e_{k_i})$ , $Sig_{SK_i}(e_{k_j})$ , $ID_j$ . Note that this step is similar to S2.

2) Sending Phase

Step M3: $S \rightarrow U_1$ : $e_{k_1}, e_{k_2}, \cdots, e_{k_n}$ , $Sig_{SK_1}(e_{k_1}), Sig_{SK_2}(e_{k_2}), \cdots, Sig_{SK_n}(e_{k_n})$ , $ID_1, ID_2, \cdots, ID_n$ .

Step M4: $U_1$ randomly generates the communication key $K_c$ .

Step M5: The encrypted e-mail message $C = E_{K_c}(M)$ is encryption under the chosen key $K_c$ , where $M$ is the content message of the e-mail.

Step M6: The $p^*_{i,j}$ for each receivers is computed by applying $p^*_{i,j} = E_{k_{i,j}}(K_c)$ for all $j = 1, 2, \cdots, n$ . The generation of $k_{i,j}$ will follow the rule of $k_{i,j} = (e_{k_i})^{k'_j} \bmod n = (g^{k'_i})^{k'_j} \bmod n = g^{k'_i k'_j} \bmod n$ .

Step M7: Each $p_{i,j}$ for all $j = 1, 2, \cdots, n$ is computing using the following equations, $p_{i,1} = \sum_{j=1}^{n} p^*_{i,j}$ , $p_{i,2} = \sum_{j=2}^{n} p^*_{i,j}$ , $\cdots$, and $p_{i,n} = \sum_{j=n}^{n} p^*_{i,j} = p^*_{i,n}$ , such that the decreasing order relation $P_{i,1} > P_{i,2} > P_{i,3} > \cdots > P_{i,n-1} > P_{i,n}$ is satisfied. Moreover $T = p_1$ is the maximal value in the set of $\{P_{i,1}, P_{i,2}, P_{i,3}, \cdots, P_{i,n-1}, P_{i,n}\}$ . A polynomial $f(x)$ is then constructed by the originator as follows, $f(x) = T + (x - k_{i,1}) \times (x - k_{i,2}) \times \cdots \times (x - k_{i,n})$ .

Step M8: Then set $\{P_{i,1}, P_{i,2}, P_{i,3}, \cdots, P_{i,n-1}, P_{i,n}\}$ is encrypted to be a sub-packet $\Gamma$ by the way of $\Gamma = E_T(P_{i,1} \,||\, P_{i,2} \,||\, P_{i,3} \,||\, \cdots \,||\, P_{i,n-1} \,||\, P_{i,n})$ .

Step M9: Sum up $x_{i,j} = \sum_{j}^{n} ID_j$ . Define the initial value $x_{-1} = -1$ . Compute $x_{i,1} = \sum_{j=1}^{n} ID_j$ , $x_{i,2} = \sum_{j=2}^{n} ID_j$ , $\cdots$, and $x_{i,n} = \sum_{j=n}^{n} ID_j = ID_j$ .

Step M10: Compute $\alpha_{i,j}$ 's for $j = 1$ to $j = n$ using the following equations, $\alpha_{i,1} = \prod_{j=n-1}^{1} (T - x_{i,j})$ , $\alpha_{i,2} = \prod_{j=n-1}^{2} (T - x_{i,j})$ , $\cdots$ , $\alpha_{i,n-1} = \prod_{j=n-1}^{n-1} (T - x_{i,j})$ , and $\alpha_{i,n} = 1$ .

Step M11: Construct a basic e-mail packet lock $Z$ as the format of $Z = \sum_{j=1}^{n} \alpha_{i,j} p_{i,j}$ .

Step M12: Compute a varied e-mail packet lock in bit-wise exclusive-or operation of follow: $L = Z \oplus E_T(T)$ .

Step M13: $U_1 \rightarrow S$ : $C$ , $L$ , $f(x)$ , $\Gamma$ , $Y$ , $t$ , where $Y = Sig_{PK_1}(h(ID_1 \| ID_2 \| \ldots \| ID_n \| M \| t))$ . The parameter $t$ is a timestamp at the time which the e-mail is sent form sender $U_1$ to his e-mail server $S$ .

3) Receiving Phase

Step M14: $S \rightarrow U_i$: $C$ , $L$ , $f(x)$ , $\Gamma$ , $Y$ , $t$ , $ID_1, ID_2, \cdots, ID_n$ .

Step M15: Find the maximal sub-packet by computing $f(k_{i,j}) = T$ , and let $\zeta = E_T(T)$ .

Step M16: Decrypt the set $\{P_{i,1}, P_{i,2}, P_{i,3}, \cdots, P_{i,n-1}, P_{i,n}\}$ by using the following equation, $D_T(\Gamma) = (P_{i,1} \| P_{i,2} \| P_{i,3} \| \cdots \| P_{i,n-1} \| P_{i,n})$ .

Step M17: Find the e-mail packet lock $Z$ by computing $L \oplus \zeta = (Z \oplus E_T(T)) \oplus \zeta = Z$ .

Step M18: Sum up $x_{i,j} = \sum_{j}^{n} ID_j$ . Let the initial value be $x_{-1} = -1$ . Compute

$$x_{i,1} = \sum_{j=1}^{n} ID_j \text{ , } x_{i,2} = \sum_{j=2}^{n} ID_j \text{ , } \cdots \text{, and } x_{i,n} = \sum_{j=n}^{n} ID_j = ID_j \text{ .}$$

Step M19: Compute the $\alpha_{i,j}$ 's for $j = 1$ to $j = n$ using the following equations,

$$\alpha_{i,1} = \prod_{j=n-1}^{1}(T - x_{i,j}) \text{ , } \alpha_{i,2} = \prod_{j=n-1}^{2}(T - x_{i,j}) \text{ , } \cdots \text{ , } \alpha_{i,n-1} = \prod_{j=n-1}^{n-1}(T - x_{i,j}) \text{ , and}$$

$\alpha_{i,n} = 1$ .

Step M20: Compute the sub-packet $p_{i,j}^*$ as per following formula,

$$\begin{aligned} p_{i,j}^* &= p_{i,j} - p_{i,j+1} \\ &= \left\{ \left[ \frac{Z}{\alpha_{i,j}} \right] mod(T - x_{i,j-1}) \right\} \\ &\quad - \left\{ \left[ \frac{Z}{\alpha_{i,j+1}} \right] mod(T - x_{i,j}) \right\}, \end{aligned}$$

where $j = 1, 2, \cdots, n$ .

$$p_{i,1}^* = p_{i,1} - p_{i,2} = \sum_{j=1}^{n} p_{i,j}^* - \sum_{j=2}^{n} p_{i,j}^* \text{ ,}$$

$$p_{i,2}^* = p_{i,2} - p_{i,3} = \sum_{j=2}^{n} p_{i,j}^* - \sum_{j=3}^{n} p_{i,j}^* \text{ ,}$$

$$\vdots$$

$$p^*_{i,n-1} = p_{i,n-1} - p_{i,n} = \sum_{j=n-1}^{n} p^*_{i,j} - \sum_{j=n}^{n} p^*_{i,j} ,$$

$$p^*_{i,n} = p_{i,n} .$$

Step M21: Decrypt communication key $K_c = D_{k_{i,j}}(p^*_{i,j})$.

Step M22: Recover the original content message $M = D_{K_c}(C)$.

Step M23: The $U_i$ computes the value $Y' = Sig_{PK_I}(h(ID_1 || ID_2 || \ldots || ID_n || M || t))$ and checks if $Y'$ equals to the value in the signature $Y$.

## 4. Security and Complexity Analysis

The security of the proposed scheme is analyzed, including replay attack, sender impersonation attack, unknown key-share attack, forgery attack and insider attack. Then, the computation complexity of the proposed scheme is discussed.

### 4.1. Security Analysis

**Replay Attack.** The replay attack on e-mail systems means that a certain user who previously established a common key with the sender exploits the preceding key materials to evade victim users' verification procedures. Then the victim users will receive the bogus information from this malicious user without discovering the misbehavior. In the proposed scheme, the messages in Step S14 and M13 contain the time stamp $t$. The sender and receivers can find out this time stamps in their memory or storage device. When a repeated time stamp is found on the received message, receiver can find out this misbehavior and discard the received messages.

**Sender Impersonation Attack.** The sender impersonation attack means that an adversary impersonates a legitimate sender to send a forged message to a receiver. In the proposed scheme, the receiver checks the signature $Y$ signed on by the sender in Step S13 and M13. Due to the properties of cryptographically secure one-way hash function, it is hard to find a collision corresponding to the forged content. In addition, an adversary who does not learn the sender's secret key cannot produce a correct signature for the forged message. Therefore, the sender impersonation attack cannot be engaged successfully.

**Unknown Key-Share Attack.** This attack can be considered as a special case of impersonation attacks. An adversary makes duplicates of the preceding authentication message transmitted between the sender and receiver to cheat a victim user to construct a short-term key. Then, the victim user considers the adversary as an authorized user and

sends him messages, confined to specific authorized users. In the proposed scheme, the sender signs on a digest related to the e-mail in Step S13 and M13. The input value of the signature $Y$ includes the sender's and receiver's identifications, the content in the e-mail, and the timestamp $t$. According to the properties of a cryptographically secure one-way hash function, it is hard to reversely derive the input and find a collision. Moreover, the short-term session key is encrypted by the receiver's public key. If an adversary tries to impersonate the sender with the preceding authentication message, users can check the signature $Y$ to discover the adversary.

**Forgery Attack.** The forgery attack on e-mail systems means that an adversary sends bogus message for authentication. In the proposed scheme, the sender sends the message in Step S13 and M13, which are signed on by the sender's secret key. The receiver can check the validity of the message through the sender's public key. Hence, any adversary cannot successfully engage a forgery attack in the proposed scheme.

**Insider Attack.** The insider attack means that malicious operators of e-mail servers can learn the short-term session key shared between the sender and the receiver. The malicious node can use the short-term session key to eavesdrop the e-mail content or send the bogus message. In the proposed scheme, the short-term session key is only known to the sender and the receiver. Even if a malicious operator of the e-mail server collects the messages transmitted between the sender and the receiver, he cannot derive the short-term session key.

## 5. Computation Complexity

The ratio of average time consumption $\alpha$ is defined in the known cryptographic algorithms [2], [25], i.e. the secret-key systems, DES, Triple DES, and AES, and public-key system, RSA. Suppose that $m$ is the number of e-mail receivers. According to the results in [14], the time consumption results in RSA are around 80 times of that in Triple-DES. Also, it takes the time even around 258 times slower than that in AES. Therefore, the ratio of average time consumption $\alpha_1$ for RSA and Triple DES equals 80,

$$\alpha_1 = \frac{\text{Average time consumption of RSA}}{\text{Average time consumption of Triple DES}} = 80,$$

and the ratio of average time consumption $\alpha_2$ for RSA and AES equals 258,

$$\alpha_2 = \frac{\text{Average time consumption of RSA}}{\text{Average time consumption of AES}} = 258.$$

The results with the scheme in [15] is compared to the results with the proposed scheme in this paper, the comparison for the average time consumption, the number of rounds for modular operation, one-way hash function operation, XOR operation are given in Table 2. For example, m is the number of receivers equals to 50. The result of the average time consumption in Sending Phase by using the method from the CRT-based scheme [15] equals to 4160, the other comparisons are given in Table 3 and Figure 3. The result of the average time consumption in Receiving Phase by using the method

from the CRT-based scheme equals to 240. On the contrary, the result of the average time consumption in both Sending Phase and Receiving Phase by using the proposed scheme also equals to 53, the other comparisons are given in Table 3 and Figure 3. The other comparisons are given in Table 4, Table 5, Table 6, Figure 4, Figure 5 and Figure 6. Therefore, the proposed scheme outperforms the CRT-based secure e-mail scheme [15].

**Table 2.** Computation comparison of the e-mail security protocols

| Compared Items \ Protocols | | The scheme of previous works in [15] | | The proposed scheme | |
|---|---|---|---|---|---|
| Communication Phases | | Sending Phase | Receiving Phase | Sending Phase | Receiving Phase |
| The average time consumption | $\alpha_1 = 80$ | $80 \times (m+2)$ | 240 | $m+3$ | $m+3$ |
| | $\alpha_2 = 258$ | $258 \times (m+2)$ | 774 | m+3 | m+3 |
| The number of rounds for modular operation | | $m+4$ | 4 | $m+1$ | $2m+1$ |
| The number of rounds for one-way hash function operation | | 1 | 1 | 1 | 1 |
| The number of rounds for *XOR* operation | | 0 | 0 | 1 | 1 |

Note that α is the ratio of average time consumption of the known cryptographic algorithms [14], and m is the number of receivers, and m is the number of receivers.

**Table 3.** The ratios of average time consumptions using $\alpha_1 = 80$ are compared in Sending Phase

| Compared Items \ Schemes | The scheme of previous works in [15] | The proposed scheme |
|---|---|---|
| $m$=1 | 240 | 4 |
| $m$=10 | 960 | 13 |
| $m$=15 | 1360 | 18 |
| $m$=20 | 1760 | 23 |
| m=25 | 2160 | 28 |
| m=30 | 2560 | 33 |
| m=35 | 2960 | 38 |
| m=40 | 3360 | 43 |
| m=45 | 3760 | 48 |
| m=50 | 4160 | 53 |

Note: $m$ is the number of receivers.

In Sending Phase, the ratios of average time consumptions using $\alpha_1 = 80$ are compared in Table 3, and the comparison results are depicted in Fig. 3.

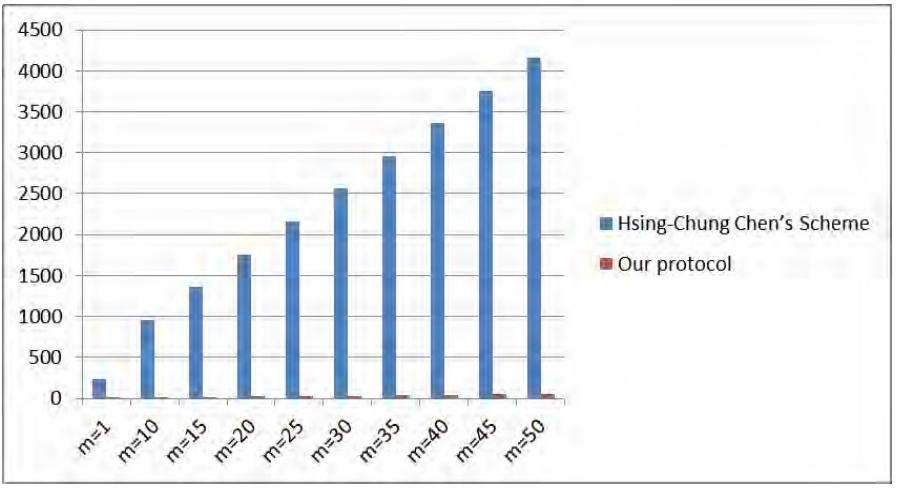



**Fig. 3.** The ratios of average time consumptions using $\alpha_1 = 80$ are compared in Sending Phase

In Receiving Phase, the ratios of average time consumptions using $\alpha_1 = 80$ are compared in Table 4, and the comparison results are depicted in Fig. 4.

**Table 4.** The ratios of average time consumptions using $\alpha_1 = 80$ are compared in Receiving Phase

| Protocols / Compared Items | The scheme of previous works in [15] | The proposed scheme |
|---|---|---|
| m=1 | 240 | 4 |
| m=10 | 240 | 13 |
| m=15 | 240 | 18 |
| m=20 | 240 | 23 |
| m=25 | 240 | 28 |
| m=30 | 240 | 33 |
| m=35 | 240 | 38 |
| m=40 | 240 | 43 |
| m=45 | 240 | 48 |
| m=50 | 240 | 53 |

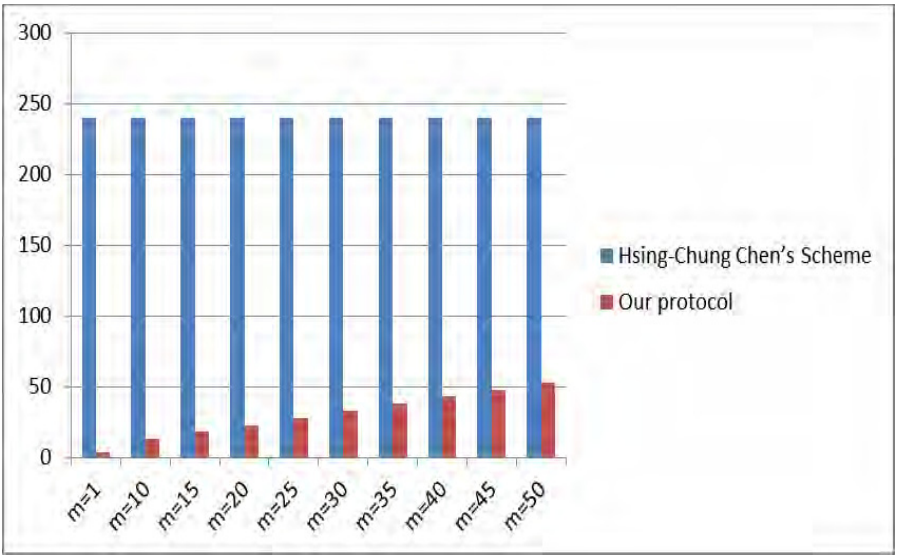



**Fig. 4.** The ratios of average time consumptions using $\alpha_1 = 80$ are compared in Receiving Phase

In Sending Phase, the numbers of modular operation round are compared in Table 5, and the comparison results are depicted in Fig. 5.

**Table 5.** The comparison results of number of modular operation in Sending Phase

| Schemes<br>Compared Items | The scheme of previous works in [15] | The proposed scheme |
|---|---|---|
| m=1 | 5 | 2 |
| m=10 | 14 | 11 |
| m=15 | 19 | 16 |
| m=20 | 24 | 21 |
| m=25 | 29 | 26 |
| m=30 | 34 | 31 |
| m=35 | 39 | 36 |
| m=40 | 44 | 41 |
| m=45 | 49 | 46 |
| m=50 | 54 | 51 |

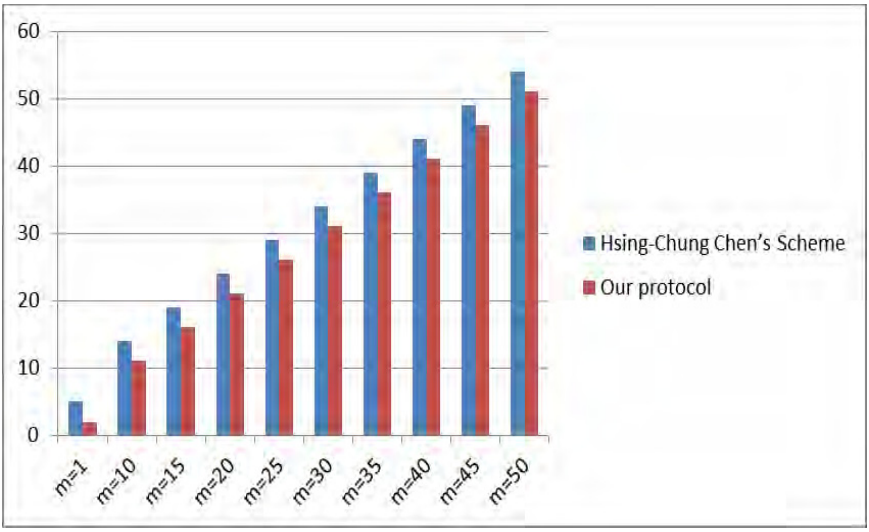



**Fig. 5.** The comparison results of number of mod operation in Sending Phase

In Receiving Phase, the numbers of modular operation round are compared in Table 6, and the comparison results are depicted in Fig. 6.

**Table 6.** The comparison results of number of modular operation in Receiving Phase

| Schemes / Compared Items | The scheme of previous works in [15] | The proposed scheme |
|---|---|---|
| m=1 | 4 | 3 |
| m=10 | 4 | 21 |
| m=15 | 4 | 31 |
| m=20 | 4 | 41 |
| m=25 | 4 | 51 |
| m=30 | 4 | 61 |
| m=35 | 4 | 71 |
| m=40 | 4 | 81 |
| m=45 | 4 | 91 |
| m=50 | 4 | 101 |

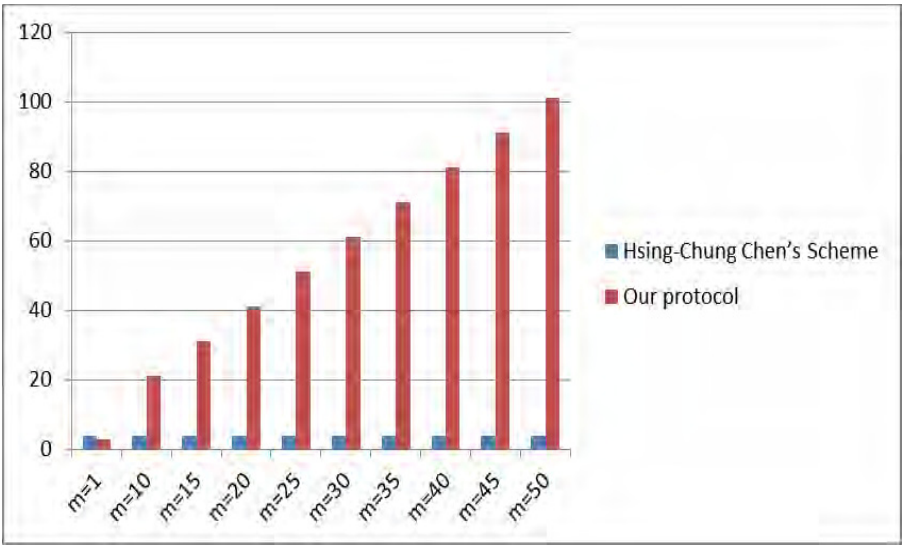



**Fig. 6.** The comparison results of number of modular operation in Receiving Phase

Finally, In Table 2, the number of rounds for one-way hash function operation between both comparison schemes is same. The number of rounds for XOR operation in the proposed scheme equals one, and the number of rounds for XOR operation in the scheme of previous works in [15] equals zero. The values of the last two items in the comparison, Table 2, are too small to be ignored.

## 6. Conclusions

In this paper, a novel secure e-mail system is proposed. The protocol is constructed by ID-based FNS Multicast mechanism with the hybrid cryptographic algorithms of public-key and secret-key system. A secure multicast key protocol is proposed a solution to the e-mail systems for distributing a session key accompanied the sent e-mail to the specific group. Due to the concerning for the securely transferring the session key, the proposed protocol adopts ID-based FNS to replace CRT is proposed. In the manner, not only the e-mail construction in multicast system can be efficiently retained, but also the easy key management and fast computation to process a multiple secure e-mail delivery can be proficiently achieved. The results with the CRT-based secure e-mail scheme is compared to the results with the proposed scheme in this paper, the comparison for the average time consumption, the number of rounds for modular operation, one-way hash function operation, XOR operation are given. According to the results of comparisons in Table 3, Table 4, Table 5, Table 6, Figure 3, Figure 4, Figure 5 and Figure 6, the proposed scheme outperforms the CRT-based secure e-mail scheme.

**Acknowledgments.** This work was supported in part by Asia University, Taiwan, under Grant 101-asia-28, also by the National Science Council, Taiwan, Republic of China, under Grant NSC 102-2221-E-468-007.

## References


1. Basagiannis S., Petridou S., Alexiou N., Papadimitriou G., Katsaros P.: Quantitative analysis of a certified e-mail protocol in mobile environments: A probabilistic model checking approach. Computers & Security, Vol. 30, 257-272. (2011)
2. Chen H.C., Marsha A.V., Weng C.E. Kung T.L.: Cognitive RBAC in Mobile Heterogeneous Networks. Computer Science and Information Systems, Vol. 10, No. 2, 779-806. (2013)
3. Fujisaki E., Okamoto T.: Secure Integration of Asymmetric and Symmetric Metric Encryption Schemes. Advances in Cryptology –CRYPTO'99, LNCS, Vol. 1666, 537-554. (1999)
4. Atkins D., Stallings W., Zimmermann P.: PGP Message Exchange Formats. Internet Draft. (1995)
5. Balenson D.: Privacy Enhancement for Internet Electronic Mail: Part III: Algorithms, Modes, and Identifiers. RFC 1423. (1993)
6. Galvin J., Murphy G., Crocker S., Freed N.: MIME Object Security Services. RFC 1848. (1995)
7. Elkins M.: MIME Security with Pretty Good Privacy (PGP). Internet Draft, 1995.
8. Schneier B.: E-mail Security with PGP and PEM: How to Keep Your Electronic Mail Private. (1995)
9. Sun H.M., Hsieh B.T., Hwang H.J.: Secure E-mail protocols providing perfect forward secrecy. IEEE Communications Letters, Vol. 9 , No. 1, 58 – 60. (2005)
10. Joyia, A., Ghafoor, A., Sajjad, M., Choudhary, M.Q.: Secure and privacy enhanced email system as a cloud service. In Proceedings of 2013 Eighth International Conference on Digital Information Management (ICDIM), 73–78. (2013)
11. Shamir A.: Identity-Based Cryptosystems and Signature Schemes. Lecture Notes in Computer Science Volume 196, 47-53. (1985)
12. McCullagh, N.: Securing e-mail with identity-based encryption. IT Professional, Vol. 7, No. 3, 64, 61 – 63. (2005)
13. Kihidis, A., Chalkias, K., Stephanides, G.: Practical Implementation of Identity Based Encryption for Secure E-mail Communication, In Proceedings of 2010 14th Panhellenic Conference on Informatics (PCI 2010), 101 – 106. (2010)
14. Chen H.C., Wang, S.J., Wen J.H.: Packet Construction for Secure Conference Call Request in Ad Hoc Network Systems. Information Sciences, Vol. 177, Issue 24, 5598–5610. (2007)
15. Chen H. C.: Secure multicast key protocol for electronic mail systems with providing perfect forward secrecy. Security and Communication Networks, Vol. 6, No. 1, 100–107. (2013)
16. Chen H. C. et. al.: Secure Multicast Key Protocol for E-Mail System Using Factorial Number Structure. In Proceedings of 2013 Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS 2013), 611-616. (2013)
17. Zhang M.Q., Xiao H.Y., Yang X.Y.: ID-Based Fair Multi-party Exchange Protocol. 2010 International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Volume 2, 402 – 405. (2010)
18. Zhang M., Takagi, T.: Efficient Constructions of Anonymous Multireceiver Encryption Protocol and Their Deployment in Group E-mail Systems with Privacy Preservation. IEEE Systems Journal, Vol. 7, No. 3, 410 – 419. (2013)
19. Denning D.E.: Cryptography and Data Security. Addison-Wesley, Reading, Mass. (1982)
20. Kent S. T.: Security Requirements and Protocols for a Broadcast Scenario. IEEE Trans. Communications, Vol. 29, No. 6, 778-786. (1981)
21. Phan R.C.W.: Cryptanalysis of E-mail Protocols Providing Perfect Forward Secrecy. Computer Standards & Interfaces, Vol. 30, No. 3, 101-105. (2008)

22. Yoon E.J., Yoo K.Y.: Cryptanalysis of Robust E-mail Protocols with Perfect Forward Secrecy. IEEE Communication Letter, Vol. 11, No. 5. (2007)
23. Menezes A.J., Van Oorschot P.C., Vanstone S.A.: Handbook of Applied Cryptography, CRC Press. (1997)
24. Chang C.C., Wu Y.C., Chang S.C.: A Novel E-mail Protocol Using Three-party Password-authenticated Key Exchange, In Proceedings of International Conference on Security Technology (SECTECH'08), 150-154. (2008)
25. Wang Y., Hu M.: Timing evaluation of the known cryptographic algorithms," In Proceedings of International Conference on Computational Intelligence and Security. (2009)

**Hsing-Chung Chen (Jack Chen)** received the Ph.D. degree in Electronic Engineering from National Chung Cheng University, Taiwan, in 2007. During the years 1991-2007, he had served as a Mobile Communication System Engineer at the Department of Mobile Business Group, Chunghwa Telecom Co., Ltd. From Feb. 2008 to Feb. 2013, he was the Assistant Professor of the Department of Computer Science and Information Engineering at Asia University, Taiwan. Since February 2013–present, he is the Associate Professor at the same University. He is also the Research Consultant of Department of Medical Research at China Medical University Hospital, China Medical University Taichung, Taiwan. Currently, he is interested in Information Security, Cryptography, Role-based Access Control, Computer Networks and Wireless Communications. He was Program Co-Chair of numerous conferences. Dr. Chen was the Editor-in-Chief of Newsletter of TWCERT/CC from July 2012 to June 2013.

**Cheng-Ying Yang** received the M.S. degree in electronic engineering from Monmouth University, New Jersey in 1991, and Ph.D. degree from the University of Toledo, Ohio in 1999. He is a member of IEEE Satellite & Space Communication Society. Currently, he is employed as an associate professor with the Department of Computer Science, University of Taipei, Taiwan. His research interests are performance analysis of digital communication systems, error control coding, signal processing, and computer security.

**Hui-Kai Su** received the B.S degree from I-Shou University, Taiwan, in 1999. He received the M.S. degree and the Ph.D. degree from National Chung-Cheng University, in 2001 and 2006 respectively. He was an Assistant Professor at the department of computer science and information engineering, Nanhua University, Taiwan, during 2006 and 2009. He joined the department of electrical engineering, Formosa University, in the spring of 2009. Currently he is an Associate Professor in the department. His research interests include multimedia network applications, P2P network applications, IP/MPLS network survivability, network QoS control and management, embedded systems, etc.

**Ching-Chuan Wei** was born in Taiwan in 1966. He received his B.S., M.S. and Ph.D. degrees from the Department of Communication Engineering, National Chiao Tung University, Taiwan. Currently, he is a Professor serving for the Department of Information and Communication Engineering, Chaoyang University of Technology. His research interests focus on the biomedical signal analysis and processing.

**Chao-Ching Lee** received the BS degree in Information System from Asia University, Taiwan, in 2011, and the MS degree in Computer Science and Information Engineering from Asia University, Taiwan, in 2013. His research interests include Information Security, Computer Network and Mobile Computing.

*Received: September 24, 2013; Accepted: January 28, 2014.*

Computer Science and Information Systems 11(3):1113–1126 DOI: 10.2298/CSIS130916067P

# Study on Network Architecture of Big Data Center for the Efficient Control of Huge Data Traffic

Hyoung Woo Park[1,2], Il Yeon Yeo[1], Jongsuk Ruth Lee[1,2] and Haengjin Jang[1]

[1] KISTI Supercomputing Center,245 Daehak-ro, Yuseong-gu, Daejeon, South Korea
{hwpark, ilyeon9, jsruthlee, hjjang}@kisti.re.kr
[2] Faculty of Korea University of Science and Technology (UST), 217 Gajong-ro, Yuseong-gu, Daejeon, South Korea

**Abstract.**The network architecture of typical data centersis characterized bytiered networks andaggregation-based traffic controls.The emergence of big data makes it difficult for these data centers to incorporate big data service. The tier and aggregation based traffic management systems can magnify the seriousness of the traffic congestion and extend the congested region when big data moves around in the data center. As a consequence, big data has been forcing data centers to change their architecture dramatically. In this paper, we first address the important paradigm shifts of network architecture caused by big data traffic. We then show the new network architecture which resulted from our experience of the CERN LHC data service. Finally, we illustrate the effect of the throughput improvements of the proposed network architecture using a NS2simulation.

**Keywords:** big data traffic QoS, big data network architecture, big data-front networking, edge traffic separation, big data paradigm shifts.

## 1. Introduction

If we look into the network architecture of data centers with respect to traffic control, the network architecture features tiered networks and aggregation-based traffic control. Tiered networks mean that networks of data centers consist of backbone networks, sub-networks, sub of sub-networks, and so on. Fig.1. shows the typical tiered network and illustrates that the traffic of the lower tier network is to be automatically aggregated at the upper tier network. Therefore, the tiered network traffic [2][3][4] has a tendency to rapidly flood over all of the networks of the data center. Some studies [5][6] were conducted in order to avoid such situations by making tools to provide multi paths under the tiered network. Tiered networks and traffic aggregation have been useful for data centers to economically construct the network and to efficiently control traffic until now. As the era of big data has arrived, the tiers and the aggregation systems are not functioning well any more. In the big data environment, the tier and aggregation based network architecture magnifies the seriousness of the traffic congestion and may extend the congested region because of the way big data moves around in the data center.

Especially, in case of science big data, scientists tend to move big data from the origin site of the big data to the nearest data center because it is hard for scientists to

analyze big data remotely due to the long delay time during the read/write process of big data. After moving big data to the nearest data center, scientists analyze big data with thousands of CPUs that are connected by a very high speed local network in the data center. One well-known big science data is the data from the Large Hadron Collider such as LHC [7] in Swiss CERN. CERN LHC generates multi-peta($10^{15}$) bytes of data per year. It is said that a peta-byte of LHC data analysis needs approximately 3000 CPUs. Therefore, science big data centers utilize Grid computing technology to collect thousands of CPUs scattered in the data center and to orchestrate all the gathered-CPUs working together as if they are single supercomputer. As a result, Grid computing can increase traffic in many parts of the data center network. Big data traffic that is caused by big data transmission and big data processing disturbs data center networks more frequently than we imagine. The impact of both traffics is so strong that it can suffocate other services of the data center for quite a long time.

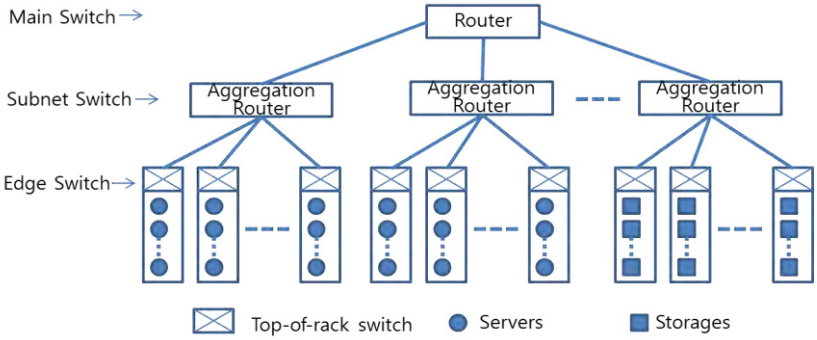



**Fig. 1.** Typical network architecture of the legacy data center

This paper is the extended version of a previous paper [1]. That paper mainly illustrated the phenomena and the impact of big data traffic from our experience only and coarsely proposed the necessity of a new network architecture for big data. But this paper focuses on showing the details of paradigm shifts due to big data traffic and an analysis of the impact in detail by simulation. For example, this paper shows which part of data centers is destined to change and shows what the architecture of big data centers is like after the changes. In this paper, we try to show the reason why those paradigm shifts happen through the simulations. We will continue this study further for the enhancement of R&E infrastructure [13].

Finally, this paper is consisted of 4 parts. We first described problems caused by big data traffic in local network of big data center. Second, we showed some research activities related with the problem. Third, we suggested some paradigm shifts as a new approach to solve the problem. Finally, we showed the result of simulations for proves of our approach.

## 2. Problems Caused by Big Data Traffic

First, we introduce some problems that we experienced during big data service. It took 6 months for us to get just 200 TByte data from KEK institute in Japan via Internet even

though we have a 10 Gbps international link. During the transmission, the LANs of our data center as well as WAN suffered from the traffic. The Grid computing for data analysis also caused local traffic bursts for long periods. The load for QoS processing of network devices and IP packet filtering of the firewall became gigantic because of the huge volume of big data. Therefore, we had to quickly buy more expensive network devices such as high performance routers as soon as we launched big data service.

The typical features from the perspective of the network administrator for big data traffic are long burst traffic, jumbo IP packet frame and low priority. The LHC data that we serviced as a big data is one of the most well-known big data. The size of the data is almost peta($10^{15}$)-byte scale data. Therefore, it always took a long time to move it. So, long burst traffic on the network of big data centers is the most typical feature of big data traffic. The second feature, Jumbo frame means a 9K byte packet. It is recommended for the high performance transmission of big data. The size of packets in ordinary Internet usage is generally less than 1.5K byte. Therefore, the effect of packet loss with big data is more serious than that of ordinary Internet packet loss. The third feature, low priority in QoS control of big data traffic, means that the big data traffic should be dropped first, not ordinary traffic, when a congestion of public networks happens. That is because Internet Service providers don't want big data to disturb ordinary Internet traffic.

The scope of the problem mentioned in this paper is limited to the local area traffic of the legacy data center. Generally, storages and file servers are located at the lowest subnet in the tiered network architecture of the legacy data center. Therefore, most parts of the local network of the legacy data center are suffering whenever big data moved from the local storage to the computing servers for the analysis of the data. The aim of the new network architecture is to reduce both of the congested area and the congestion time caused by big data traffic. Therefore, Problems caused by big data traffic can be enumerated by long time of the network congestion, the vast range of the congested region and the strong aggressiveness of big data for occupying the network bandwidth.

Considering big data traffic in the LAN of the legacy data centers once more, their tiered network architecture and the aggregation based traffic management are not the best strategy for traffic management any longer due to the increase of big data traffic. Aggregation based traffic management demands data centers raise the network bandwidth of the data center or enhance QoS function in networking devices. This expenditure continuously increases according to the increase of the volume of big data as we mentioned above. The summary of the problem is that big data analysis as well as big data transfer drops the quality of the data center service because data centers use grid computing for the collection of thousands of CPUs spread in the data center.

## 3. The Related Researches for the Separation of Big Data Traffic

CERN LHC data is one of big data and CEN LHC produces multi peta($10^{15}$)-byte data per year. Therefore, it is difficult for a single data center to analyze peta-scale of scientific big data such as CERN LHC data [15][16] within a single data center. Therefore, The CERN LHC data should be moved to multiple data centers over the world. There are 10 data centers [17][18] for the analysis of CERN LHC data. They are called CERN LHC Tier1 centers. Our center (GSDC in KISTI) is one of them. 10 Tier1

centers consist of global infrastructure for LHC data share and analysis computing. Therefore, it is essential for 10 Tier1 centers to work together as if they are single system [19][20][21]. This single system is called WLCG (World-wide LHC Computing Grid). For the analysis of peta-scale data, 10 Tier1 centers have been researching on separating LHC data traffic from the legacy Internet data traffic because peta-scale of LHC data that produced annually severely suffers other traffic. Therefore, related researches that described in this paper are focused on the research activities for building additional infrastructure for the separation of LHC data traffic. We are going to insist that the structure for the separation of LHC data traffic should be extended into the local network architecture of the data center if a data center wants to service CERN LHC data.

To survey related research about the problems mentioned above, we first studied research for the construction of the dedicated network for big data transfer. Among well-known dedicated networks for big data, there is LHCOPN [9], LHCONE [10] and ScienceDMZ [8]. LHCOPN is operated by the research community of CERN LHC data. LHCOPN is a kind of the dedicated optical network and it is built globally by ten data centers and services at 10Gbps for peta-byte data transfer. LHCONE is a kind of dedicated Internet for CERN LHC data. It provides services between Tier 1 centers and Tier2/Tier3 centers. ScienceDMZ has been implemented for local networks of data centers, which use vLAN for building virtually dedicated networks. The difference between LHCONE and ScienceDMZ is that the aim of LHCONE is to support the data transfer between LHCOPN and research organizations. But, ScienceDMZ is mainly used for local networks within a data center. We also surveyed Grid computing technology as related research because we found Grid computing made big data move for big data processing. Grid computing builds a dedicated cluster based computing farm for big data analysis. We surveyed Grid computing, virtual computing [14] and cloud computing [11][12].All of them are used for gathering thousands of CPUs. Therefore, it is inevitable for the legacy traffic and big data traffic to brim over in legacy data centers. If other data centers' resources are collected to use, the range of the network congestion is further extended.

## 4. The Paradigm Shifts of Network Architecture for Big Data Center

Thus, this paper suggests the new network architecture of data centers for big data. New requirements for the network architecture of big data centers can summarized as follows. First, it must avoid collision between big data traffic and the other traffic when big data move around. Second, it must minimize the region of the big data traffic residence in the data center for the reduction of the influence of big data traffic. Finally, it must reduce the QoS cost that is exponentially rising in proportion to the increase of the volume of big data. We suggest three paradigm shifts of the network architecture to meet the requirements above. These are a paradigm shift on resource provisioning, a paradigm shift on service provisioning, and a paradigm shift on QoS provisioning.

### 4.1. The 1st Paradigm Shift on Resource Provisioning

The 1st paradigm shift is related to the point of separating big data traffic from the ordinary traffic. The volume of big data is more than a million times of the size of the ordinary data. Therefore, it is impossible to separate big data traffic by simply allocating a virtual circuit because the size of big data traffic is beyond full utilization of the physical network device. Big data traffic always demands full allocation of the capability of the network device for an extended period. For convenience, this paper uses the term, BDC, for Big Data Center. BDC is also used for future data centers. IDC or Internet Data Center is used for legacy data centers.
Fig.2 and Fig.3 show the difference before and after the 1st paradigm shift on resource provisioning. Fig.1 illustrates the shared use of virtual resources by the separation of logical networks in the legacy data center. Fig.2 indicates the proprietary use of physical resources by dynamic allocation.

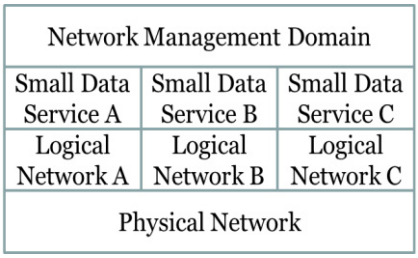



**Fig. 2.** The share use of virtual resource by the separation of logical network in legacy data center

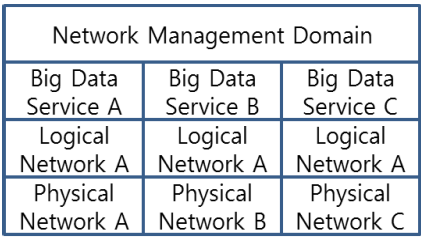



**Fig. 3.** The proprietary use of physical resource by dynamic allocation for big data center

### 4.2. The 2nd Paradigm Shift on Service Provisioning

The 2nd paradigm shift on service provisioning addresses the change of the sequence of the service processing. In the legacy data center, legacy data is usually attached the computing server. It is hidden to users. But, in the era of big data, big data should be first, the server for big data processing will be invisible. Users don't to need to know which computers service their job. The typical differences caused by the 2nd paradigm shift on service provisioning are a change from client/server computing to data-driven computing in computing architecture and a change from menu-driven service to user-

defined service in service architecture. Fig.4 and Fig.5 show the impact of the 2nd paradigm shift on service provisioning.

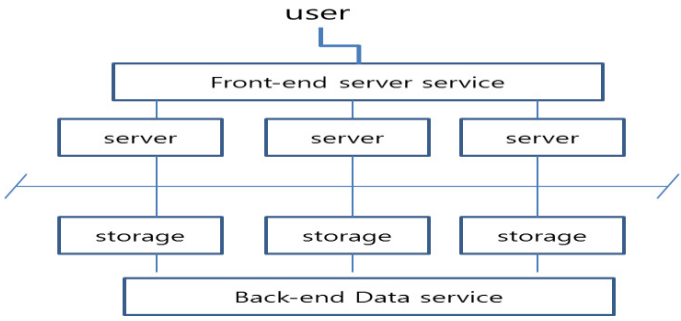



**Fig. 4.** Traditional client-server service architecture for menu-based service in legacy data center

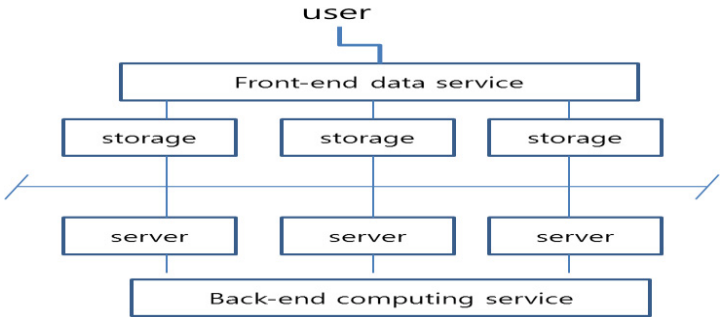



**Fig. 5.** User-defined service by data-driven computing architecture for big data center

### 4.3. The 3rd Paradigm Shift on QoS Provisioning

Finally, the 3rd paradigm shift on QoS provisioning is related to QoS initiative. In a legacy data center, QoS initiative belongs to the Internet service provider or network administrator. These kinds of QoS management costs are high in dealing with big data because they have always tried to solve QoS problems by purchasing more expensive network devices which have more performance than that of the existing devices. This paradigm shift suggests moving QoS initiative from ISP to users or end systems. Fig. 6 shows the overload of QoS at each tier in the tiered network architecture. The increase of load of the lower tier increases the load of top tier dramatically. Fig. 7 illustrates the possibility of reduction of QoS load in each tier if part of the load for QoS control is moved from the network devices to end systems. Fig.8 shows traffic flow when traffic separation occurs at end systems.

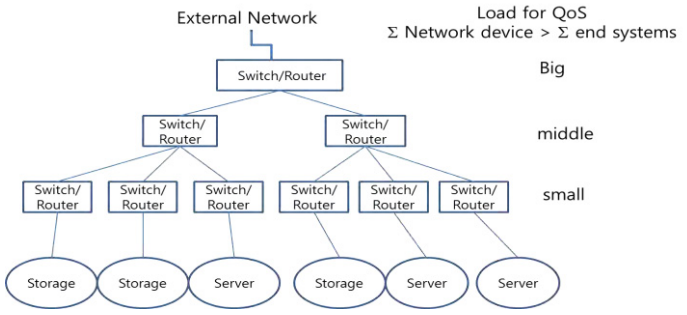



**Fig. 6.** Tree-like centralized QoS provisioning in traditional data center

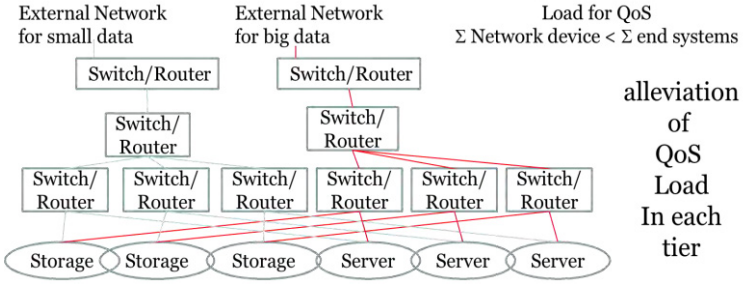



**Fig. 7.** Traffic separation based QoS provisioning by end system

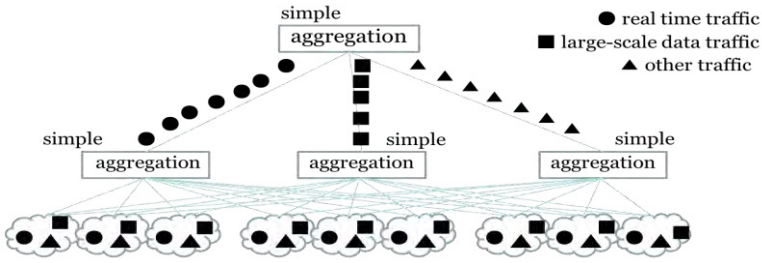



**Fig. 8.** Traffic separation based QoS provisioning by end system

### 4.4. The Candidate Network Architecture for Big Data Center

To accomplish these paradigm shifts, we designed anew network architecture for big data centers. We first divided the data center network into 3 parts for the separation of traffic. Part 1 is for big data transfer and sharing, part 2 is for big data analysis computing, and part 3 is for user access and job control. Part1 and part 3 are configured with a public IP address and a private IP address. Part 2 is configured with only a private IP address for security. We also suggest dual interconnection between each part using front end networking and back end networking. These dual interconnections can eliminate the traffic collision between big data traffic and small data traffic. User access

is allowed only by front networking. Big data is serviced only by back end networking Fig. 9 shows the implementation example of the new approach.

The key benefit of our approach is a dramatic reduction of the cost for the big data network operation and management. Traffic separating at the end systems reduces the requirement of high performance routers and extends the life of the legacy network devices.

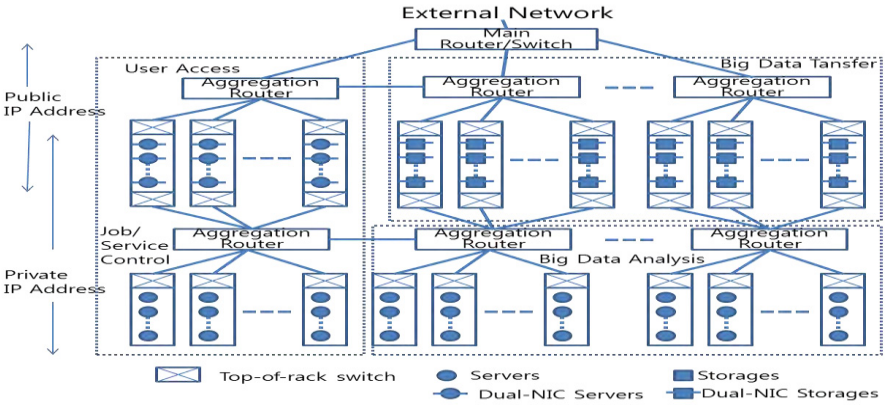



**Fig. 9.** The candidate architecture for big data center, empowered by organic interrelationship among functional areas

## 5. Simulations for the Analysis of the Impact of Big Data Traffic

We did the simulation in order to prove the advantage of the paradigm shifts on network architecture mentioned above. It is a well-known fact that traffic separation improves the throughput of the congested network. Therefore, we simulated to show how big data and small data interact when they are co-existing. For this purpose, we set some conditions for the simulations. We set a 9k-byte packet for big data traffic because the 9k-byte frame is strongly recommended by CERN LHC data center. It is also called jumbo frame. For small data, we use the 1.5 k-byte frame. Most ordinary packets are less than 1.5k byte. Big data transfer usually uses TCP protocol. Fig. 10 showed the configuration for the simulation. N0 is a congested node. N1 is a destination node for all of the source nodes. Each link delay is set at 2ms for the simulation of local data center traffic. Except packet size and link delay, we use the default parameter of NS2. We set the volume of traffic to be the same for a reasonable comparison. Table 1 shows the combination of 9k-byte frame and 1.5k byte frame. Simulation is to run for 90 seconds.

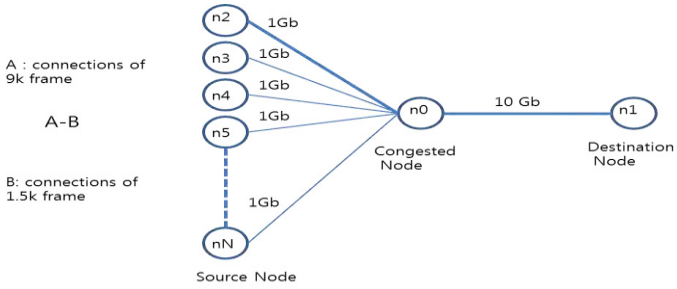



**Fig. 10.** Simulation for the study of the interrelation between big data traffic and small data traffic

**Table 1.** The list of the various combinations between big data traffic and small data traffic

| type of frame | Combinations | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| - no. of sessions with 9k frame | 21 | 18 | 15 | 12 | 9 | 6 | 3 | 0 |
| - no. of sessions with 1.5k frame | 0 | 18 | 36 | 54 | 72 | 90 | 108 | 126 |
| - total No. of sessions | 21 | 36 | 51 | 60 | 81 | 96 | 111 | 126 |

Fig. 11 shows the amount of the data received at node 1 during the simulation. X-axis stands for the combination for 9k-byte frame and 1.5k-byte frame. The former number is for the number of 9k-byte frames, the latter number for the number of 1.5k-byte frames. We find that the amount of transferred data is decreased according to the increase of the number of total sessions. Therefore, total packet loss in Fig. 12 increased according to the increase of the number of total sessions. One of the interesting results of simulation is shown by Fig.13. The number of packets dropped in a single session can be reduced even though the number of sessions increases when sessions have the same kind (size) of packets. In other words, it is proven that it is better traffic management to classify and to group the traffic into the similar traffic packets.

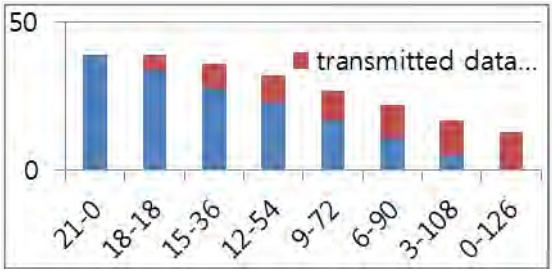



**Fig. 11.** The total amount of the data received at node 1 during the simulation. Y-axis unit is Mega Byte

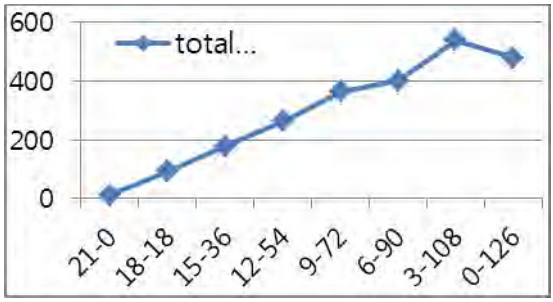



**Fig. 12.** Total packet drops during simulation. Y-axis unit is number of drops

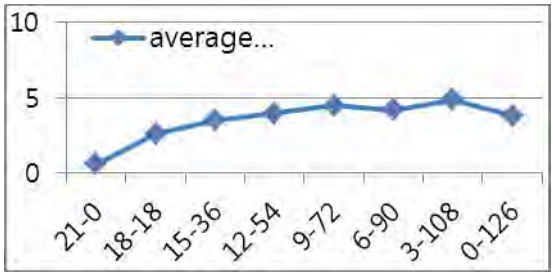



**Fig. 13.** Average packet drops per sessions. Y-axis unit is number of drops

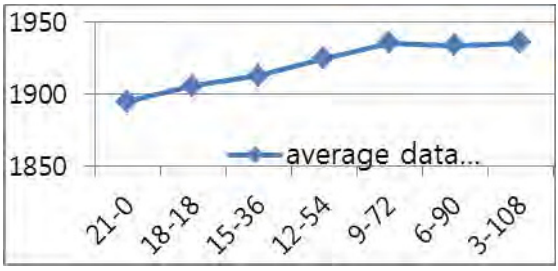



**Fig. 14.** Average data transmitted by sessions with 9k frames. Y-axis unit is Kilo Byte

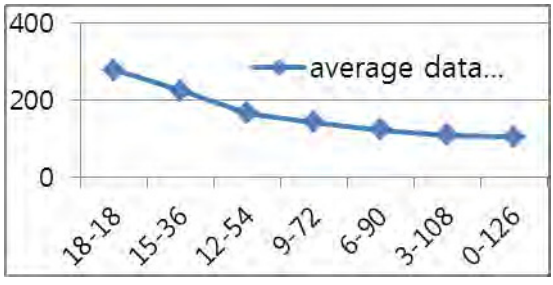



**Fig. 15.** Average data transmitted by sessions with 1.5k frames. Y-axis unit is Kilo Byte

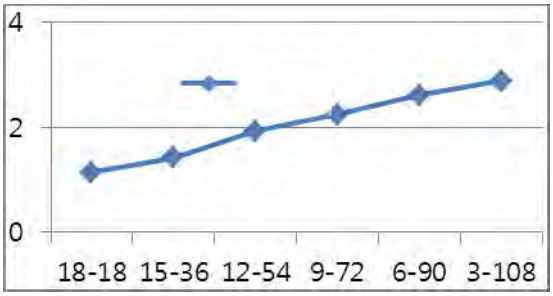



**Fig. 16.** Relative ratio between transmitted data by the unit size of 9k frame and transmitted data by the unit size of 1.5k frame

Fig. 14 also shows an interesting result. It shows that the amount of transmitted data by the 9k-byte frame is continuously increasing compared with the amount of transmitted data by the 1.5k-byte frame which is decreasing. The decrease is shown in Fig. 15. These phenomena illustrate that big data is stronger than small data in comparison of aggressiveness to occupy network bandwidth. This result can be a proof that the separation of big data traffic from ordinary data traffic is necessary. Fig.16 shows the performance ratio between the performance obtained by small frame and performance obtained by big frame on the same simulation. Small frame means 1.5K byte Packet Data Unit and big frame means 9K byte PDU. Therefore, Fig. 16 also indicates that the aggressiveness of big data is higher than small data. The ratio of the aggressiveness varies from 114% to 288%.

For more detail explain of results, we describe meaning of parameters among result Figures showed in this paper. “A-B” marked in the left of Fig. 10, A-B means the number of a pair (or a set of A and B). One(for A) is the number of sessions for 9k-byte frame transmitted by source nodes and the other (for B) is the number of sessions for 1.5K-byte frame transmitted by the rest of source nodes that are not joined 9k-byte frame transmission. In other words, A is denoted for the number of TCP sessions that transmits 9k-byte frame transmission and B is denoted for the number of TCP sessions that sends data with 1.5k-byte frame. This “A-B” form is used for the value of X-axis in the Fig.12, Fig.13, Fig.14, Fig.15 and Fig. 16. For example, “21” stands for A and “0” stands for B at the first value (21-0) of X-axis in Fig. 11. And, “0” means for A and “126” means for B at the last value (0-126) of X-axis in Fig. 11. This denotation style of the value of X-axis is applied to the all of result Figures from Fig.11 to Fig. 16. The meaning for the value of Y-axis is denoted at the bottom of each Figure. In Fig. 12, the value of Y-axis stands for the number of total packet drops whenever we simulated under the condition of each A-B combination that is denoted at the value of X-axis. The meaning of the result showed in Fig. 12 indicated that packet drops is proportional to the number of sessions that were joined in the simulation and at the same time Fig. 12 indicated that it is relatively less related with the amount of the total data that are transmitted during the simulation. Because we configure NS2 simulation program to send same amount of data for all of simulation even though the pair value of A-B changed. This result is very important results for the management of big data traffic. In Fig. 13, we can find more confidence on that reducing the number of sessions is better for the management of big data traffic. In Fig. 13, the value of Y-axis stands for the average value of packet drops at each combination of A (the number of 9k-byte-frame

used sessions) and B (the number of 1.5K-byte-frame used sessions). Therefore, we also reach same result that showed in Fig. 12. The most important result is also showed in Fig. 13. That is, the average packet loss at the both ends of X-axis is lower than average packet loss at the middle of X-axis. It means that average packet loss is decreased abs the ratio of homogeneity of packet size is increased. This result can be also the proof of our proposed architecture. This result is also appeared in Fi. 14. But, the meaning of the value for Y-axis is changed from packet drops to throughput.

## 6. Conclusion

The goal of this study is to develop new network architecture for big data centers. As we mentioned above, big data traffic will have a major influence in creating paradigm shifts of the system and network architecture of big data centers. These paradigm shifts are related to resource provisioning, service provisioning and QoS provisioning. Therefore, big data will change the architecture of data centers fundamentally. Due to these paradigm shifts, the tiered architecture of IDC will be changed into full-matrix architecture for BDC, and we can also expect that dynamic physical resource allocation will be preferred to the allocation of virtual systems, the decision power of the network path will belong to users not network providers, and the demand for expensive backbone routers can be reduced by edge traffic separation. An interesting feature of our new approach for network architecture is a kind of recycling-friendly architecture because the proposed architecture requires a plentiful number of legacy network cables and legacy low-end network devices instead of buying expensive and cutting-edge network devices.

According to our investigation, the future network architecture of the big data center will be a dual matrix architecture in which the big data part will be located at the front and the center of the architecture in order to reduce the number of interactions between the big data traffic and the legacy traffic. Therefore, the thing that we are going to study further is how we can carry out the transforming of the current network architecture of our data center (GSDC: Global Science Data Center) from tier architecture to data-centered architecture. It is difficult for data centers to change their network architecture without interruption of service. We are building the proposed architecture in parallel. We will first construct the big data share part, and then move the big data analysis part. Finally we will upgrade the user service part. Dual interconnection among the 3 parts will be parallel implemented. Therefore, we will spend much time in rearranging the network devices in order to group similar kinds of traffic onto same network. Finally, we hope this architecture and our experience will help legacy data centers introduce big data service.

**Acknowledgements**. This work was supported by the program of the Construction and Operation for Large-scale Science Data Center, 2014funded by the KISTI and by the program of the Global hub for Experiment Data of Basic Science, 2014 funded by the NRF.

## References


1. Park, H, Yeo I, Lee, J, Jang, H: Study on big data center traffic management based on the separation of large-scale data stream. In the proceedings of the seventh International Conference on Innovative Mobile and Internet Service in Ubiquitous Computing. pp. 591-594. Asia University, Taichung, Taiwan (2013)
2. Kandula, S, Sengupta, S, Greenberg, A, Patel, P, Chaiken, R: The nature of Data Center Traffic Measurement & Analysis. In the proceedings of IMC'09. pp. 202-208. Chicago, Illinois, USA (2009)
3. Benson, T, Anand, A, Akella, A, Zhang, M: Understanding Data Center Traffic Characteristics. ACM SIGCOMM Computer Communication Review, volume 40 issue 1, 92-99 (2010)
4. Benson, T, Akella, A, Maltz, A: Network Traffic Characteristics of Data Centers in the Wild. In the proceedings of IMC'10. pp. 267-280. Melbourne, Australia (2010)
5. Al-Fares, M, Radhakrishnan, S, Raghavan, B, Huang, N, Vahdat, A: Hedra: Dynamic Flow Scheduling for Data Center Networks. In the NSDI'10 Proceedings of the 7th USENIX conference on Networked systems design and implementation.(2010)
6. Benson, T, Anand, A, Akella, A, Zhang, M: MicroTE Fine grained Engineering for Data Centers. In the proceedings of ACM CoNEXT 2011 on emerging Networking experiments and Technologies, ACM. Tokyo, Japan, Article No. 8. (2011)
7. CERN LHC: http://public.web.cern.ch/public/en/lhc/lhc-en.html
8. ScienceDMZ: http://fasterdata.es.net/science-dmz/
9. LHCOPN: http://lhcopn.web.cern.ch/lhcopn/
10. LHCONE: http://lhcone.net/
11. Agrawal, D, Das, S, Abbadi, A: Big Data and Cloud Computing: Current State and Future Opportunities. In the proceedings of EDBT 2011. pp. 530-533. Uppsala, Sweden (2011)
12. Cisco Cloud Computing - Data Center Strategy, Architecture, and Solutions, Point of View White Paper for U.S. Public Sector, http://www.cisco.com/web/strategy/docs/gov/CiscoCloudComputing_WP.pdf
13. Park, H, Kim, S, Lee, J, Jang, H, Cho, K: u-TransitNET: Researches on the user-controlled path-shortening for the reduction of the path-delay in detoured R&E networks. INFORMATION, An International Interdisciplinary Journal, volume 4, 933-944 (2009)
14. http://en.wikipedia.org/wiki/Virtuality_(computing)
15. Juve, G, Rynge, M, Deelman, E, Vockler, J, Berriman, G: Comparing FutureGrid, Amazon EC2, and Open Science Grid for Scientific Workflows. Computing in Science & Engineering, Vol. 15, issue 4, 20-29, (2013)
16. Adamova, D, Horkey, J: An optimization of the ALICE Xrootd storage cluster at the Tier-2 site in Czech Republic. In Journal Conference on Computing in High Energy and Nuclear Physics 2012, NewYork, USA, 1-14, (2012)
17. WLCG(Worldwide LHC Computing Grid) Tier 1 sites, http://lcg-archive.web.cern.ch/lcg-archive/public/tiers.htm
18. Aiftimiei, C, Ceccanti, A, Dongiovanni, D, Meglio, A, Giacomini, F: Improving the quality of EMI Releases by leveraging the EMI Testing Infrastructure. Journal of Physics: Conference Series, Vol. 396, No. 5, 1-12, (2012)
19. EMI (European Middleware Initiative): http://www.eu-emi.eu
20. OSG(Open Science Grid): http://www.opensciencegrid.org/
21. Rodriguez, A, Gouveia, V, Meneses, D, Capannini, Aimar, A, Meglio, A: Article 1. Multi-platform Automated Software Building and Packaging. Journal of Physics: Conference Seriese, Vol. 396, No. 5,1-7, (2012)

**Hyoung Woo Park** is a Principal Researcher in KISTI Supercomuting center in South Korea. He obtained his Ph. D. in Computer Networks at SungKyunKwan University in South Korea. He has participated lots of R&D projects including the construction of national R&D network (KREONET), the implementation of National Grid Computing Infrastructure, and so on. Currently, He joins the project for construction of peta-scale science data Grid center for the global collaboration researches on CERN LHC data, KEK Belle data etc.

**Il Yeon Yeo** is a senior researcher at Global Science Data hub Center, KISTI. He obtained his Master degree in Electronic Engineering at Kyungpook National University. He is working for ALICE which is one of the WLCG experiment. He is interested in Grid Computing and Information Retrieval.

**Jongsuk Ruth Lee** is the corresponding author of this paper. She is a principal researcher and a head of the Department of Advanced Application Environment Development, National Institute of Supercomputing and Networking, Korea Institute of Science and Technology Information (KISTI), Daejeon, South Korea. She is also an adjunct faculty member at University of Science & Technology of Korea, South Korea. She received her Ph.D. in Computer Science and Software Engineering from the University of Canterbury, New Zealand. She, as a researcher, worked for University of Canterbury, NZ from 1998 to 2002, and Korea Electronics and Telecommunications Research Institute (ETRI) from 1992 to 1993. She is the author or co-author of more than 160 research publications, including more than 32 domestic and international patents. Dr. Lee is on the editorial board and a reviewer for various domestic and international journals. Her research interests include parallel computing, distributed simulation, simulation based cyber learning, and grid computing. Board Member of KIPS and KSII.

**Haengjin Jang** is a director of WLCG Tier1-Korea which is the peta-scale science data Grid Center for the global collaboration researches on CERN LHC data, KEK Belle data etc. He obtained Ph.D. in Computer Science at CHONBUK national university in South Korea. He joined the construction of national supercomputing center since 1988. He was a member of IT expert committee that is managed by Korea Telecommunication Technology Association and worked for Korea business Grid association as the vice chair.

*Received: September 16, 2013; Accepted: February 25, 2014.*

Computer Science and Information Systems 11(3):1127–1141 DOI: 10.2298/CSIS130921068W

# An Efficient Wormhole Attack Detection Method in Wireless Sensor Networks

Guowei Wu[1], Xiaojie Chen[1], Lin Yao[1], Youngjun Lee[2], and Kangbin Yim[2]

[1] School of Software, Dalian University of Technology,
Dalian, 116620 China
wgwdut@dlut.edu.cn, 747070908@qq.com, yaolin_yl@hotmail.com
[2] Dept. of Information Security Engineering, Soonchunhyang University,
Asan, 336-745 Korea
dog3hk@gmail.com, yim@sch.ac.kr

**Abstract.** Wireless sensor networks are now widely used in many areas, such as military, environmental, health and commercial applications. In these environments, security issues are extremely important since a successful attack can cause great damage, even threatening human life. However, due to the open nature of wireless communication, WSNs are liable to be threatened by various attacks, especially destructive wormhole attack, in which the network topology is completely destroyed. Existing some solutions to detect wormhole attacks require special hardware or strict synchronized clocks or long processing time. Moreover, some solutions cannot even locate the wormhole. In this paper, a wormhole attack detection method is proposed based on the transmission range that exploits the local neighborhood information check without using extra hardware or clock synchronizations. Extensive simulations are conducted under different mobility models. Simulation results indicate that the proposed method can detect wormhole attacks effectively and efficiently in WSNs.

**Keywords:** wormhole attacks, wireless sensor network, local neighborhood, network topology.

## 1. Introduction

Wireless sensor networks (WSNs) consist of a large number of low-cost and resource constraint sensor nodes to perform distributed sensing tasks. Sensor nodes in WSNs collaborate with each other to transmit messages in a multi-hop manner. WSNs are used for various tasks such as surveillance, widespread environmental sampling, security, and health monitoring [23][2]. WSNs are characterized by their infrastructure-less nature, ease of deployment and independence to any pre-existing architecture [24]. Since the open nature of wireless communication, WSNs are prone to be attacked in various ways, such as Denial of Service (DOS) attack, the wormhole attack, the Sybil attack, selective forwarding attack, etc. [22].

In this paper, the wormhole attack [1][5][10] is taken into consideration. The wormhole attack is a kind of tunneling attack, which is very dangerous and damaging to defend against even though the routing information is confidential, authenticated or encrypted [9]. The adversary doesn't need to have knowledge about the routing protocols or compromise the sensor nodes. In wormhole attack, two malicious nodes are connected through

a low-latency link, namely wormhole link. A low latency can be realized through a network cable, other kind of wired link technology or just a long-range out-of-band wireless transmission [20]. Once the wormhole link is established, the adversary eavesdrops on packets at one end of the link, tunnels them through the wormhole link and replays the packets at the other end of the link. This makes the sensor nodes around the two ends of the wormhole link seem like neighbor nodes as though they are multi-hops away from each other actually.

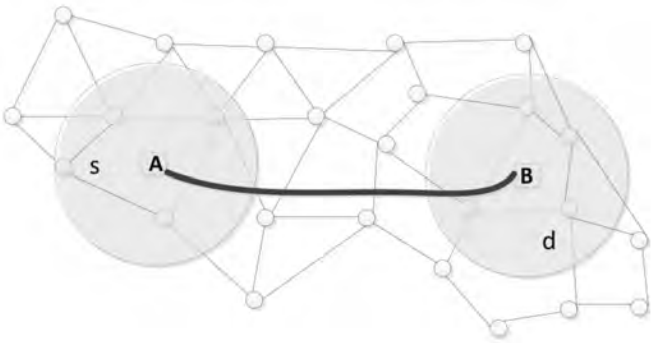



**Fig. 1.** The Minimum Key Set Route

An example of wormhole attack is given in Fig.1. Node A and B are two malicious nodes placed by the adversary connected via a network cable. So node A and B are the two end points of the wormhole link. Node A receives packets, tunnels them through the wormhole link and replays the packets at node B and vice versa. As a result, nodes in the neighborhood of node A will assume that all nodes in the neighborhood of node B are their neighbors and vice versa. For example, source node s can take a one-hop path to send packets to destination node $d$ via the wormhole link instead of a multi-hop path.

A number of protocols have been proposed to defend against wormhole attacks in wireless networks by adopting synchronized clocks, positioning devices, or directional antennas [19]. In this paper, we introduce novel approaches for detecting wormhole attacks and propose an efficient wormhole detection algorithm, which is named Transmission Range based Method (TRM). With the existence of wormhole, the network topology is destructed and normal routes are misled. Unlike many existing techniques, it does not use any specialized hardware, making it extremely useful for real-world scenarios. Most importantly, however, the algorithm can always prevent wormholes, irrespective of the large transmission range, by checking the local neighborhood information to decide whether the network topology is true or faked, while its efficiency is not affected even by the dynamic topology. We also provide an analytical evaluation of the algorithm's correctness through simulation experiments that demonstrates its efficiency in terms of computation complexity and processing delay. The remainder of this paper is organized as follows. In Section 2, related works are discussed. The wormhole attack detection method is presented in Section 3. The performance of our method is evaluated through simulation experiments in Section 4. At last, we conclude our work in Section 5.

## 2. Related Works

Wormhole attack is very destructive since the neighborhood information is confused. Any routing protocol relying on network topology information can't work normally. Periodic protocols like Secure Efficient Ad hoc Distance vector routing protocol (SEAD) [6] will malfunction because the routing table information is different from the real network topology due to the wormhole. On-demand protocols like Dynamic Source Routing protocol (DSR) [11] will have false route establishment because the route quest and route reply message in the route discovery stage will contain the wormhole link. So all the routes established by these network routing protocols are attracted to the wormhole and the adversary can launch further attack like selective forwarding attack, black hole attack and etc. What is worse, the wormhole attack is easily deployed to some extent. The adversary has no need to compromise any node in the network and don't need to deal with the cryptographic keys. The integrity, authenticity and confidentiality are still reserved in the existence of wormhole. All the adversary has to do is to place two malicious nodes in good positions in the network and make them receive and send packets.

Because of the reason, the detection of wormhole attack has become an essential issue and various methods have been proposed to detect the wormhole. In [7], Hu et al. introduce the general mechanism of packet leashes to detect wormhole attacks. Two types of leashes are used: geographic leashes and temporal leashes. A leash is any information that is added to a packet designed to restrict the packet's maximum allowed transmission distance. However, to form a leash, each node must know its own location and have synchronized clocks. In [8], the End-to-end Detection of Wormhole Attack (EDWA) is proposed in wireless ad-hoc networks. The source node estimates the minimum hop count to the destination and compares the hop count value received from the reply packet to detect the wormhole. Obviously, each node should measure its geographical location through a GPS. There are some solutions based on the discovery and maintenance of node neighborhood. For instance, LITEWORP [12] uses secure two-hop neighbor discovery and local monitoring of control traffic to detect nodes involved in the wormhole attack. It provides a countermeasure technique that isolates the malicious nodes from the network thereby removing their ability to cause future damage. MobiWorp [13] is further proposed to complement LITEWORP by introducing some location-aware mobile nodes.

Most existing solutions are based on the network topology. Lazos et al. [14] present a graph-based framework to tackle wormhole attacks. Making use of geometric random graphs induced by the communication range constraint of the nodes, the authors present the necessary and sufficient conditions for detecting and defending against wormholes. In [16], the authors propose a wormhole detection approach with only local connectivity information. The algorithm uses only connectivity information to look for forbidden substructures in the connectivity graph. In [4] a distributed connectivity-based wormhole detection method is proposed. Each node collects its k-hop neighborhood and checks whether the boundary of its k-hop neighborhood sub-graph has one or two circles. Its basic idea is based on the observation that the neighborhood that encloses a wormhole link will have two cycles and single cycle otherwise. In [3], authors develop a simple distributed algorithm for wormhole detection in wireless ad hoc and sensor networks, using only the communication graph, and not making unrealistic assumptions. Their algorithm works well in relatively dense and regular networks but results in many false positives in sparse or random networks. In [15], each node locally collects its neighborhood informa-

tion and reconstructs the neighborhood sub-graph by Multi-Dimensional Scaling (MDS). Potential wormhole nodes are detected by validating the legality of the reconstruction. Then, a refinement process is introduced to filter the suspect nodes and to remove false positives. In the paper [21], wormhole attack detection is proposed based on Round-Trip Time (RTT) between successive nodes and congestion detection mechanism. If the RTT between two successive nodes is higher than the threshold value, a wormhole attack is suspected. If a wormhole is suspected, node's transitory buffer is probed to determine whether the long delay between the nodes is due to wormhole or not, as delays can be caused due to congestion or by queuing delays.

## 3. Proposed wormhole detection method

Detecting wormholes in WSNs is essential since they can make the routing protocols malfunction. In this paper, a highly efficient wormhole detection method named TRM is developed, which uses the local neighborhood information to calculate the transmission range.

### 3.1. Network model

In order to prepare for the discussion of the wormhole detection, the network model is presented first. In the network model, a WSN with N sensor nodes is considered, which can be denoted by a directed graph $G = (V, E)$. In this graph, $V$ is the set of vertices indicating the sensor nodes and E is the set of direct edges indicating the wireless links in the graph. The graph takes a Unit Disk Graph (UDG) [17] as its connectivity model. In UDG, each node is modeled as a disk of unit radius in the plane, which indicates the transmission range of a single node. Each node is a neighbor of all nodes located in its disk. Nodes are randomly distributed in the specified area. Two types of nodes are considered in the network: normal nodes and malicious nodes placed by the adversary. Malicious nodes differ from normal nodes in their transmission range, power and calculation capability.

### 3.2. Adversary model

As described in Section 1, one end of the wormhole eavesdrops on packets, tunnels them through the wormhole and replays them at the other end of the wormhole. The adversary can place many pairs of malicious nodes to deploy wormholes across the whole network. The adversary's goal is to attract as more routes through the wormhole link as possible. And as long as the wormholes are placed carefully, the majority of the network routes can be attracted to the wormhole link. To introduce our wormhole detection method, some assumptions must be made first. These three assumptions following lay a foundation for our wormhole detection method.

1. The wormhole link is long enough so the regions of the two end points don't overlap with each other [17]. For example, A and B in Fig.1 are well separated from each other, i.e., they are multi-hops away.

2. There is some time t when the wormhole is absent, so the sensor nodes have enough time to establish their neighbors.

3. The wormhole is closed [25]. The wormhole attacks are divided into three groups (closed, half open, and open) according to the format of the tunnel and attacker's capability. In this paper, we focus on the closed wormhole attack.

### 3.3. Principle and analysis

In order to explain our wormhole detection method, its principle analysis is presented first. In the network, each node pair can establish a link because their distance is less than or equal to the transmission range $r$. For any node $m$, the neighbor set of $m$ is denoted by $N(m)$. For example, if a node $B$ can receive packets from node $A$ with one hop, node $B$ is a neighbor of node $A$ and meets $B \in N(A)$. The principle is to check the neighbor topology by using the geometric relationship of nodes' locations under the constraint of the communication range of the two involved sensor nodes.

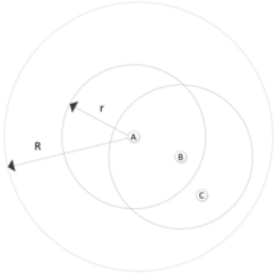



**Fig. 2.** Neighbor Nodes without Wormhole

The principle is illustrated in Fig.2 by studying the geometric relationship among nodes in the network without wormholes. Node $A$ and $B$ are two neighbor nodes to be checked. Node $C$ meets $C \in N(B)$ but $C \notin N(A)$. The transmission range of node $A$, $B$ and $C$ is $r$. When node $A$ adjusts its transmission range to $R = 2r$ in Fig.2, all the neighbors of node $C$ become neighbors of node $A$. So it meets that $C \in N(A)$ and $N(B) \subseteq N(A)$.

The geometric relationship among nodes in the network under wormhole attack is totally different as shown in Fig.3. Node $A$ and $B$ are two neighbor nodes which are connected via the wormhole link. Node $C$ and $D$ both meet that $C, D \in N(B)$. Node $A$, $B$ and $D$ are mutually neighbors due to the wormhole link as described in Section 1.

Node $B$ and $D$ lay in node $A$'s neighbor list due to the wormhole link. Node $C$ is far from the wormhole end point and thus free from wormhole attack. The transmission range of these four nodes is $r$ at first. Then the transmission range of node $A$ is expanded to $R = 2r$. Node $D$ is node $A$'s neighbor connected by the wormhole link. However, since node $A$ and $B$ are multi-hops away from each other, node $C$ is still not a neighbor of node $A$ even though the radius of node $A$ is doubled. After increasing the radius of

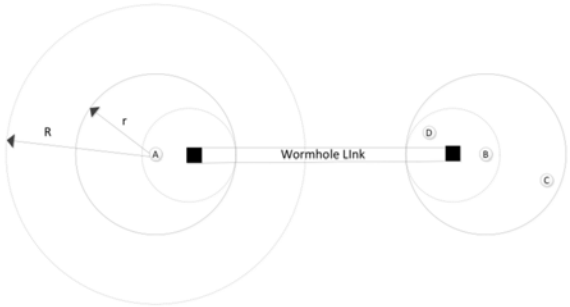



**Fig. 3.** Neighbor Nodes with Wormhole

node $A$, one of node $B$'s neighbors is still not a neighbor of node $A$. So it meets that $D \in N(A)$ and $C \notin N(A)$. As a result, not all the neighbors of node $B$ turn into neighbors of node $A$, which meets that $N(B) \not\subset N(A)$. And this can be used to check whether there exists a wormhole between two sensor nodes.

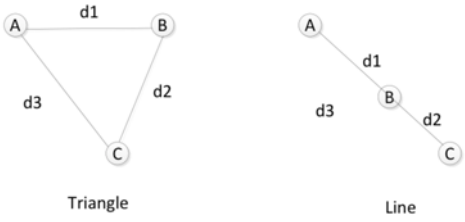



**Fig. 4.** Neighbor Nodes with Wormhole

Then we make some calculations to prove that the above principle is feasible. As shown in Fig.4, the distance between $A$ and $B$ is denoted by $d1$; the distance between $B$ and $C$ is denoted by $d2$, the distance between $A$ and $C$ is denoted by $d3$. There are two cases of node-relative position: triangle and line. According to the neighbor relationship described above and their transmission range $r$, it is obvious that $d1 \leq r$ and $d2 \leq r$. In the triangle case, it can be seen that $d3 < d1 + d2 \leq 2r$. In the line case, it can be seen that $d3 = d1 + d2 \leq 2r$. So we can get $d3 \leq 2r$. Since the radius of node A is $R = 2r$, node $C$ is within node $A$'s transmission range. And it meets that $d3 \leq 2r$ for $\forall C \in N(B)$. Therefore, we can get the formula $N(B) \subseteq N(A)$. When the network is under the wormhole attack, the actual distance of the two neighbor nodes $A$ and $B$ may be very far away. It may meet that $d3 > 2r$ for $\forall C \in N(B)$. Node $C$ is still not a neighbor of node $A$ after expanding its radius to $2r$. But due to the wormhole, some node like $D$ in Fig.3 may still be a neighbor of node $A$, which means that $C \notin N(A)$, $\exists C \in N(B)$. Therefore, we can get $N(B) \not\subset N(A)$. Now we can get the conclusion that:

1. When there is a wormhole and the transmission range of node $A$ is $R$, there must exist a node $C \in N(B)$ but $C \notin N(A)$.

2. When there is no wormhole and the transmission range of node $A$ is $R$, all nodes $C \in N(B)$ meet $C \in N(A)$.

### 3.4. Detection procedure

Based on the principle of detecting wormholes, detailed detection procedure will be presented in this section. Two neighbor nodes such as node $A$ and $B$ are to be checked which has its neighbor list $N(A)$ and $N(B)$ separately. The neighbor list information can be exchanged between neighbors through periodic beacon messages. After nodes $A$ and $B$ exchange the neighbor list information, the detection procedure will begin. Node $A$ notifies all its neighbors in $N(A)$ through its beacon messages that will increase its transmission radius. The neighbor nodes receiving this notification will not change their transmission radius in the next beacon time. Then node $A$ increases its transmission range to $2r$ and updates its neighbor list $N(A)$. Finally, node $A$ compares $N(A)$ and $N(B)$:

1. If the neighbor lists $N(A)$ and $N(B)$ satisfy $N(B) \subseteq N(A)$, then there is no wormhole link between node $A$ and $B$.
2. If the neighbor lists $N(A)$ and $N(B)$ satisfy $N(B) \not\subset N(A)$, then there is a wormhole link between node $A$ and $B$.

The node $A$ and $B$ in Fig.4 is used as two tested nodes to describe the main wormhole detection procedure of TRM algorithm. The flow of wormhole detection is shown in the Fig.5. In our model, every node has a current list of its neighbors. Moreover, the neighbor list is regularly updated. Each node can request its neighbors to get their neighbor lists by transmitting a beacon message to its neighbors. Finally, each node can know one-hop neighbor information and two-hop neighbors as well. After a node starts the wormhole detection process, the node first broadcasts a beacon message including a packet to notify its neighbors, which will increase the transmission range. All nodes receiving this notification will not change their transmission range in the next beacon period. After sending the message, the transmission range of node $A$ is increased to $2r$. If the neighbors of node $B$ are still neighbors of node $A$, node $A$ will search from the neighbor list in the next beacon period. If $B$'s one neighbor, node $C$, is still not a neighbor of $A$, a wormhole will be detected. From Fig.5, we can see that communication links between nodes are required to establish in the primary stage. Then a node adopts the neighbor discovery mechanism to establish the link with other node. During the discovery stage, every node will send its own neighbor list to its neighbors by sending beacon frames. By this way, each node can get its neighbor information within two hops. Finally, the network topology will be established. The beacon information will be transmitted at regular intervals. After changing the radius, a test node will update its neighbor node list in the next beacon time. By comparing its current neighbor list with the previous list, a test node can find the existence of false topology that does not exist in a normal network. Then the wormhole is detected.

In some wormhole detection methods based on statistical analysis, the algorithm calculates the link frequency statistics for some time to determine the presence of a wormhole. This method must work after the routes are established and transmission is observed for some time. TRM algorithm can begin execution before the route establishment phase causing a large number of packets to be transmitted to the base station. In this way, wormholes can be detected before the network traffic to be sent. Then the administrator of the

network can eliminate the bad effects of wormholes. The description of the algorithm is shown in Table 1.

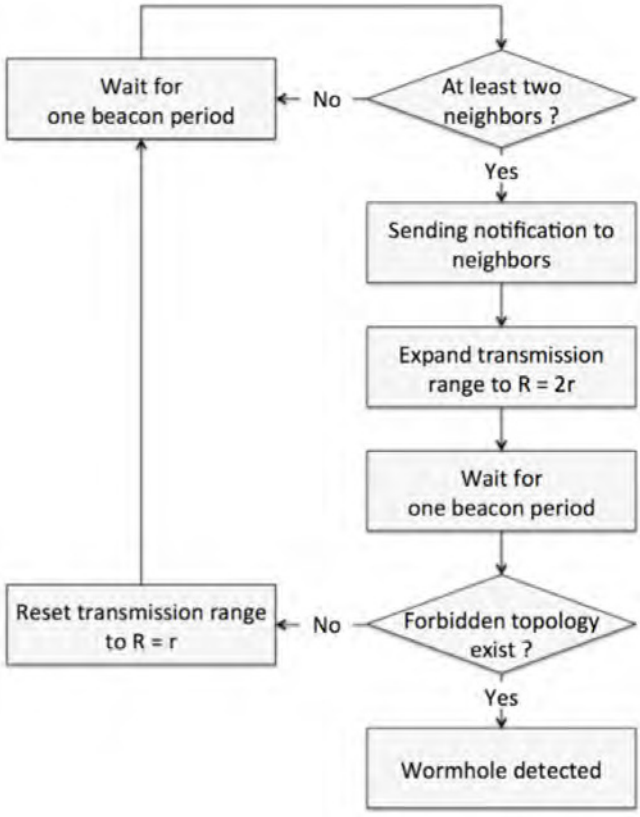



**Fig. 5.** Wormhole Detection Process

### 3.5. Complexity and feasibility analysis

In order to demonstrate that our algorithm is a lightweight one, the complexity of the wormhole detection is analyzed from the aspects of time complexity and space complexity. The time complexity is the time consumed by executing the algorithm. In order to obtain the time complexity, the consumption time of detecting a pair of wormhole nodes is calculated firstly. Suppose there is a wormhole between node $A$ and node $B$. The algorithm needs to find a node in $N(B)$ but not in the neighbor list of node $A$. Since the number of neighbors is a constant $c$, the time complexity of wormhole detection is $O(C)$, i.e., $O(1)$. Secondly, the consumption time of detecting all pairs of wormhole nodes is calculated. At this time, every node and its neighbors should be checked. When the num-

**Table 1.** Transmission Range based Method to Detect Wormholes

| Line | Description |
|---|---|
| 1 | **Given:** Network $N$ with node radius $r$, wormhole number $c = 0$ |
| 2 | **While** check every node $m$ in $N$ **do** |
| 3 | Expand radius of $m$ to $R = 2r$ |
| 4 | **For** each node $n$ in $N(m)$ **do** |
| 5 | **If** there exists once $d \in N(n)$ **and** $d \notin N(m)$ |
| 6 | **then** $c + 1$ |
| 7 | **end for** |
| 8 | **end while** |

ber of nodes is limited such as $n$ and the number of its neighbors is $c$, the time complexity of TRM is $O(cn)$, i.e., $O(n)$.

The space complexity is defined as the storage space. In the TRM, the space complexity is influenced by the number of nodes in the network. According to our algorithm, except the neighbor list, no extra data structures are required to store in TRM. Suppose there is a wormhole between node $A$ and node $B$. Because only neighbor information is stored, the space complexity is obviously $O(1)$. When all the $n$ nodes in the network are checked, the space complexity is $O(n)$. The feasibility of the algorithm is that every node must have its neighbor nodes. Suppose $n$ nodes are distributed in a square region with the side length $d$ and the transmission radius $r$. According to TRM algorithm, the number of nodes in each row is $\sqrt{n}$ lying on a line of length $d$. The distance between two neighbor nodes is $\frac{d}{\sqrt{n}-1}$. Every node can communicate with each other as long as the distance between neighbor nodes is less than the node's transmission radius. So it should be met $\frac{d}{\sqrt{n}-1} \leq r$, which is easy to implement. However, there may be some particularly isolated nodes, which doesn't make sense for the wormhole attacker. In summary, the feasibility of the proposed algorithm has been verified.

## 4. Simulation analysis

In order to verify the performance of our wormhole detection method, various experiments have been carried out. In the simulated system scenario, the wireless sensor network consists of 100 sensor nodes. First, we show the great damage of wormhole attack to the network. Among the entire nodes, ten source nodes and ten destination nodes are selected randomly. Then routes are established between those source and destination nodes. The routes are set up using the basic Shortest Path Algorithm for simplicity. Then it can be seen in the simulation as shown in Fig.5 that the routes are badly corrupted due to the existence of wormhole. The routes are broken since the routes cross the wormhole end points. In this way, the traffic can be attracted to the wormhole link and the adversary can mount further attack like sinkhole attack or just eavesdrop on the information:

In the experiments, the nodes are distributed randomly in 5x5, 10x10, 15x15, and 20x20 square separately. The node transmission range is 2 meters and nodes are distributed randomly, which forms a unit disk graph for universality. The wormholes are also placed in a random way.

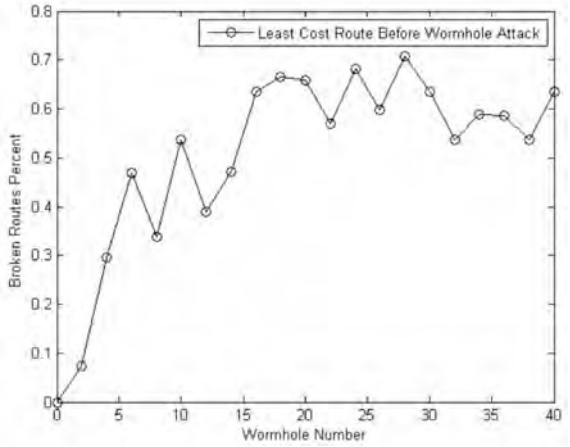



**Fig. 6.** The Broken Routes Percent by Wormholes

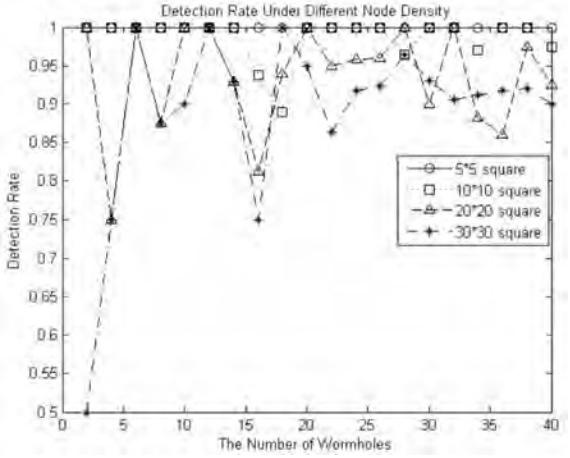



**Fig. 7.** Simulation Results of Wormhole Detection Rate

In Fig.7, the wormhole detection rate is calculated as the number of wormholes increases from 2 to 40. The detection rate is also compared under different system scenario in which the networks with the same number of wormholes have different node densities. Network distribution area 5x5 corresponds to the greatest node density. And the node density decreases as the network distribution area increases to 10x10, 20x20, and 30x30. It can be seen from Fig.6 that the bigger the node density, the higher the detection rate. The detection rate is perfectly 100% when the side length of the network is square since it's easy to detect wormholes when a node has many neighbors. A node's detection failure can be complemented by another neighbor node. The detection rate is not 100% because some neighbor nodes around the wormhole can't detect the wormhole link. The detection may fail because the node has nearly no neighbor to check the local neighborhood information using our method. This situation, which is of low probability in practical application, happens in very spare network or some isolated sensor nodes. Moreover, there is no worth for the adversary to attack such isolated sensor nodes because little traffic will be caused to use by the attack.

To compare the performance of TRM with other wormhole detection method, two other kinds of detection methods are simulated in the experiments. The Transmission Time based Mechanism (TTM) [13] detects wormhole attacks during the route setup procedure by computing transmission time between every two successive nodes along the established path. The Four Way Handshaking algorithm (FWH) [18] uses a simple four-way handshaking messages to exchange. It can be seen from Fig.8 when the wormhole length is smaller than 10, our method can achieve the highest detection rate. When the wormhole length is 2, the transmission time of two neighbor nodes created by wormhole link is not too long to be detected. The FWH algorithm is also affected by the time. Our TRM has nothing to do with the time and detect the wormholes according to the geometric relationship of nodes as described in section 3. So our TRM can have high detection rate all the time in different network scenarios.

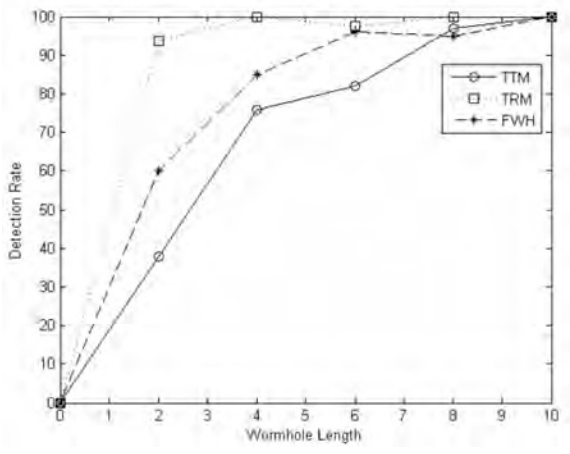



**Fig. 8.** Detection Rate of Different Detection Methods

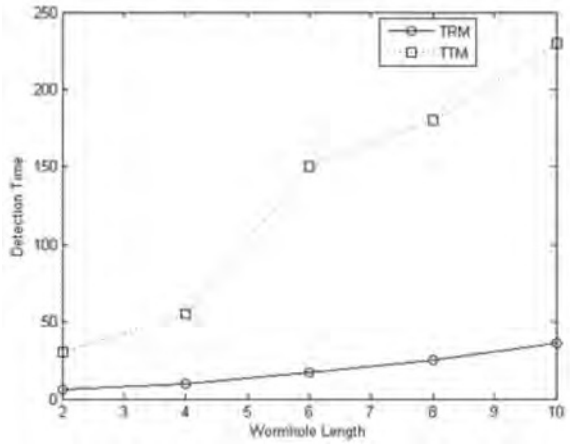



**Fig. 9.** Detection Time Comparison

In Fig.9, the detection time of TRM and TTM algorithm is compared. The actual average transmission time between one-hop nodes is ten milliseconds. However in TTM, the RTT between two nodes connected through the wormhole link is calculated since the two endpoints of wormholes are far away. In TTM, the detection result is obtained through calculating transmission time. So the detection time is longer when the wormholes are far away. It can be seen in Fig.9 that the detection time increases greatly as the length between wormholes increases. In TRM, however, the wormholes are detected by checking the false neighbor topology. The wormholes can be found out by calculating the geometrical relationship between nodes. In this way, the computation is of low complexity and more quick. At the same time, since the node's neighbor list has nothing to do with the length of wormholes, the wormhole length doesn't affect the detection time. So the detection time doesn't increase greatly as the length of wormhole increases.

## 5. Conclusions

Wormhole attack in WSNs has been drawing more and more attention since it can disrupt normal network routing protocols. However, in previous work of wormhole detection, most of them need either extra hardware or clock synchronizations and suffer from high complexity. In this paper, an efficient wormhole detection method is proposed, which is based only on local neighborhood information. Through judging the node's position, we can determine whether the node is in the local network topology affected by the wormhole link.

In the detection procedure, the neighborhood information of each node is updated and exchanged periodically between neighbors along with the increment of the transmission range. A local topology that has a wormhole link finally reports a mismatch of the neighborhood information between nodes. According to the analysis, the algorithm gives $O(n)$ for both of the time complexity and the space complexity.

The simulation results also demonstrate that our wormhole detection method can achieve a high wormhole detection rate. For the simulation, we organized a wireless sen-

sor network with 100 sensor nodes and deployed up to 40 wormholes in it with different density. In case of a denser network with more wormholes, the detection rate was getting higher. In the performance comparison with other detection methods, the proposed TRM gave much bigger detection rate for wormholes with shorter lengths.

In the future, the proposed algorithm is required to enhance the performance for coarse networks and consider the separated nodes as well as optimizing the procedure even for dense networks. Performance of the proposed TRM algorithm also should be evaluated for various network conditions such as the case that the network has frequent link breaks between nodes as a common problem in a practical environment.

**Acknowledgments.** This research was sponsored in part by the Fundamental Research Funds for the Central Universities (No. DUT13JS10). This work was also supported in part by the Soonchunhyang University Research Fund.

## References

1. Agrawal, S., Jain, S., Sharma, S.: A survey of routing attacks and security measures in mobile ad-hoc networks. Journal of Computing 03(01), 41–48 (2011)
2. Akyildiz, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor networks: a survey. Computer Networks 38(04), 393–422 (2002)
3. Ban, X., Sarkar, R., Gao, J.: Local connectivity tests to identify wormholes in wireless networks. In: Proceedings of the 12th ACM International Symposium on Mobile Ad Hoc Networking and Computing. pp. 65–78 (2011)
4. Dong, D., Li, M., Liu, Y., Liao, X.: Wormcircle: connectivity-based wormhole detection in wireless ad hoc and sensor networks. In: Proceedings of the 15th International Conference on Parallel and Distributed Systems. pp. 72–79 (2009)
5. Hu, Y.: Wormhole attacks in wireless networks. IEEE Journal on Selected Areas in Communications 24(02), 370–380 (2006)
6. Hu, Y., Johnson, D., Perrig, A.: Secure efficient distance vector routing for mobile wireless ad hoc networks. Ad Hoc Networks 01(01), 175–192 (2003)
7. Hu, Y., Perrig, A., Johnson, D.: Packet leashes: A defense against wormhole attacks in wireless networks. In: Proceedings of 22nd Annual Joint Conference of the IEEE Computer and Communications. pp. 1976–1986 (2003)
8. Hu, Y., Perrig, A., Johnson, D.: An end-to-end detection of wormhole attack in wireless ad-hoc networks. In: Proceedings of 31st Annual International Computer Software and Applications Conference. pp. 39–48 (2007)
9. Jhaveri, R., Patel, D., Jatin, D., Parmar, D., Shah, B.: Manet routing protocols and wormhole attack against aodv. International Journal of Computer Science and Network Security 10(04), 12–18 (2010)
10. Jhaveri, R., Patel, S., Jinwala, D.: Dos attacks in mobile ad hoc networks: A survey. In: Proceedings of Advanced Computing & Communication Technologies. pp. 535–541 (2012)
11. Johnson, D., Maltz, D., Broch, J.: Dsr: The dynamic source routing protocol for multi-hop wireless ad hoc networks. In: Perkins, C. (ed.) Ad Hoc Networks, pp. 139–172. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA (2001)
12. Khalil, I., Bagchi, S., Shroff, N.: Liteworp: a lightweight countermeasure for the wormhole attack in multihop wireless networks. In: Proceedings of the International Conference on Dependable Systems and Networks. pp. 612–621 (2005)
13. Khalil, I., Bagchi, S., Shroff, N.: Mobiworp: mitigation of the wormhole attack in mobile multihop wireless networks. Ad Hoc Networks 06(03), 344–362 (2008)

14. Lazos, L., Poovendran, R., Meadows, C., C., S., L., C.: Preventing wormhole attacks on wireless ad hoc networks: a graph theoretic approach. In: Proceedings of the IEEE Wireless Communications and Networking Conference, Broadband Wirelss for the MassesReady for Takeoff. pp. 1193–1199 (2005)
15. Lu, X., Dong, D., Liao, X.: Mds-detection using local topology in wireless sensor networks. International Journal of Distributed Sensor Networks 2012, 1–9 (2012)
16. Maheshwari, R., Gao, J., Das, S.: Detecting wormhole attacks in wireless networks using connectivity information. In: Proceedings of the 26th IEEE International Conference on Computer Communications. pp. 107–115 (2007)
17. Maheshwari, R., Gao, J., Das, S.: Detecting wormhole attacks in wireless networks using connectivity information. In: Proceedings of 26th IEEE International Conference on Computer Communications. pp. 107–115 (2007)
18. Nat-Abdesselam, F., Bensaou, B., Yoo, J.: Detecting and avoiding wormhole attacks in optimized link state routing protocol. In: Proceedings of IEEE Wireless Communications and Networking Conference. pp. 3117–3122 (2007)
19. Patel, K., Manoranjitham, T.: Detection of wormhole attack in wireless sensor network. International Journal of Engineering Research & Technology 02(05), 366–369 (2013)
20. Poovendran, R., Lazos, L.: A graph theoretic framework for preventing the wormhole attack in wireless ad hoc networks. Wireless Networks 13(01), 27–59 (2007)
21. Sebastian, M., Kumar, A.: A novel solution for discriminating wormhole attacks in manets from congested traffic using rtt and transitory buffer. I. J. Computer Network and Information Security 05(08), 28–38 (2013)
22. Sharma, K., Ghose, M.: Wireless sensor networks: an overview on its security threats. IJCA, Special Issue on Mobile Ad-hoc Networks pp. 42–45 (2010)
23. Sohrabi, K., Gao, J., Ailawadhi, V., Pottie, G.: Protocols for self-organization of a wireless sensor network. IEEE Personal Communications 07(05), 16–27 (2000)
24. Triki, B., Rekhis, S., Boudriga, A.: A novel secure and multipath routing algorithm in wireless sensor networks. In: Proceedings of 2010 International Conference on Data Communication Networking. pp. 1–10 (2010)
25. Wang, W., Bhargava, B., Lu, Y., Wu, X.: Defending against wormhole attacks in mobile ad hoc networks. Wireless Communications and Mobile Computing 06(04), 483–503 (2006)

**Guowei Wu** received B.E. and Ph.D. degrees from Harbin Engineering University, China, in 1996 and 2003, respectively. He was a Research Fellow at INSA of Lyon, France, from September 2008 to March 2010. He has been an Associate Professor in School of Software, Dalian University of Technology (DUT), China, since 2003. Dr. WU has authored three books and over 20 scientific papers. His research interests include embedded real-time system, cyber-physical systems (CPS), and wireless sensor networks.

**Xiaojie Chen** received B.E. and Master degrees from Dalian University of Technology, China, in 2010 and 2013, respectively. He is an engineer in China Unicom. His research interests include embedded real-time system, cyber-physical systems (CPS), and wireless sensor networks.

**Yao Lin** received B.E. and Master degrees from Harbin Engineering University, China, in 1998 and 2001, respectively, and received Ph.D. degree from Dalian University of Technology, China in 2011. She has been a lecturer in School of Software, Dalian University of

Technology (DUT), China, since 2004. She has co-authored one book and over ten scientific papers. Her research interests include pervasive computing, cyber-physical systems (CPS), and wireless sensor networks.

**Youngjun Lee** received B.E. degree from Dept. of Information Security Engineering, Soonchunhyang University, Korea, in 2013. He is currently pursuing his Master degree. His research interests include malware analysis, secure hardware design, and CPS security and testing.

**Kangbin Yim** received his B.S., M.S., and Ph.D. from Ajou University, Korea in 1992, 1994 and 2001, respectively. He is currently a Full Professor in the Department of Information Security Engineering and the founding director of the R&BD Center for Security and Safety Industries (SSI) in Soonchunhyang University. He has served as the executive board member of Korea Institute of Information Security and Cryptology, Korean Society for Internet Information and The Institute of Electronics Engineers of Korea. He also has served as editor of the journals such as JIT, MIS, IJCM, JCPS, JISIS and JoWUA. His research interests include vulnerability assessment, malware analysis, embedded systems security, and software-hardware co-design and evaluation. Related to these topics, he has worked on more than fifty research projects and published more than a hundred research papers.

*Received: September 21, 2013; Accepted: February 27, 2014.*

Computer Science and Information Systems 11(3):1143–1156 DOI: 10.2298/CSIS130925069W

# The Performance Analysis of Direct/Cooperative Transmission to Support QoS in WLANs

Chien-Erh Weng[1], Jyh-Horng Wen[2], Hsing-Chung Chen[3], and Lie Yang[1]

[1] Department of Electronic Communication Engineering, National Kaohsiung Marine University, Kaohsiung, Taiwan. ROC
ceweng@mail.nkmu.edu.tw
[2] Department of Electrical Engineering, Tunghai University
Taichung, Taiwan. ROC
horngwen528@gmail.com
[3] Department of Computer Science and Information Engineering, Asia University
Taichung, Taiwan. ROC
cdma2000@asia.edu.tw

**Abstract.** In the past decades, cooperative communications schemes have gained significant attention in wireless networks. The cooperative scheme leads to longer transmission time which can considerably degrade the system performance. We evaluate the saturation throughput and saturation delay of the Markov chain model with direct/cooperative schemes to support QoS in WLANs. Simulation results show that differentiating the contention window size is better than differentiating the arbitration interframe space in terms of throughput and delay.

**Keywords:** cooperative scheme, throughput, delay, Markov chain model, QoS, WLANs

## 1. Introduction

In recent years, the cooperative communications market is experiencing an explosive growth. With the introduction of relays, an auxiliary channel, the relay channel to the direct channel between the source and destination can be generated. That is, the relays help forwarding the signal from the source to the destination [1]. As a result, spatial diversity which ameliorates the frame error rate is generated via the help of relay channel. On the other hand, cooperative scheme leads to longer transmission time which can considerably degrade the system performance. There have been many performance analyses of the cooperative communication systems. Yan Zhu et al [2] showed the effectiveness of utilizing collaborative relays in a large-scale network is penalized by the elevated level of interference. G. Jakllar et al [3] showed that virtual multiple-input single-output (MISO) transmissions can improve the performance and be robustness to link failures due to mobility and interference and the advantage of using virtual antenna arrays is it does not require and additional hardware. Zhiguo Ding et al [4] proposed a spectrally efficient strategy for cooperative multiple access systems in multiple-users environment and it can achieve more robust performance than the direction

transmission. K. Lee et al [5] focused on the concept of power consumption and examined the performance of heterogeneous cooperative networks with the source that do not act as relays and relays that are dedicated to relaying functions with concern about power consumption. Most of research mainly focused on the designs of cooperative protocol schemes and how to gain benefits of spatial diversity based on information theory. In order to evaluate the system performance, a suitable analytic model that combines the traditional direction transmission and the cooperative transmission from the medium access control (MAC) perspectives should be exploited and with the population of multimedia applications, including the transport of voice, audio and video over WLANs, there is a clear need to support quality of service (QoS) guarantees. In this paper, we utilize the Markov chain model with direct/cooperative transmission scheme to support QoS guarantees from the MAC perspectives to analyze the saturation throughput and saturation delay.

The rest of this paper is organized as follows. An overview of the system model is depicted in Section II. The performance analysis of the model is depicted in the Section III. The simulation results are shown in Section IV. Finally, Section V gives the conclusions.

## 2. The System Model

To analyze the performance of the Markov chain model, we follow the considerations of [11]. We assume a fixed number Ni of contending stations in the network and a given station in the priority i class (i = 0, 1, ... , n-1). Let b(*i, t*) be the stochastic process representing the backoff timer of a given station at slot time t (note that the backoff timer is stopped when the station senses that channel is busy). The value of the backoff timer is uniformly chosen in the range (0, $W_{i,j-1}$) and depends on the station's backoff stage *j*. For convenience, we define that

$$W_{i,j} = \begin{cases} 2^j CW_{i,min} & 0 \le j < m \\ CW_{i,max} & m \le j \le m + r \end{cases} \quad (1)$$

where $CW_{i,min}$ is the minimum contention widow for the priority *i* class and $CW_{i,max}$ is the maximum contention widow for the priority *i* class, and *m* is the maximum backoff stage. Moreover, let s(*i, t*) be the stochastic process representing the backoff stage *j* of the station at time t. On this condition, we can describe the state of each station in the priority *i* class is as {*i, j, k*}, where *j* stands for the backoff stage and *k* stands for the backoff timer.

There is another state in our model, additional idle state, denoted by {*i,* -1}. The backoff procedure is activated whenever a station has a frame to transmit and senses the channel is busy or whenever the transmitting station infers a failed transmission. If the station verifies its current transmission is successful and senses the channel is idle for arbitration inter-frame spacing in priority *i* class (AIFS[*i*]) duration, it enters into the {*i,* −1} state. If the station is at {*i,* -1} state, whenever it senses the channel is idle for AIFS[*i*] duration, it transmits its frame without entering the backoff procedure.

The state transition diagram of the Markov chain model in the priority *i* class shown in Fig. 1 has the following transition probabilities:

The station transmits its frame without entering the backoff procedure if it senses that its previous transmission was successful and the channel is idle for AIFS[*i*] duration.

$$P\{i,-1|i,-1\}=\left(1-P_{i,dir}\right)\left(1-P_{i,b}\right). \tag{2}$$

The station defers the transmission of a new frame and enters stage 0 of the backoff procedure if it detects a collision occurred or it senses the channel is busy.

$$P\{i,0,k|i,-1\}=\left(P_{i,b}+P_{i,dir}-P_{i,b}P_{i,dir}\right)/W_{i,0},\ 0\le k\le W_{i,0}-1. \tag{3}$$

The backoff timer is stopped when the station senses that channel is busy.

$$P\{i,j,k|i,j,k\}=P_{i,b},\ 0\le j\le m+r,\ 0\le k\le W_{i,j}-1. \tag{4}$$

The backoff timer decreases when the station senses that the channel is idle.

$$P\{i,j,k|i,j,k+1\}=1-P_{i,b},\ 0\le j\le m+r,\ 0\le k\le W_{i,j}-2. \tag{5}$$

The station chooses a backoff delay of stage 0 if its current transmission was successful and it senses that the channel is busy when it tries to transmit a new frame.

$$\begin{aligned}P\{i,0,k|i,j,0\}&=\frac{\left(1-P_{i,dir}\right)P_{i,b}}{W_{i,0}},\ 0\le j\le \ell-1,\ 0\le k\le W_{i,0}-1.\\ P\{i,0,k|i,j,0\}&=\frac{\left(1-P_{i,coop}\right)P_{i,b}}{W_{i,0}},\ \ell\le j\le m+r-1,\ 0\le k\le W_{i,0}-1.\end{aligned} \tag{6}$$

Where $\ell$ is the backoff stage to distinguish the strategy adopting cooperative transmission, the parameters $P_{i,dir}$ is the probabilities in the priority *i* class for receiving incorrect frame at the destination via the traditional direction transmission, $P_{i,b}$ is the probability that the station in the priority *i* class senses that the channel is busy.

The station enters into the {*i*, -1} state if it verifies its current transmission is successful and senses the channel is idle for AIFS[*i*] duration.

$$\begin{aligned}P\{i,-1|i,j,0\}&=\left(1-P_{i,dir}\right)\left(1-P_{i,b}\right),\ 0\le j\le \ell-1.\\ P\{i,-1|i,j,0\}&=\left(1-P_{i,coop}\right)\left(1-P_{i,b}\right),\ \ell\le j\le m+r-1.\end{aligned} \tag{7}$$

The station choices a backoff delay of next stage *j* after an unsuccessful transmission at stage *j*−1.

$$\begin{aligned}P\{i,j,k|i,j-1,0\}&=P_{i,dir}/W_{i,j},\ 1\le j\le \ell.\\ P\{i,j,k|i,j-1,0\}&=P_{i,coop}/W_{i,j},\ \ell+1\le j\le m+r.\end{aligned} \tag{8}$$

When the station has reached the last stage of backoff procedure, it would drop the current frame and enter $\{i, 0, k\}$ state if it detects its current transmission is still failed and the channel is busy during an AIFS[$i$] duration.

$$P\{i,0,k|i,m+r,0\} = P_{i,b}/W_{i,0},\ 0 \le k \le W_{i,0}-1. \tag{9}$$

When the station has reached the last stage of backoff procedure, it would drop the current frame and enter $\{i, -1\}$ state if it detects its current transmission is still failed and the channel is idle for AIFS[$i$] duration.

$$P\{i,-1|i,m+r,0\} = 1-P_{i,b}. \tag{10}$$

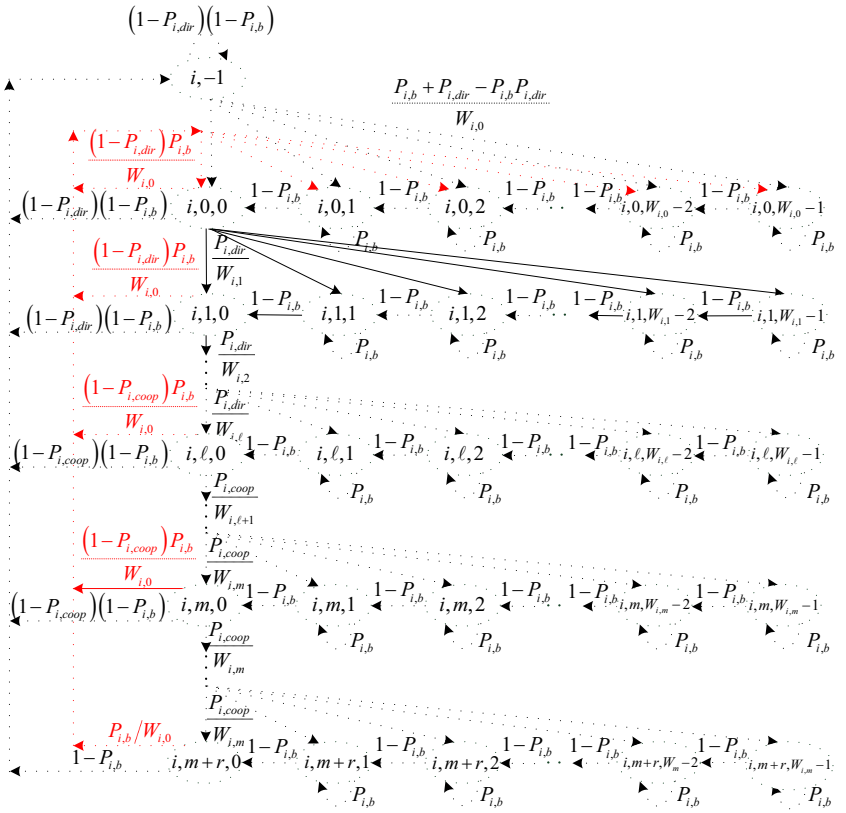



**Fig. 1.** Markov chain model with direct/cooperative strategy for the priority i

The parameters $P_{i,dir}$ and $P_{i,coop}$ are the probabilities in the priority $i$ class for receiving incorrect frame at the destination via the traditional direction transmission and the cooperative transmission, respectively. Note that the unsuccessful reception of frames at the destination is considered to result from either the frame collision or the channel noise. Thus, the parameters $P_{i,dir}$ and $P_{i,coop}$ can be expressed as

$$P_{i,dir} = 1-\left(1-FER_{dir}\right)\left(1-P_{i,c}\right),$$
$$P_{i,coop} = 1-\left(1-FER_{coop}\right)\left(1-P_{i,c}\right), \tag{11}$$

where $P_{i,c}$ is the probability that the transmitted frame collides for the priority $i$ class. $FER_{dir}$ and $FER_{coop}$ are the frame error rates at the destination via the traditional direction transmission and the cooperative transmission, respectively.

We have to calculate the probability that a station in the priority $i$ class is at state $\{i,j,k\}$.

Let $b_{i,j,k} = \lim_{t\to\infty} P\{s(i,t)=j, b(i,t)=k\}$ be the stationary distribution of the Markov chain [6]. In steady-state we have following relations:

$$b_{i,j,0} = P_{dir}^{j} b_{i,0,0}, \quad 0 \le j \le \ell.$$
$$b_{i,j,0} = P_{dir}^{\ell} P_{coop}^{j-\ell} b_{i,0,0}, \quad \ell+1 \le j \le m+r. \tag{12}$$

$$b_{i,j,k} = \frac{1}{1-P_{i,b}} \frac{W_{i,j}-k}{W_{i,j}} b_{i,j,0}. \tag{13}$$

$$b_{i,-1} = \frac{1-P_{i,b}}{P_{i,b}+P_{i,dir}-P_{i,b}P_{i,dir}} b_{i,0,0}. \tag{14}$$

Let $\tau_i$ be the probability that a station in the priority $i$ class transmits its frame during a slot time. A station in the priority $i$ class transmits its frame when its backoff timer reaches zero, regardless of the backoff stage, i.e. the station is at any of the $b_{i,j,0}$ states or at the $b_{i,-1}$ state. Therefore, we have

$$\begin{aligned}\tau_i &= b_{i,-1} + \sum_{j=0}^{m+r} b_{i,j,0} \\ &= \left(\frac{1-P_{i,b}}{P_{i,b}+P_{i,dir}-P_{i,b}P_{i,dir}} + \frac{1-P_{i,dir}^{\ell+1}}{1-P_{i,dir}} + P_{i,dir}^{\ell} P_{i,coop} \frac{1-P_{i,coop}^{m+r-\ell}}{1-P_{i,coop}}\right) b_{i,0,0}\end{aligned} \tag{15}$$

Let $N_i$ (i = 0, 1, ..., n− 1) denote the number of station in the priority $i$ class and $P_t$ denote the probability that there is at least one transmission in a slot time, i.e., there is at least one station transmits during a slot time. Therefore, we have

$$P_t = 1-\prod_{h=0}^{n-1}\left(1-\tau_h\right)^{N_h}. \tag{16}$$

Let $P_{i,s}$ denote the probability that the transmission is successful during a slot time for the priority $i$ class, i.e., a transmission is assumed to be successful when only one station transmits. So we have

$$P_{i,s} = n_i \tau_i \left(1-\tau_i\right)^{N_i-1} \prod_{h=0,h\neq i}^{n-1} \left(1-\tau_h\right)^{N_h}. \tag{17}$$

Let $P_{i,b}$ be the probability that the station in the priority $i$ class senses that the channel is busy when it is trying to decrease its backoff timer in a slot time. The probability $P_{i,b}$ that the station in the priority $i$ class senses that the channel is busy is given by

$$P_{i,b} = 1-\left(1-\tau_i\right)^{N_i-1} \prod_{h=0,h\neq i}^{n-1} \left(1-\tau_h\right)^{N_h}. \tag{18}$$

Moreover, let us introduce the parameter $P_{i,r}$ that is the probability for the priority $i$ class of the traditional direction transmission considering at least one transmission happens. Thus, we have

$$\begin{aligned} P_{i,r} &= \frac{b_{i,-1} + \sum_{h=0}^{\ell-1} b_{i,h,0}}{b_{i,-1} + \sum_{j=0}^{m+r} b_{i,j,0}} \\ &= \frac{\dfrac{1-P_{i,b}}{P_{i,b}+P_{i,dir}-P_{i,b}P_{i,dir}} + \dfrac{1-P_{i,dir}^{\ell}}{1-P_{i,dir}}}{\dfrac{1-P_{i,b}}{P_{i,b}+P_{i,dir}-P_{i,b}P_{i,dir}} + \dfrac{1-P_{i,dir}^{\ell+1}}{1-P_{i,dir}} + P_{i,dir}^{\ell} P_{i,coop} \dfrac{1-P_{i,coop}^{m+r-\ell}}{1-P_{i,coop}}} \end{aligned} \tag{19}$$

## 3. Performance Analysis

In this section, the purpose of our analysis is to evaluate the saturation throughput and the delay performances of Markov chain model with traditional direction and cooperative transmission strategies. Based on the previous description, we can derive the close forms for system performance metrics of saturation throughput and delay.

### 3.1. Throughput Analysis

Let $S_i$ denote the normalized saturation throughput of a given priority $i$ class [7]. We can express it as (20). The parameter $E[T_{P,i}]$ is the average duration of transmitting payload information successfully in a slot time for the priority $i$ class, which is derived as

$$\begin{aligned} S_i &= \frac{E\left[T_{P,i}\right]}{E\left[T_B\right] + \sum_{i=0}^{n-1} E\left[T_{S,i}\right] + E\left[T_{C,i}\right] + E\left[T_{E,i}\right]}. \\ E\left[T_{P,i}\right] &= P_{i,s}P_t\left[P_{i,r}\left(1-FER_{dir}\right) + \left(1-P_{i,r}\right)\left(1-FER_{i,coop}\right)\right]T_{payload}, \end{aligned} \tag{20}$$

where $T_{payload}$ is the average duration to transmit the payload information. The parameter $E[T_B]$ is the average duration of non-frozen backoff timer. And the parameters $E[T_{S,i}]$, $E[T_{C,i}]$ and $E[T_{E,i}]$ are the average duration for the priority $i$ class of the successful transmission, the transmitted frame colliding and the transmitted frame is error due to the channel noise, respectively. Those parameters can be derived as

$$E[T_B] = (1-P_t)\sigma. \tag{21}$$

$$E[T_{S,i}] = P_{i,s}P_t\left[P_{i,r}(1-FER_{dir})T^s_{i,dir} + (1-P_{i,r})(1-FER_{coop})T^s_{i,coop}\right]. \tag{22}$$

$$E[T_{C,i}] = P_{i,c}P_t\left[P_{i,r}T^c_{i,dir} + (1-P_{i,r})T^c_{i,coop}\right]. \tag{23}$$

$$E[T_{E,i}] = P_{i,s}P_t\left[P_{i,r}FER_{dir}T^s_{i,dir} + (1-P_{i,r})FER_{coop}T^s_{i,coop}\right]. \tag{24}$$

The parameters of above equations can be obtained as follows. σ is the size of a slot time. As mentioned before, $FER_{dir}$ and $FER_{coop}$ are the frame error rates at the destination via the traditional direction transmission and the cooperative transmission, respectively. $T^c_{i,dir}$ and $T^c_{i,coop}$ are the average durations for the priority $i$ class that the channel is captured with a successful transmission via the traditional direction transmission and the cooperative transmission, respectively. Similarly, $T^c_{i,dir}$and $T^c_{i,coop}$ are the average duration for the priority $i$ class that the channel is captured with a collision. Note that the average time to detect the error frame is considered the same as that to receive the frame successfully. The values of the above durations depend on the channel access method and are defined as follows.

$$\begin{aligned}
T^s_{i,dir} &= T_{header} + T_{payload} + \delta + SIFS + T_{ACK} + \delta + AIFS[i],\\
T^s_{i,coop} &= 2\left(T_{header} + T_{payload} + \delta + SIFS\right) + T_{ACK} + \delta + AIFS[i].\\
T^c_{i,dir} &= T_{header} + T_{payload} + \delta + AIFS[i],\\
T^c_{i,coop} &= 2\left(T_{header} + T_{payload} + \delta\right) + AIFS[i].
\end{aligned} \tag{25}$$

RTS/CTS mechanism,

$$\begin{aligned}
T^s_{i,dir} &= T_{RTS} + \delta + SIFS + T_{CTS} + \delta + SIFS + T_{header} + T_{payload} + \delta + SIFS + T_{ACK} + \delta + AIFS[i],\\
T^s_{i,coop} &= T_{CRTS} + \delta + SIFS + T_{CTS} + \delta + SIFS + 2\left(T_{header} + T_{payload} + \delta + SIFS\right) + T_{ACK} + \delta + AIFS[i].\\
T^c_{i,dir} &= T_{RTS} + \delta + AIFS[i],\\
T^c_{i,coop} &= T_{CRTS} + \delta + AIFS[i].
\end{aligned} \tag{26}$$

The parameters $T_{header}$, $T_{ACK}$, $T_{RTS}$, $T_{CRTS}$ and $T_{CTS}$ are the durations to transmit the header, ACK frame, RTS frame, CRTS frame and CTS frame, respectively. And δ is the propagation delay.

### 3.2. Delay Analysis

Saturation delay $D_i$ is the average delay (defined as the time from the generation of a frame to the source is acknowledged by the destination) for the priority $i$ class under the saturation condition and includes the interframe spaces (such as SIFS), the channel

access delay (due to backoff, collisions, etc.) and the transmission delay [8]. Let $X_i$ be the random variable representing the total number of backoff slots for the priority $i$ class without considering the case that the backoff timer is stopped when the channel is sensed busy. The probability that the frame in the priority $i$ class is successfully transmitted at the ($j$+1)th transmission and the average number of backoff slots that the station needs to transmit a frame successfully at the $j$th retry is $\sum_{h=0}^{j}\frac{W_{i,h}-1}{2}$. Thus, we have

$$E\left[X_i\right]=\left(\sum_{j=0}^{\ell}P_{i,dir}^{j}+\sum_{j=\ell+1}^{m+r}P_{i,dir}^{\ell}P_{i,coop}^{j-\ell}\right)\sum_{h=0}^{j}\frac{W_{i,h}-1}{2}P_{i,suce}, \tag{27}$$

where $P_{i,suce}$ is the probability for receiving correct frame at the destination for the priority $i$ class which can be derived as

$$P_{i,suce}=P_{i,s}P_t\left[P_{i,r}\left(1-FER_{dir}\right)+\left(1-P_{i,r}\right)\left(1-FER_{coop}\right)\right]. \tag{28}$$

The probability that channel is sensed idle is (1–$P_{i,b}$). Let Fi be the random variable representing the total number of backoff slots when the backoff timer is stopped for the priority $i$ class. Thus, we can regard E[$X_i$] and E[$F_i$] as the total number idle and busy slots that the frame encounters during backoff procedure , respectively. We have

$$E\left[F_i\right]=\frac{P_{i,b}}{1-P_{i,b}}E\left[X_i\right]. \tag{29}$$

Let E[$BD_i$] denote the average backoff delay that the station in the priority $i$ class experiences before accessing the channel. We have

$$E\left[BD_i\right]=E\left[F_i\right]\left(P_{i,bs}T_{i,s}+P_{i,bc}T_{i,c}\right). \tag{30}$$

The parameters $P_{i,bs}$ and $P_{i,bc}$ are the probabilities that the transmission is successful and the transmitted frame collides on the condition that the channel is busy, respectively. We have

$$\begin{aligned}P_{i,bs}&=\frac{\left(N_i-1\right)\tau_i\left(1-\tau_i\right)^{N_i-2}}{P_{i,b}}\prod_{h=0,h\neq i}^{n-1}\left(1-\tau_h\right)^{N_h},\\ P_{i,bc}&=1-P_{i,bs}.\end{aligned} \tag{31}$$

$T_{i,s}$ and $T_{i,c}$ are the total durations that the channel is captured with a successful transmission and a collision for the priority $i$ class, respectively.

$$\begin{aligned}T_{i,s}&=P_{i,r}\left(1-FER_{dir}\right)T_{i,dir}^{s}+\left(1-P_{i,r}\right)\left(1-FER_{coop}\right)T_{i,coop}^{s},\\ T_{i,c}&=P_{i,r}T_{i,dir}^{c}+\left(1-P_{i,r}\right)T_{i,coop}^{c}.\end{aligned} \tag{32}$$

Let E[$N_{i,retry}$] denote the average number of retries for the priority $i$ class which is derived as

$$E\left[N_{i,retry}\right]=\left(\sum_{j=0}^{\ell}p_{i,dir}^{j}+\sum_{j=\ell+1}^{m+r}p_{i,dir}^{\ell}p_{coop}^{j-\ell}\right)P_{i,suce} \tag{33}$$

As mentioned before, the saturation delay includes the interframe spaces, the channel access delay and the transmission delay. Thus, the delay for the priority *i* class can be derived as

$$\begin{aligned} D_i &= E\left[X_i\right]\sigma + E\left[BD_i\right] + E\left[N_{i,retry}\right]\left(T_{i,c} + T_o\right) + T_{i,s} \\ &= E\left[X_i\right]\sigma + E\left[F_i\right]\left(P_{i,bs}T_{i,s} + P_{i,bc}T_{i,c}\right) + E\left[N_{i,retry}\right]\left(T_{i,c} + T_o\right) + T_{i,s}. \end{aligned} \tag{34}$$

$T_o$ is the duration that a station has to wait when its frame transmission collides before sensing the channel again.

### 3.3. Cost Function Analysis

The optimal performance is achieving by maximizing the throughput and minimizing the delay. There is always a tradeoff between throughput and delay [9]. Thus, we introduce the concept of cost function C [10] that is the tradeoff between throughput and delay to determine the cooperative transmission strategy. The larger value of cost function means that the system performance is better because the throughput is higher and delay is smaller. The cost function is defined as the ratio of the saturation throughput (*S*) to the saturation delay (*D*), which can be derived as

$$C = S/D. \tag{35}$$

## 4. Numerical Results

In this section, we show the results that utilizing the optimal cooperative transmission strategy. The parameters of our analysis are as follows: Frame payload = 1023 bytes, ACK = 14 bytes, RTS = 20 bytes, CTS = 14 bytes, SIFS = 10 us, DIFS = 50 us, propagation delay = 1 us, $CW_{min}$ = 32, $CW_{max}$ = 1024. For demonstration purposes, we adopt four priority classes, i.e., $i$ = 4. And we utilize the default parameter values which are defined in IEEE 802.11e standard.

The saturation throughput performances under different channel conditions are depicted in Fig. 2 to Fig. 3. From the results, we know that the IEEE 802.11e EDCA priority mechanism is quite effective in throughput. The contention window differentiation can provide the different probability of accessing the channel. A station with lower values of backoff parameters ($CW_{min}$ and $CW_{max}$) has higher probability of winning the contention in comparison to station with higher values. Thus, AC_3 class has the highest priority because it has the lowest backoff parameters. The throughput of AC_0 is the same as AC_1 because the parameter $T_{Paylaod}$ of each AC is the same and the contention window parameters of AC_0 is the same as AC_1, i.e., the $E[T_{P,i}]$ of AC_0 is the same as AC_1.

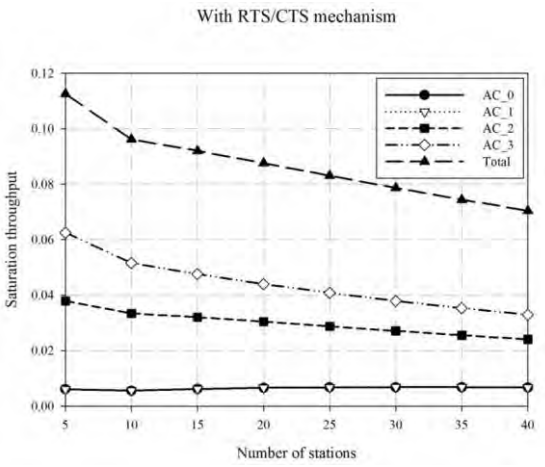



**Fig. 2.** Throughput with RTS/CTS mechanism (FERdir = 0.9)

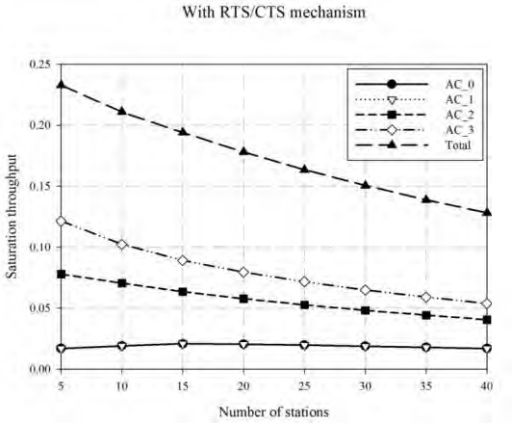



**Fig. 3.** Throughput with RTS/CTS mechanism ($FER_{dir}$ = 0.6)

The delay performances under different channel conditions are depicted in Fig. 4 to Fig. 5. After every busy channel period, each station has to wait for the duration equal to its AIFS value. If the AIFS values are different, there is a time in which the stations with shorter AIFS values (the higher-priority) may access the channel, while the stations with longer AIFS values (lower-priority) are prevented from accessing the channel. Thus, the delay of AC_0 is higher than that of AC_1 because the value of AIFS[AC_0] is larger. The delay of AC_3 class is higher than that of AC_2 class because the parameter $P_{i,b}$ of the AC_3 class is much higher than that of AC_2 (about 1.6 times). Hence, the backoff delay (i.e., E[$BD_i$]) that the station in the priority AC_3 class experiences before accessing the channel is longer than the priority AC_2.

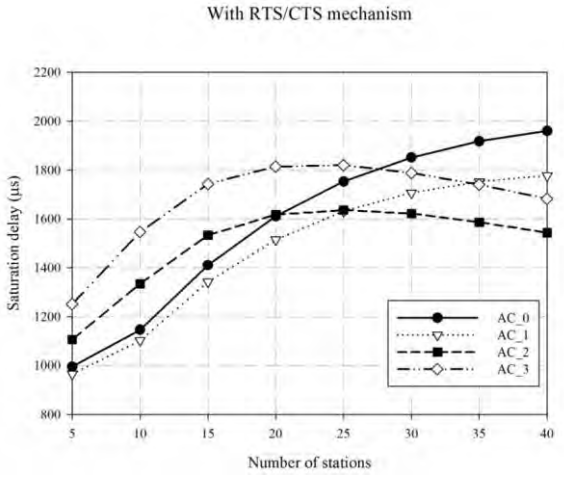



**Fig. 4.** Delay with RTS/CTS mechanism ($FER_{dir}$ = 0.9)

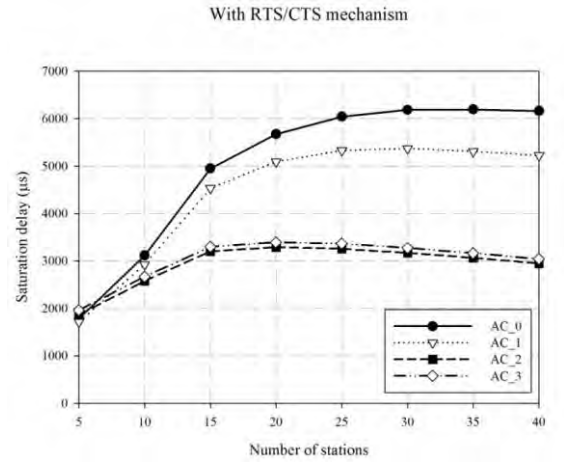



**Fig. 5.** Delay with RTS/CTS mechanism (FERdir = 0.6)

The cost function performances under different channel conditions are depicted in Fig. 6 to Fig. 7. We know that the EDCA mechanism provides the different priorities for differentiate services by using different backoff parameters and AIFS values. Thus, we can adjust the parameters to provide the different priorities with differentiated services to get better cost function.

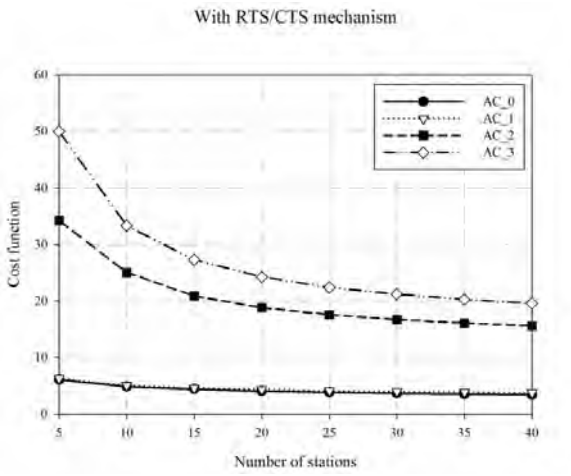



**Fig. 6.** Cost function with RTS/CTS mechanism ($FER_{dir}$ = 0.9)

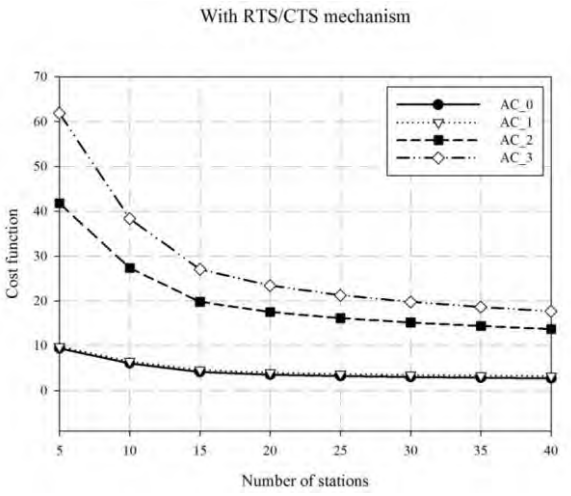



**Fig. 7.** Cost function with RTS/CTS mechanism ($FER_{dir}$ = 0.6).

## 5. Conclusions

In this paper, the Markov chain model with traditional direction and cooperative transmission strategies is proposed to analyze saturation throughput and saturation delay. In general, cooperative communication can reduce the frame error rate; while the rerouting delay due to the additional signal transmitted from the relay to the destination can considerably degrade the system performance. To obtain optimal performance, the cost function is introduced to tradeoff the system performance to determine the strategy for adopting the cooperative transmission.

The theoretical analysis of this paper is very general, and we did not consider the multi-rate transmission for the QoS requirements. We can extend the model to support a multi-rate transmission and derive the numerical analysis in the future.

## References

1. Liao, C.C., Hsu, Y.P. and Feng, K.T.: Performance analysis of cooperative communications from MAC layer perspectives. In Proceedings of the IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications, Tokyo, Japan,1–5, Sep. (2008)
2. Zhu, Y. and Zheng, H.: Understanding the impact of interference on collaborative relays. In IEEE Transactions on Mobile Computing, vol. 7, no. 6, 724–736, Jun. ( 2008)
3. Jakllari, G., Krishnamurthy, S.V., Faloutsos, M., Krishnamurthy, P.V., and Ercetin O.: A cross-layer framework for exploiting virtual MISO links in mobile ad hoc networks. In IEEE Transactions on Mobile Computing, vol. 6, no. 6, 579–594, Jun. (2007)
4. Ding, Z., Ratnarajah, T., and Cowan, C.C.F.: On the diversity-multiplexing tradeoff for wireless cooperative multiple access systems. In IEEE Transactions on Signal Processing, vol. 55, no. 9, 4627–4638, Sep. (2007)
5. Lee, K.D. and Leung, V.C.M.: Evaluations of achievable rate and power consumption in cooperative cellular networks with two classes of nodes. In IEEE Transactions on Vehicular Technology, vol. 57, no. 2, 1166–1175, Mar. (2008)
6. Xiao, Y.: Performance analysis of priority schemes for IEEE 802.11 and IEEE 802.11e wireless LANs. In IEEE Transactions on Wireless Communications, vol. 4, no. 4, pp. 1506−1514, Jul. (2005)
7. Kong, Z.N., Tsang, D. H. K., Bensaou, B. and Gao, D.: Performance analysis of IEEE 802.11e contention-based channel access. In IEEE Journal on Selected Areas in Communications, vol. 22, no. 10, 2095−2106, Dec. (2004)
8. Xiao, Y.: Performance analysis of priority schemes for IEEE 802.11 and IEEE 802.11e wireless LANs. In IEEE Transactions on Wireless Communications, vol. 4, no. 4,1506–1515, Jul. (2005)
9. Jin, J., Wang Q. and Yang, H.: Cross-layer Design of Optimal Contention Period for Mobile WiMAX Systems. In Proceedings of the 4th International Conference on Wireless Communication, Networking and Mobile Computing, Dalian, China, 1–4, Oct. (2008)
10. Chehri, A., Fortier, P., and Tardif, P. M.: Throughput-delay trade-off for slotted aloha multiple access with capture effect. In Journal of Computer Science, vol. 5, no. 9, 630–634, (2009)
11. Xiao, Y.: Performance analysis of priority schemes for IEEE 802.11 and IEEE 802.11e wireless LANs. In IEEE Transactions on Wireless Communications, vol. 4, no. 4, 1506−1514, Jul.( 2005)

**Chien-Erh Weng** received the M.S. degree in Electrical Engineering from the National Yunlin University of Science & Technology, Yunlin, Taiwan, and the Ph.D. degree in electrical engineering from the National Chung Cheng University, Chiayi, R.O.C., in 2000 and 2007, respectively. Since Sep. 2010, he joined the Department of Electronic Communication Engineering at National Kaohsiung Marine University, Kaohsiung, Taiwan, R.O.C., as an Assistant Professor. His research interest is in the field of performance study of UWB communication systems, wireless sensor networks and cooperative radio networks.

**Jyh-Horng Wen** received his Ph.D. degree in Electrical Engineering from the National Taiwan University, Taipei, in 1990. Since February 1991, he has been with the Institute of Electrical Engineering, National Chung Cheng University, Chia-Yi, Taiwan, first as an Associate Professor and, since 2000, as a Professor. He is an Associate Editor of the Journal of the Chinese Grey System Association. Since 2007, Prof. Wen was also the Director of Department of Electrical Engineering for Tung-Hai University. His current research interests include computer communication networks, cellular mobile communications, personal communications, spread-spectrum techniques, wireless broadband systems, and gray theory. Prof. Wen is a member of the IEEE Communication Society, the IEEE Vehicular Technology Society, the IEICE Communication Society, the International Association of Science and Technology for Development, the Chinese Grey System Association, and the Chinese Institute of Electrical Engineering.

**Hsing-Chung Chen** received the Ph.D. degree in Electronic Engineering from National Chung Cheng University, Taiwan, in 2007. During the years 1991-2007, he had served as a Mobile Communication System Engineer at the Department of Mobile Business Group, Chunghwa Telecom Co., Ltd. From Feb. 2008 to Feb. 2013, he was the Assistant Professor of the Department of Computer Science and Information Engineering at Asia University, Taiwan. Since February 2013–present, he is the Associate Professor at the same University. He is also the Research Consultant of Department of Medical Research at China Medical University Hospital, China Medical University Taichung, Taiwan. Currently, he is interested in Information Security, Cryptography, Role-based Access Control, Computer Networks and Wireless Communications. He was Program Co-Chair of numerous conferences. Dr. Chen was the Editor-in-Chief of Newsletter of TWCERT/CC from July 2012 to June 2013.

**Lie Yang** received the B.S. degree in Electronic Engineering from the National Formosa University, Yunlin, Taiwan, in 2012 and he is working towards the master's degree in Electronic Communication Engineering at National Kaohsiung Marine University, Kaohsiung, Taiwan. His current research is wireless sensor networks.

*Received: September 25, 2013; Accepted: January 26, 2014.*

Computer Science and Information Systems 11(3):1157–1172 DOI: 10.2298/CSIS130927070Z

# Weibo Clustering: A New Approach Utilizing Users' Reposting Data in Social Networking Services

Guangzhi Zhang[1], Yunchuan Sun[2], Mengling Xu[1], and Rongfang Bie[1]

[1]College of Information Science and Technology, Beijing Normal University, 100875 Beijing, China
{zgz, xml}@mail.bnu.edu.cn, rfbie@bnu.edu.cn
[2]Business School, Beijing Normal University, 100875 Beijing, China
yunch@bnu.edu.cn

**Abstract.** As one of the most popular Social Networking Services (SNS) in China, Weibo is generating massive contents, relations and users' behavior data. Many challenges exist in how to analyze Weibo data. Most works focus on Weibo clustering and topic classification based on analyzing the text contents only. However, the traditional approaches do not work well because most messages on Weibo are very short Chinese sentences. This paper aims to propose a new approach to cluster the Weibo data by analyzing the users' reposting behavior data besides the text contents. To verify the proposed approach, a data set of users' real behaviors from the actual SNS platform is utilized. Experimental results show that the proposed method works better than previous works which depend on the text analysis only.

**Keywords:** behavior data, clustering, data mining, microblog, Weibo, Social Networking Services.

## 1. Introduction

Social Networking Services (SNS) are changing the world. In the era of Web 1.0, most netizens are just tourists to retrieve information from the Internet. Nowadays, this is not the case. Massive messages are generated by the netizens and massive public highlighting opinions are emerging. The era of "Information Explosion" has been transformed in to that of "Opinion Explosion" with the support of Social Networking Services, such as Microblog, Weibo (a kind of microblog in China), and etc. The content on the Internet, such as the text, image, audio and video, and etc., is the primary resource in the "Information" era. However, for one "opinion", only content is not enough [1, 2]. The social relation (e.g. follow, group, etc.) and users' behaviors (e.g. repost, comment, "@", etc.) play more important roles in forming an "opinion".

Most works on SNS are based on analyzing the text contents for there have been numerous of successful approaches on text mining. Unfortunately, these traditional approaches which are designed to process normative and long enough texts don't work well on SNS platforms because most of the messages are short texts. Even more, there are many kinds of data besides short texts on SNS platforms which can't be processed

with these traditional approaches [3]. Further, many hot messages on SNS even have no text, but only image or video etc.

Social links are more important in forming an opinion on the SNS platform. A schema theory is proposed to help the semantic analysis for the links among objects in [4, 5], which can be utilized in SNS platform. Social relations are more and more frequently used in recent researches and applications [6]. However, there're two problems about social relations. 1) It is hard to discover all social relations among users for its high dynamic changing and sometimes the overall relations are needed in analysis. 2) Social relations are somewhat "static"- it's somewhat "inharmonious" when compared with SNS's highly variable "dynamic". It would be more exciting if new "helper" like relations could be found [7].

In the SNS society, opinions are gradually formed in the dissemination process and every behavior of users contributes to this process. Indeed, we can construct the Web of opinions by extracting opinions from the users' behavior data [8], where opinions can be regarded as events correspondingly. Reposting is a strong opinion expression in SNS (especially in microblog); because it shows that users have a strong wish to recommend the reposted messages to their friends. In other words, one person reposting a message shows his/her strong interest on the topic.

This paper proposes a method to cluster the Weibo messages, utilizing users' interest distribution in different messages which is mined from the reposting data. The experiment results show it performs better than traditional works. The paper is structured as follows. Section 2 introduces the related works and Section 3 introduces the technology background and proposes a new method to cluster the Weibo data. Experiments and analysis are presented in Section 4. Finally, we conclude the paper in Section 5.

## 2. Related Works

Clustering is to organize data into sensible clusters, and is one of the most fundamental modes for understanding and learning a data set. K-means is one of the well-known and simple clustering algorithms proposed 50 years ago. In last decades, some useful research directions, such as semi-supervised clustering, ensemble clustering and so on, have been proposed [9]. K-means++ improves both the accuracy and speed of K-means by choosing the initial seeds, which satisfies users better in some specific fields [10]. In fact, K-means++ is exactly the vital inspiration of our new proposed algorithms.

TF-IDF scheme proposed by Salton and McGill in 1983 [11], is widely used to characterize documents information retrieval systems based on the vector space model. Many classical and modified TF-IDF based approaches were presented for text mining in various fields, such as topic detection and tracking in [12] (proposing a term frequency smoothing method which weaves time slices) and [13] (presenting a multi-document summarizer, which generates summaries using cluster centroids), web pages retrieval [14] (proposing several approaches to refining the TF-IDF by using one page's hyperlinked neighboring pages), image detection [15], and object matching in Google videos [16] and so on. Especially, [17] proposes a perspective of TF-IDF measures for text categorization based on term weighting theories and information theory. There are also lots of researches based on TF-IDF for different purposes, such as introducing

multi-language knowledge integration into social media datasets from Facebook and Twitter for clustering [18, 19] (enriching data representation by employing machine translation to increase the number of features from different languages, but it's useless for Chinese microblogs because of the metaphor and social background), quality-biased ranking for the high-quality contents by a regression approach which incorporates various features [20], content summarization from these collections of posts on a specific topic [21], feature selection for microblog mining [22], real-time topical news recommendation [23], hash-tag retrieval [24] (they all require the relative standard format) and so on.

LDA (Latent Dirichlet Allocation), a generative probabilistic model using TF-IDF for collections of discrete data, is a quite popular model for microblog mining [25]. Reference [26] characterizes microblogs with topic models based on "Labeled LDA", a partially supervised learning model. A modified model called "MB-LDA" is proposed on topic mining in [27], which introduces the "@" and "RT" (Retweet, Repost) into the LDA model to mine the latent relations in the conversations whose test data come from Twitter in English. Short text in microblogs brings big challenges to microblog mining utilizing traditional methods. Reference [28] proposes a method based on hidden topics analysis and text clustering to discover news topics in microblogs. Although the experimental results show this method works well on large-scale microblog dataset, the small length of news in microblog cannot ensure completeness of the whole event.

Some other literatures put forward many creative ways to cluster microblog, including using semantic knowledge [4, 5 and 29] and affinity propagation [30]. Using the results of clustering, many more interesting works have been done to deepen the research on microblog, such as identifying topical authorities [31].

As a typical measurement, TF-IDF earns big success in many fields, including microblog mining. The TF-IDF based K-means algorithms also work quite well in microblog clustering. This paper deploys a clustering framework for microblog clustering based on K-means++， and propose a new RepSim measurement to measure its distance. To test the effectiveness of the proposed method we take the TF-IDF for comparison on the same data set with the same indicators.

## 3. Methods and Design

Microblog data is a kind of typical big data, including contents, relations, and users' behavior records. Considering the features of big data, approaches aiming to do something with the microblog data should be high-efficiency and simple enough (remember the saying "Keep It Simple and Stupid"). In this paper we attempt to find out a measurement for clustering to represent the similarity between two microblogs, which are effective and simple.

After some previous experiments, we find that the users' reposting records data meet our expectation. We here define a new "RepSim" (Reposting Similarity) distance measurement for the similarity computation between Weibos using the users' reposting records data without considering the contents of the Weibo itself, employ K-means++ to cluster Weibo data, while carefully choosing its initial centers, and then we randomly select 100 hot microblogs posted recently from Weibo for the effectiveness test.

Meanwhile, TF-IDF is applied to the same dataset, to compare with the RepSim's results. Three indicators, Cosine, Jaccard and Tanimoto, are used to evaluate the effectiveness of proposed method.

We describe our framework in detail in this part.

### 3.1. Clustering Framework Based on K-means++

The K-means method is a widely used clustering technique that seeks to minimize the average squared distance between two points in the same cluster. Its simplicity and speed are very appealing in practice, but it cannot guarantee general accuracy currently. K-means++ improves both the accuracy and speed of K-means by choosing the initial seeds. We propose that our clustering framework is based on K-means++, choosing the initial seeds according to author's experience with the aim to make sure the results more stable and credible, and is also relatively fair to TF-IDF and RepSim at the same time.

The K-means++ technological process is shown as follows:

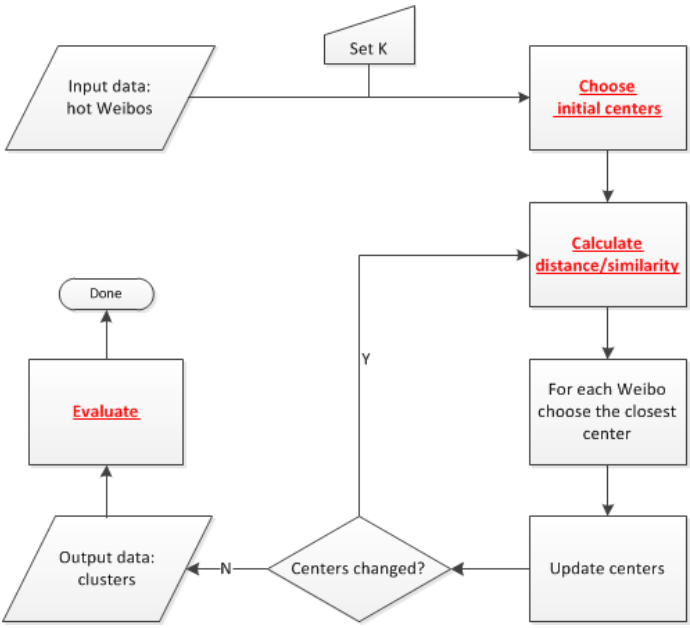



**Fig. 1.** The flow chart of K-means++ used in this paper

The three key points for the algorithm is how to choose the k initial centers, what the similarity or distance definition is, and how to evaluate the clustering effectiveness, which are colored red in Fig. 1.

### 3.2. New Similarity Measurement "RepSim"

New Proposed "RepSim" means "Reposting Similarity", which calculates the degree of similarity between microblog $M_i$ and $M_j$ via the ratio of shared people in all who have reposted the two microblogs. As mentioned above, "reposting" stands for "interest". Meanwhile, one person holds his/her interests stable relatively during a certain period. According to the survey about reposting, it is true that a person is interested in a microblog if he/she reposts it, and two microblogs might have something in common if both of them are reposted by one person. So it has great probability that the two microblogs belong to one cluster when clustering the set of microblogs. That means, the more reposting people $M_i$ and $M_j$ share, the higher probability the two microblogs have the similar topics or characteristics. Hence RepSim can measure the Weibo's similarity from the perspective of probability. We define RepSim as following:

$$\mathrm{RepSim}_{i,j} = \frac{|R_i \cap R_j|}{\sqrt[2]{|R_i| * |R_j|}}. \quad (1)$$

$R_i(R_j)$ is the set of people who repost$M_i$ ($M_j$).We use square root in the denominator so as to process the balance of huge difference between their reposting times.

For example, there aretwo microblog messages reposted by people, $R_1$ = {A, B, C, D, E}, $R_2$ = {C, E, F}, we can calculate RepSim of the two messages by:

$$\mathrm{RepSim}_{1,2} = \frac{|\mathrm{R}_1 \cap \mathrm{R}_2|}{\sqrt[2]{|\mathrm{R}_1| * |\mathrm{R}_2|}} = \frac{|\{\mathrm{C}, \mathrm{E}\}|}{\sqrt[2]{5 \times 3}} = \frac{2}{\sqrt[2]{15}} \approx 0.5164$$

In fact, RepSim performs quite differently between different microblogs. The following scatter-gram shows the distribution on the 100 hot microblogs dataset.

Fig. 2 shows the distribution of RepSim of the 100 hot microblogs dataset, and there are several points to note: 1) RepSim between most microblogs (over 91%) is less than 1%; 2) we divide the 1% into 10 parts with the step of 0.1%, so the distribution through the histogram in the right side is relatively homogeneous; 3) it's hard to get a large RepSim, but RepSim has a clear discrimination for Weibo clustering.

Classical "TF-IDF" is a widely used method to characterize documents information retrieval systems based on the vector space model. TF-IDF is a notable measurement to express the similarity between two microblogs' text (only for text). The TF-IDF formula is:

$$\mathrm{TF}_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}. \quad (2)$$

$n_{i,j}$ is the frequency of the particular word in the document *k*, and the denominator is the total number of words in the document. The greater $\mathrm{TF}_{i,j}$ is, the more significant this word is in the document *k*.

$$\mathrm{IDF}_i = \log \frac{|D|}{|\{d : t_i \in d\}|}. \quad (3)$$

$|D|$ is the total number of documents, and $|\{d : t_i \in d\}|$ is the number of the documents that include the word *t*. That means, the greater $\mathrm{IDF}_i$ is, the more unusual this word is to all documents.

$$\text{TF-IDF}_{i,j} = \text{TF}_{i,j} \times \text{IDF}_i. \quad (4)$$

Now, from the above equation, we can get the conclusion that: the greater $\text{TF-IDF}_{i,j}$ is, the more representative this word is in the document $k$. Therefore, TF-IDF is a notable measurement to express the similarity between two microblogs' text (only for text).

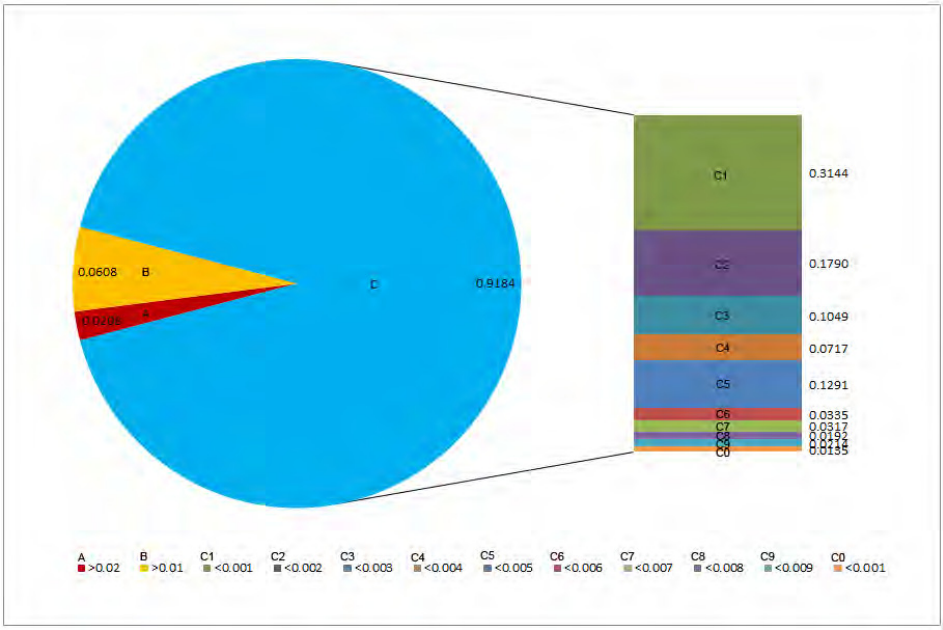



**Fig. 2.** The distribution of RepSim on the 100 hot microblogs dataset

For example, we have a set of text documents and want to find which document is most relevant to the article "Chinese bee breeding". A simple way to start is eliminating documents that do not contain the three words "Chinese", "bee" and "breeding" at the same time. To further distinguish them, we may count the frequency each term occurs in each document, called Term Frequency (TF), and compare them.

However, because the term "Chinese" is so common, which has appeared too many times in the set, this will tend to incorrectly emphasize documents which happen to use the word "Chinese" more frequently without giving enough weight to the more meaningful terms "bee" and "breeding". The term "Chinese" is not a good keyword to distinguish relevant and non-relevant documents and terms when compared with the less common words "bee" and "breeding". Hence a factor, Inverse Document Frequency (IDF),is proposed, which diminishes the weight of terms that occur too frequently in the documents set while increases those that occur rarely.

So we can see that the TF value increases proportionally to the times a word appears in the document, but the value IDF is offset by the frequency of the word in the corpus, which helps to control for the fact that some words are generally more common than

others. And TF-IDF is the product of two statistics, term frequency and inverse document frequency, which presents the contribution of a certain word.

Let's focus on the instance of "Chinese bee breeding". Suppose the article has 1000 words, and words "Chinese", "bee", "breeding" all appear 20 times, so the TFs of these words are 0.02. After that, we find 25 billion web pages, and 6.23 billion web pages contain the word "Chinese", 0.0484 billion web pages contain the word "bee", and 0.0973 billion web pages contain the word "breeding". So TF, IDF and TF-IDF are presented in the following sheet:

**Table 1.** TF, IDF and TF-IDF values of three candidate words

| Words | Web pages(bil) | TF | IDF | TF-IDF |
|---|---|---|---|---|
| Chinese | 6.23 | 0.02 | 0.60 | 0.01 |
| Bee | 0.05 | 0.02 | 2.71 | 0.05 |
| Breeding | 0.10 | 0.02 | 2.41 | 0.09 |

We can see that the TF-IDF value of "bee" is the highest one. So it is obvious that "bee" is the keyword of the article, which is more representable than other two words.

### 3.3. Polymerization Degree for Evaluation

The standards mentioned in this section are based on the training set for evaluating the degree of polymerization within the cluster or between the clusters, with the indicators of Cosine, Jaccard and Tanimoto.

Cosine is a simple and popular indicator for evaluating the similarity between vectors. Training data in this paper is vectors showed in Table 2.

$$\text{Cosine}(x, y) = \frac{x^t \cdot y}{\|x\| \|y\|}. \quad (5)$$

The $x^t$ is the transpose of the vector $x$, and $\|x\|$ is the Euclidean norm of $x$, and it is the same to $\|y\|$.

The Jaccard index, also known as the Jaccard similarity coefficient, is a statistic used for comparing the similarity and diversity of sample sets, which is defined as the size of the intersection divided by the size of the union of the sample sets.

$$d(i, j) = \frac{r+s}{q+r+s}. \quad (6)$$

$$\text{Jaccard}(i, j) = \frac{q}{q+r+s} = 1 - d(i, j) \quad (7)$$

Where q is the number of vector elements which are not zero at the same time, r and s are the number of vector elements when one is zero and the other is nonzero.

Various forms of functions described as Tanimoto Similarity and Tanimoto Distance occur in the literature and on the Internet. Sometimes Tanimoto is called generalized Jaccard. We calculate the Tanimoto with the formula as following, which is mathematically different from the Jaccard.

$$\text{Tanimoto}(x, y) = \frac{x^t \cdot y}{x^t \cdot x + y^t \cdot y - x^t \cdot y}. \quad (8)$$

Where $x^t$ is the transpose of the vector $x$, and the same to $y$.

We evaluate the degree of polymerization via the following two dimensionalities: within the cluster and between the clusters.

The degree of polymerization within the cluster is calculated based on the similarity formulas described above, containing the item's combination within the cluster. At last, a mean value presents the degree of polymerization within the cluster. We define it as following:

$$\text{Polymerization\_Int}(M) = \frac{\sum_{c \in Combinaton\ \ (M)} Sim}{|Combination\ \ \_Int\,(M)|}. \quad (9)$$

For example, for the cluster {a, b, c} generated by the K-means++ algorithm, the similarities between every two elements in the cluster are as follow:

*Sim(a, b)* = 0.5, *Sim(b, c)* = 0.6, *Sim(a, c)* = 0.7.

So the polymerization degree within the cluster is:

$$\begin{aligned} \text{Polymerization}_{\text{Int}\,(M)} &= \frac{\sum_{c \in Combinaton\ \ (M)} Sim}{\left|Combination_{Int\,(M)}\right|} \\ &= \frac{Sim(a,b) + Sim(b,c) + Sim(a,c)}{C_3^2} \\ &= \frac{0.5 + 0.6 + 0.7}{3} \\ &= 0.6 \end{aligned}$$

The degree of polymerization between clusters is quite similar with the degree within the cluster, except that the combination is between different clusters, rather than within the same cluster. We define it as following:

$$\text{Polymerization\_Ext}(M) = \frac{\sum_{c \in Combinaton\ \ (M)} Sim}{|Combination\ \ \_Ext\,(M)|}. \quad (10)$$

"Int" means "within the cluster", while "Ext" means "between the clusters". These two terms will be used later in this paper.

Finally, we define the polymerization of one time's clustering via formula (11), whose results are used as the global evaluation indicator.

$$\text{Polymerization}(M) = \frac{\text{Polymerization}\ \ \_\text{Int}\,(M)}{\text{Polymerization}\ \ \_\text{Ext}\,(M)}. \quad (11)$$

## 4. Experiments and Analysis

The experiments are designed as follows to cluster and evaluate Weibo with two indicators, RepSim and TF-IDF. We use K-means++ algorithm to cluster the set of microblogs, and the "distance" in the K-Means++ algorithm are RepSim and the cosine value of TF-IDF vectors. We calculate the TF-IDF value of all the words appearing in

the set. For each microblog, a vector of TF-IDF value of each word appearing in the microblog is available. After that the cosine distance between any two microblogs could be calculated by the vector we got and then K-means++ algorithm runs with the vector, thus our set of microblogs could be separated into K clusters. As for the RepSim, we calculate the RepSim between every two microblogs as the distances in the K-means++ algorithm. Thus, the set of microblogs can also be divided into K clusters.

After clustering, we evaluate the polymerization degree of these two methods. The training data is the standard data for calculating the polymerization degree, and we analyze the statistics at the end of the experiments. First, the polymerization degree between any two pieces of microblogs is computed with the training data vectors, and three kinds of computing methods are Cosine, Jaccard, and Tanimoto. So with the formula of (9), (10), and (11), the polymerization degrees are available, which is important for us to evaluate the results.

### 4.1. Data Set and Preprocessing

We design a test system to supervise testees to separate microblogs into different classifications in dataset. Seven categories are adopted in our training: Politics, Commerce, Social Focus, ESC (Educational Scientific and Cultural), Sports, Recreation and Health. Testees are well trained and supervised during the whole test process, thus the data training results are credible. We calculate the mean value of all testees' data as our test data:

$$\mathrm{M}_i = \left[m_{i,1}, m_{i,2}, \cdots, m_{i,7}\right]^T$$

$$\mathrm{m}_{i,j} = \frac{\sum_{p \in P} C_{i,j,p}}{|P|}. \tag{12}$$

Where $\mathrm{C}_{\mathrm{i,j,p}}$ is the choice of person $p$,

$$C_{i,j,p} = \begin{cases} 1 & if\ person\ p\ choose\ the\ lable \\ 0 & else \end{cases}. \tag{13}$$

And P is the set of persons who participate in the experiment.

Table 2 shows part of the training results. Since more than one classification options can be selected for a microblog, some sums of the training vector values are greater than 1.

**Table 2.** Examples of training set

| Mid | Politic | Com | Social | ESC | Sport | Recreate | Health |
|---|---|---|---|---|---|---|---|
| 1 | 0.7 | 0.2 | 0.7 | 0 | 0 | 0 | 0 |
| 2 | 0.4 | 1.0 | 0.5 | 0 | 0 | 0 | 0.2 |
| 3 | 0.2 | 0 | 1.0 | 0 | 0 | 0 | 0.1 |
| 4 | 0.2 | 0 | 0.3 | 0.9 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 1.0 | 0.5 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | 1.0 | 0 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0.1 | 1.0 |

The detail about training results and the way we select initial centers will be described in the section “Experiments and Analysis” with the Fig. 4 “The distribution of classification after trained”.

The subsequent similarity computing is based on this training set. We now give more introductions about the data training steps:

Firstly, we capture the hot microblogs from Sina Weibo, via the crawler designed by the author through the Weibo Open Platform APIs (http://open.weibo.com). In fact, we have captured over 40,000,000 high-quality users, more than 100,000,000 reposting records, hot microblogs created everyday over one year, and other data. 100 hot microblogs are selected randomly as our test data set.


**Fig. 3.** The interface of training system

Then, a simple test system is designed to training data, which is like a multiple choice test for testees. We provide good guide to testees for credible results. Fig. 3 gives a screenshot of the training system, where the blue button can submit the classifying results. In this microblog, the text shows poor information for text processing, while the images give people meaningful information.

Finally, statistics about the training results are calculated with formula (12) ~ (13). Some examples are shown in Table 2. Besides, we make a three-dimensional diagram to present both ensemble data and detail of the training data set in Fig. 4.

From the diagram, some information can be found: 1) there are more recreation, social focus and health contents than ESC, commerce, politics and sports; 2) some microblogs have one or more clear classifications, compared to the equivocal; 3) many equivocal microblogs for these seven classifications don’t act well after clustering, mainly because they are extremely confusing on the significance.

That’s the real data from the real SNS site. In general, the 100 top hot microblogs are appropriate to be the test set for Weibo clustering.

Chinese words segmentation is much more difficult than English, especially for the short text. In fact, Chinese short text in SNS (including Weibo) often presents some special features, such as ambiguity and metaphor.

There are several mature and stable open resources for Chinese words segmentation. We refer to these resources and implement a practical program. Specially, artificial detection and modification are made to enhance the accuracy of the TF-IDF based method. The purpose of this operation is to make sure it is more persuasive when compared with our RepSim.

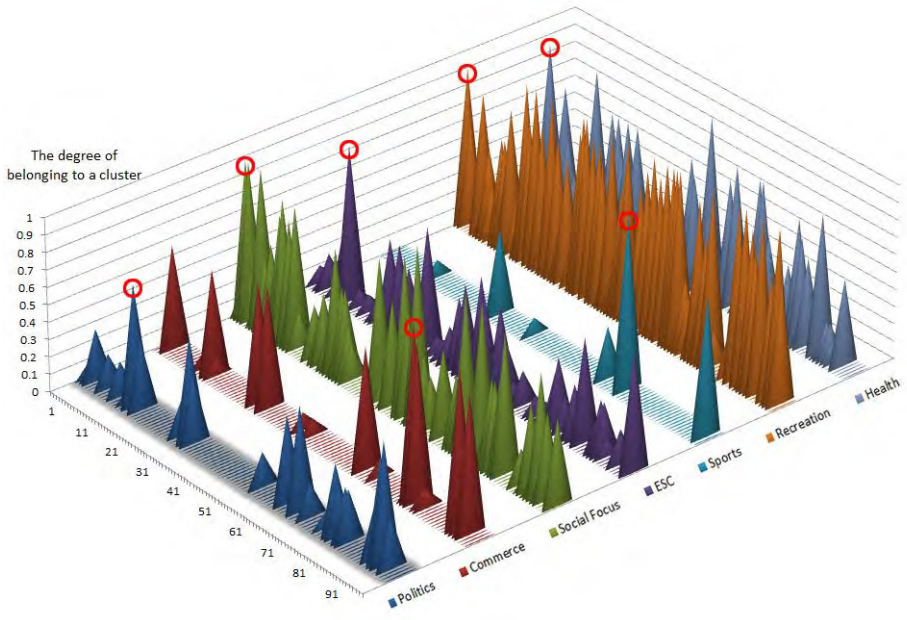



**Fig. 4.** The distribution of classification after trained

## 4.2. Clustering and Evaluation

K-means++ are adopted to do the Weibo clustering. The first thing is to select the initial centers. We select the initial centers artificially according to the distribution of classification after trained shown in Fig. 4. Another important thing is the distance computing, here we use RepSim as described and compare with TF-IDF.

The number of clusters is a skillful and experienced job. In this paper, we assume seven classifications, K=7, which is an "ideal" choice. Besides, we set two more options K=3 and K=10 for comparison.

We run the RepSim/TF-IDF based K-means++ algorithms to do the Weibo clustering on the test set, and get 3 sets of results respectively when K=3, 7 and 10. For each results set, Cosine, Jaccard and Tanimoto are calculated. All computing operations are according to formula (5) ~ (11).

Table 3 shows the results of our experiment in detail, from which we can see that:

1) Cosine, Jaccard and Tanimoto act differently but harmoniously. That means, these three indicators play a role in the evaluation and we can get credible analysis results based on them;

2) The degree of polymerization within the cluster is not always greater than the degree between clusters, which seems not so good. But it's acceptable, because that the TF-IDF and RepSim based K-means++ algorithms are simple and not improved specifically. In addition our purpose in this paper is to show the validity of RepSim based K-means++by comparing it with TF-IDF, so whether the RepSim performs better than TF-IDF is much more important to us;

3) In fact, we can find that no matter via the value of Cosine, Jaccard or Tanimoto, the RepSim is better than TF-IDF stably, no matter K=3, 7 or 10.

**Table 3.** The experiment results. "Int" means the average of indicator in the same cluster's internal; "Ext" means the average of indicator between external clusters

| K | TF-IDF | | | | | | RepSim | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Cosine | | Jaccard | | Tanimoto | | Cosine | | Jaccard | | Tanimoto | |
| | Int | Ext | Int | Ext | Int | Ext | Int | Ext | Int | Ext | Int | Ext |
| 3 | 0.51 | 0.51 | 0.28 | 0.29 | 0.40 | 0.40 | 0.55 | 0.49 | 0.35 | 0.35 | 0.36 | 0.31 |
| 7 | 0.55 | 0.49 | 0.28 | 0.26 | 0.45 | 0.38 | 0.77 | 0.49 | 0.66 | 0.29 | 0.69 | 0.34 |
| 10 | 0.58 | 0.49 | 0.29 | 0.30 | 0.48 | 0.39 | 0.72 | 0.49 | 0.49 | 0.26 | 0.61 | 0.35 |

There is a better perspective to make analysis on the evaluation results. With the use of formula (11), we get 18 polymerization values (2 methods (TF-IDF and RepSim) * 3 measurements (Cosine, Jaccard and Tanimoto) * 3 different Ks (K=3, 7 and 10)) at last.

Fig. 5 shows the 18 values with the form of histogram, and a clear contrast can be seen easily with the help of different colors. Especially, the Y-axis presents the polymerization values.

In Fig. 5, we can see that there are 9 pairs containing 2 close neighbors respectively. Take Jaccard (red histogram, while shallow for TF-IDF and deep for RepSim) for example:

1) When K=3, Jaccard based on RepSim is 0.9993, which is 0.62% better than TF-IDF's 0.9931; when K=7, Jaccard based on RepSim is 2.2604, which is 115.44% better than TF-IDF's 1.0492; when K=10, Jaccard based on RepSim is 1.8767, which is 96.27% better than TF-IDF's 0.9561. From the comparison, we can see clearly that the RepSim's global polymerization is better than TF-IDF, especially when K=7 or 10. The results of Cosine and Tanimoto are similar with Jaccard;

2) Another fact is RepSim's polymerizations when K=7 or 10 is always better than K=3 obviously, while TF-IDF's global polymerization is always just so-so and even becoming worse for Jaccard when K=10. That means, TF-IDF is somewhat powerless in Weibo clustering only based on the text, so it performs generally but also "stably". At the same time, RepSim performs much better, and quite robustly;

3) Overall, the entire polymerization when K=7 and 10 is much better than K=3. Especially, the entire polymerization of RepSim based when K=7 is quite conspicuous. This phenomenon reflects that our test training classifies the microblogs into 7 categories. The RepSim based method agrees with the reality well, because it meets the testees' choice.

In conclusion, Fig. 5 indicates that, the RepSim based method is better than TF-IDF, stably and markedly, and new approach utilizing users' reposting data is effective.

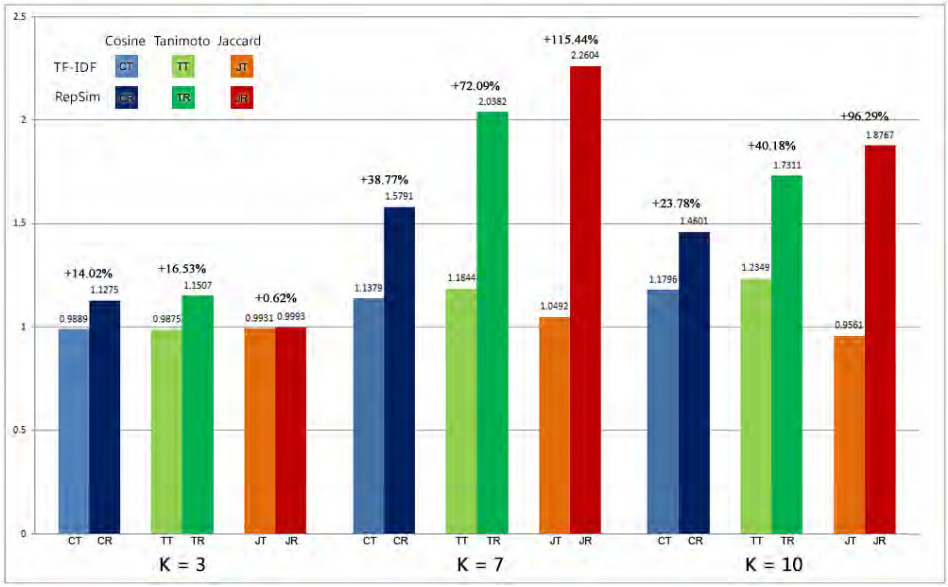



**Fig. 5.** The comparing between TF-IDF and RepSim via Cosine, Jaccard and Tanimoto when K=3, 7, 10

## 5. Conclusion

It is a fact that microblogs on the SNS platform are often very short, and text itself only cannot reflect the real interest of the author and the reposting users, so Weibo clustering based on normal methods are not effective any more. Challenges exist in developing novel approaches for Weibo clustering.

Users' reposting behavior data is a good indicator for discovering users' interests. In this paper a new similarity measurement RepSim is proposed for similarity computing between Weibos by analyzing the behavior data of reposting records. Clustering via RepSim is implemented on the hot microblogs from Sina Weibo so as to find the similar topics.

Experiment results indicate that: 1) RepSim performs well on Weibo clustering, especially comparing with the TF-IDF; 2) RepSim is stable and effective in a variety of conditions, including different evaluating standards and K.

There are several advantages about our work. Firstly, RepSim is simple enough to guarantee the real-time performance. Secondly, RepSim depends on the behaviors of users, but few relevant to the contents of microblogs. Considering two microblogs may be similar if they are reposted by the same user. RepSim is born at this moment. Without the interference of the irrelevant contents of microblog, RepSim works better in experiments.

**Acknowledgments.** This research is sponsored by National Natural Science Foundation of China (61171014, 61272475, 61371185) and the Fundamental Research Funds for the Central Universities（2013NT57）and by SRF for ROCS, SEM. Specially, Libin Jiao and Qin Hu, who are currently undergraduate students in Beijing Normal University, make contribution to this paper in data processing and thesis writing.

## References

1. Juan L, Xueguang Z, Bin C.: Research on Analysis and Monitoring of Internet Public Opinion [C]. In Proceedings of the 2012 International Conference of Modern Computer Science and Applications. Springer Berlin Heidelberg, 449-453. (2013)
2. He Z T, Zhang X Q, Zhao F W, et al.: Internet Public Opinion Monitoring Model Based on Cloud Computing [J]. Applied Mechanics and Materials, 404: 744-747. (2013)
3. Wang H.: Understanding Short Texts [M]. Web Technologies and Applications.Springer Berlin Heidelberg, 1-1. (2013)
4. Zhuge H: Schema Theory for Semantic Link Network [J]. Future Generation Computer Systems, Volume 26, Issue 3, March 2010, Pages 408-420. (2010)
5. Sun Y, Bie R, Yu X, Wang S: Semantic Link Networks: Theory, Applications, and Future Trends [J]，Journal of Internet Technology, Vol. 14 No. 3, P.365-378. (2013)
6. D Wilson, DW Supa: Examining Modern Media Relations: An Exploratory Study of the Effect of Twitter on the Public Relations–Journalist Relationship [J]. Public Relations Journal, Vol. 7, No. 3. (2013)
7. Musiał K, Kazienko P.: Social Networks on the Internet [J]. World Wide Web, 16(1): 31-72. (2013)
8. Sun Y, Yan H, Lu C, Bie R, Zhou Z: Constructing the Web of Events from Raw Data in the Web of Things [J], Mobile Information Systems. Volume 10, No. 1, 2014, pp. 105-125. (2014)
9. Jain A K.: Data Clustering: 50 Years Beyond K-means [J]. Pattern Recognition Letters, 31(8): 651-666. (2010)
10. Arthur D, Vassilvitskii S.: K-means++: The Advantages of Careful Seeding[C]. Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. Society for Industrial and Applied Mathematics, 1027-1035. (2007)
11. Salton G, McGill M J.: Introduction to Modern Information Retrieval [J]. (1983)
12. Radev D R, Jing H, Styś M, et al.: Centroid-based Summarization of Multiple Documents [J]. Information Processing & Management, 40(6): 919-938. (2004)
13. Lee S, Lee J, Park C Y, et al.: Blog Topic Analysis Using TF Smoothing and LDA [C]. In Proceedings of the 7th International Conference on Ubiquitous Information Management and Communication.ACM, 75. (2013)
14. Sugiyama K, Hatano K, Yoshikawa M, et al.: Refinement of TF-IDF Schemes for Web Pages Using Their Hyperlinked Neighboring Pages[C]. In Proceedings of the fourteenth ACM conference on Hypertext and hypermedia. ACM, 198-207. (2003)
15. Chum O, Philbin J, Zisserman A.: Near Duplicate Image Detection: Min-Hash and TF-IDF Weighting[C]. In BMVC. 810: 812-815. (2008)
16. Sivic J, ZissermanA.:Video Google: A Text Retrieval Approach to Object Matching in Videos[C]. Computer Vision, 2003.Proceedings.Ninth IEEE International Conference on. IEEE, 1470-1477. (2003)
17. Aizawa A.: An Information-theoretic Perspective of TF-IDF Measures [J]. Information Processing & Management, 39(1): 45-65. (2003)
18. Tang J, Wang X, Gao H, et al.: Enriching Short Text Representation in Microblog for Clustering [J]. Frontiers of Computer Science, 6(1): 88-101. (2012)

19. Li P, Sun Y, Chen Y, Tian Z: Estimating User Influence In Online Social Networks Subject To Information Overload [J], International Journal of Modern Physics B, Vol. 28, No. 3. (2014)
20. Huang M, Yang Y, Zhu X.: Quality-biased Ranking of Short Texts in Microblogging Services [C]. IJCNLP. 373-382. (2011)
21. Sharifi B, Hutton M A, Kalita J K.: Experiments in Microblog Summarization [C]. In Social Computing (SocialCom), IEEE Second International Conference on. IEEE, 49-56. (2010)
22. Liu Z, Yu W, Chen W, et al.: Short Text Feature Selection for Micro-blog Mining [C]. In Computational Intelligence and Software Engineering (CiSE), 2010 International Conference on. IEEE, 1-4. (2010)
23. Phelan O, McCarthy K, Smyth B.: Using Twitter to Recommend Real-time Topical News[C]. In Proceedings of the Third ACM Conference on Recommender Systems. ACM, 385-388. (2009)
24. Efron M.: Hashtag Retrieval in A Microblogging Environment [C]. In Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 787-788. (2010)
25. Blei D M, Ng A Y, Jordan M I.: Latent Dirichlet Allocation [J]. The Journal of Machine Learning Research, 3: 993-1022. (2003)
26. Ramage D, Dumais S T, Liebling D J.: Characterizing Microblogs with Topic Models [C]. In ICWSM. (2010)
27. Zhang C, Sun J, Ding Y.: Topic Mining for Microblog Based on MB-LDA Model [J]. Journal of Computer Research and Development, 48(10): 1795-1802. (2011)
28. LU Rong, XIANG Liang, LIU Ming-Rong, YANG Qing: Discovering News Topics from Microblogs Based on Hidden Topics Analysis and Text Clustering [J], PR & AI, 25(3): 382-387. (2012)
29. Hu X, Tang L, Liu H.: Enhancing Accessibility of Microblogging Messages Using Semantic Knowledge [C]. In Proceedings of the 20th ACM International Conference on Information and Knowledge Management.ACM, 2465-2468. (2011)
30. Kang J H, Lerman K, Plangprasopchok A.: Analyzing Microblogs with Affinity Propagation [C]. In Proceedings of the First Workshop on Social Media Analytics. ACM, 67-70. (2010)
31. Pal A, Counts S.: Identifying Topical Authorities in Microblogs[C]. In Proceedings of the Fourth ACM International Conference on Web Search and Data Mining. ACM, 45-54. (2011)

**Guangzhi Zhang** is currently a postgraduate student in Beijing Normal University, where he is making effort to obtain a Ph.D. degree. He devotes himself to the research of big data and the internet of things.

**Yunchuan Sun** received his PhD in 2009 from the Institute of Computing Technology, Chinese Academy of Science, Beijing, China. He is currently an associate professor in Beijing Normal University, Beijing, China. He acts as the Secretary of the IEEE Communications Society Technical Subcommittee for the Internet of Things from Jan. 2013. He is also an associate editor of the Springer journal Personal and Ubiquitous Computing. His research interests include Internet of Things, Semantic Link Network, Big Data, Knowledge Representation, Information Security, and Business Models for the Internet of Things. In recent years, he has successfully organized several special issues in some international journals like Springer Personal and Ubiquitous Computing, Elsevier Journal of Networks Computer Applications, etc. He hosts or participates in several research projects from NSFC, 863 Program of China.

**Mengling Xu** obtained her BSc from Beijing Normal University, and is currently a master graduate student at Beijing Normal University. She has published papers in the area of data mining and semantic link.

**Rongfang Bie** received her Ph.D. degree in 1996 from Beijing Normal University, where she is now a professor. She visited the Computer Laboratory at the University of Cambridge in 2003. Her current research interests include Internet of Things, Big Data, knowledge representation and acquisition, computational intelligence and model theory.

*Received: September 27, 2013; Accepted: March 7, 2014.*

Computer Science and Information Systems 11(3):1173–1190 DOI: 10.2298/CSIS130930071S

# An Approach for Selecting Candidates in Soft-handover Procedure Using Multi-Generating Procedure and Second Grey Relational Analysis

Neng-Yih Shih and Hsing-Chung Chen (Jack Chen)

Department of Computer Science and Information Engineering, Asia University
41354 Taichung, Taiwan
{shih, cdma2000}@asia.edu.tw

**Abstract.** The objective of this paper is to develop a decision-making approach for selecting candidates in soft-handover procedure in 3th or 4th generation mobile communication through grey relational analysis of the series similarity and approximation. The multi-generating and second grey relational analysis procedure is applied to select candidates in soft-handover procedure with considerations of the velocity and acceleration similarity of multi-generating data. The validation of computer simulation models illustrate how the approach can be applied in candidates selection in soft-handover, and obtain the best results of feasibility and effectiveness for user equipment (UE) in 3th or 4th generation mobile communications. Moreover, the approach could be easily applied to soft-handover procedure for the mobile communication systems. In this proposed approach is performed to select the candidate target cells by UE instead of eNodeB. It could provide a first solution to choose the candidate target cells through comparing multiple measured data for candidate-selecting with the target communication cell.

**Keywords:** generating procedure, grey relational grade, difference generating, candidates selecting, soft-handover, 3G, 4G.

## 1. Introduction

As the demand for information growing rapidly in recent years, the bandwidth-hungry applications, such as Skype, Google maps, YouTube, Facebook and etc., are gaining more popularity. In order to meet the increasing demands of emerging high-speed mobile data, Internet service providers have launched 4th Generation system such as Long Term Evolution Advanced (LTE-Advanced) over the past years [1]. It is standardized by the 3rd Generation Partnership Project (3GPP) as a major enhancement of the mobile communication standard. It was formally submitted as a candidate 4G system to ITU-T in late 2009. It was approved by the ITU as meeting the requirements of the IMT-Advanced standard, and was finalized by 3GPP in March 2011 [1]. The demand for broadband communication is not only in stationary devices but also in mobile uses. Meanwhile, current innovation in hardware, such as smartphones [2] and tablets, has profoundly changed traditional computing environment and drives the developments of next generation mobile communication. Hence, it can be seen that

people nowadays access the Internet anytime and anywhere through their smartphones, tablets, laptop and other devices.

Mobile communication should enable full accessibility to user equipment (UE) simultaneously and guarantee the Quality of Service (QoS) [3, 4, 5]. Thus, the QoS is a significant measurement tool of Internet quality in mobile users. Usually, the signal strength from serving cell and neighboring cells are viewed as the decision-making factors in handover process. When UE is on process of receiving or sending data, UE is periodically sending measurement report Reference Signal Received Power (RSRP) and Reference Signal Received Quality (RSRQ) from the serving eNodeB and the neighboring eNodeB [4]. If the monitored signal strength of the serving cell drops under the threshold or below the signal level of neighboring eNodeB, the UE have to perform a soft-handover as soon as possible to accommodate the required QoS. There are many research of handover decision are conducted solely based on RSRP [6, 7].

The Grey System Theory is proposed by Deng (2002) [8], and it is mainly applied to measure system models, analyze relations between systems, establish models, predict the system performances and make decisions [9, 10, 11]. The grey relational analysis method is used to compare the geometric relationships of the data series. It is also applied to analyze the causal relationship of the input and output variables, and to identify the major variables and secondary variables from system perspective. In the grey relational analysis method, the relationship between series is developed on the basis of geometric closeness between each data series. However, it is founded that there are some flaws in the traditional grey relational analysis method. Therefore, an approach, the multi-generating and second grey relational analysis, is proposed in this research to improve the reliability of data [12]. First, the approach shows that if the curves of data series are close to each other, then the relationship that the Grey Relational grade shows will be optimal. Furthermore, the approach does not require large amount of data or typical distribution pattern, and only perform small amount of calculations. Hence, it is regarded as an effective means to improve the mathematical statistics. The grey relational analysis, under the requirement of relatively small data sets, is adopted in research to determine the relationship between the grades. This analysis will show how this approach accurately quantifies the similarity and approximation relationship of data series.

There are several Grey Relational Grade models, including the general relational grade presented by Deng [8], the grey relational analysis of B–mode proposed by Wang [13], the slope relational grade developed by Dang [14], the absolute grey relational grade introduced by Liu [15], and the integrated relational grade constructed by Yin [16], etc. These models have indeed achieved a certain effect on the application, but most of the models aim to make improvements based on the proximity between the data series. These models consider the similarity relationship between data. However, these methods increase the complexity of the grey relational model. In this paper, we focus on the geometric similarity and approximation of series, and aim to measure the data series that has similarity in velocity and acceleration. We process not only the original data, but also three sets of data series. Comparing with the applications of traditional Deng's relational grade, our method will do two more times of calculation of grey relational grades in order to enhance comprehensive responses. Finally, the procedure is proved to be successfully applied to candidates' selection in soft-handover procedure. The calculation of simulation results show that this method is more feasible and effective than those in previous researches.

The rest of the paper is organized as following. Related works of the research are introduced in Section 2. The multi-generating and second grey relational analysis approach are illustrated in Section 3. After that, an approach for selecting candidates in soft-handover procedure is proposed in Section 4, followed by the conclusions in section 5.

## 2. Related Works

Mainly, the signal strength from serving cell and neighboring cells in existing 3th generation or 4th generation mobile communication systems are viewed as the decision-making factors for candidate target cells in soft-handover processes. However, the details of how the measurement reported by UE and what kind of related approach could be used to improve the soft-handover processes are very important in this paper. Thus, the related works of measurement report in LTE are described in Section 2.1. In addition, the grey relational grade models are also illustrated in Section 2.2.

### 2.1. Measurement Report in LTE

To simplify the experiment of handover procedure of UE, input measurements are divided into 2 signals, RSRP and RSRQ. The details [9] will be explained in the following subsections.

**Reference Signal Received Power.** Reference Signal Received Power (RSRP) is defined as the linear average over the power contribution of the resource elements that carry cell-specific reference signal within the considered measurement frequency bandwidth. The cell-specific reference signal, according to [17], can be used for RSRP measurement. RSRP can be calculated from the transmit power ($P_s$) of the serving cell, eNodeB, the path loss value ($PL_{us}$) from UE to the serving cell eNodeB, and additional shadow fading with a log-normal distribution. The RSRP can be calculated as following Equation (1).

$$RSRP(x)_m = P_s - PL_{us}(x)_m \quad (1)$$

The reporting range of RSRP is defined from -140 dBm to -44 dBm with 1 dB resolution. The detail of RSRP Measurement report is shown in [18].

**Received Signal Strength Indicator.** E-UTRA Carrier Received Signal Strength Indicator (RSSI) is the total received wideband power observed by the UE from all sources, including co-channel serving and non-serving cell, adjacent channel interference, thermal noise and so on. RSSI can be calculated by Equation (2) as below.

$$RSSI(x)_m = RSRP_{s,ue}(x)_m + RSRP_{int,noice}(x)_m \quad (2)$$

The reporting range for UTRA carrier RSRI is from -100 dBm to -25 dBm. The detail of RSSI Measurement report is shown in [9].

**Reference Signal Received Quality.** Reference Signal Received Quality (RSRQ) can be calculated by the ratio N × RSRP/RSSI, where N is the number of resource block (RB) of the E-UTRA carrier RSSI measurement bandwidth. RSSI includes thermal noise and interference received from the target eNodeB, thus RSRQ, calculated by the following Equation (3). It can show the relation of signal, interference and thermal noise.

$$RSRQ(x)_m = N \times \frac{RSRP(x)_m}{RSSI(x)_m} \tag{3}$$

The reporting range of RSRQ is defined from -19.5 dB to -3 with 0.5 dB resolutions. The detail of RSRQ Measurement report is shown in [18].

## 2.2. Grey Relational Grade Models

Grey relational analysis (GRA) [19-21] is commonly used in Asia. It is a significant evaluation model that scales the level of similarity and difference between the sequences by grey relational grade [19]. The main procedure of GRA is firstly translating the performance of all alternatives into a comparability sequence. This step is called grey relational generating. According to these sequences, a reference sequence (ideal target sequence) is defined. Then, the grey relational coefficient between all comparability sequences and the reference sequence is calculated [20]. GRA is developed based on the point-set topology, it performs an overall comparison between two sets of data rather than a comparison between two points. Thus, it is used to reduce the subjective parameter setting within the model [21].

There are four models are analyzed in this subsection. These models will be compared in this Section.

**Deng's GRG Approach.** Since 2002, Deng's Grey relation grade (GRG, for short) [8] has been widely used to solve problems under small data set in many fields. The specific steps can be summarized as follows:

*Determining the Reference Series and Comparative Series.* At first, the reference series are represented as $X_0 = \{x_0(k) | k = 1,2,\cdots,n\}$, and the comparative series are given as $X_i = \{x_i(k) | k = 1,2,\cdots,n\}$, where $i = 1,2,\cdots,m$.

*Calculate the Relational Coefficient.* The Grey correlation coefficient $\xi_i(k)$ can be expressed as following Equation (4).

$$\xi_i(k) = \frac{\min_i \min_k \Delta_i(k) + \rho \max_i \max_k \Delta_i(k)}{\Delta_i(k) + \rho \max_i \max_k \Delta_i(k)} \tag{4}$$

where $\Delta_i(k) = |x_0(k) - x_i(k)|, \quad i = 1,2,\cdots . k = 1,2 \ldots n$, is representing the absolute value of the difference, $\min_i \min_k \Delta_i(k)$ is the minimum absolute differences, $\max_i \max_k \Delta_i(k)$

is the maximum of absolute differences, and discrimination coefficient $\rho$ which can be changed from 0 to 1, and in this research, we assume that $\rho = 0.5$.

*Calculate the Deng's GRG.* After the grey relational coefficient having been derived, the grey relational grade is calculated by the average value of the grey relational coefficient as the following Equation (5).

$$\gamma(X_0, X_i) = \frac{1}{n}\sum_{k=1}^{n}\xi_i(k) \tag{5}$$

According to Equation (5), there are four axioms of the grey relational grade as below.

1) *Normality*: $0 < \gamma(X_0, X_i) \le 1$, $\gamma(X_0, X_i) = 1 \Leftrightarrow X_0 = X_i$;
2) *Integrity*: *we have* $\gamma(X_i, X_j) \ne \gamma(X_j, X_i)$ $i \ne j$ *for* $X_i, X_j \in X = \{X_s | s = 0,1,2\cdots, m; m \ge 2\}$;
3) Symmetry: *we have* $\gamma(X_i, X_j) = \gamma(X_j, X_i) \Leftrightarrow X = \{X_i, X_j\}$ *for* $X_i, X_j \in X$;
4) *Closely*: if the value of $\Delta_i(k) = |x_0(k) - x_i(k)|$ became smaller, then the value of $\xi_i(k)$ will become greater.

According to Equation (5), Grey relational coefficient $\xi_i(k)$ is function of $\Delta_i(k)$ and discrimination coefficient $\rho$. In this paper, the operator $\rho = 0.5$ is selected. The discrimination coefficient can be automatically selected, please see Ref. [22]. The reference series, the comparative series, the order of non-dimensional data, and the transformation of negative relation and positive relationship data can be pre-processed in the Grey relational generation.

**Grey Slope Similarity Incidence Approach.** In 2010, the grey slope similarity incidence model is proposed by Li-zhi Cui et al. [16]. It is constructed on the basis of grey slope incidence. It is indicated that this model can satisfy similarity and be used to calculate the correlation data of positive or negative. The reference series are represented as $X_0 = \{x_0(k) | k = 1,2,\cdots,n\}$, and the comparative series are given as below.

$$X_i = \{x_i(k) | k = 1,2,\cdots,n\}, \text{ where } i = 1,2,\cdots,m, \text{ and}$$

$$\Delta_i(k) = x_i(k) - x_i(k-1), \quad i = 0,1,2,\cdots m.\ k = 2,3...n$$

$$R_i = \max x_i(k) - \min x_i(k), \quad i = 0,1,2,\cdots m.\ k = 1,2,3...n$$

The grey relational coefficient $\gamma_{0i}(k)$ is given as the following Equation (6):

$$\gamma_{0i}(k) = \operatorname{sgn}(\Delta_0(k)\Delta_i(k))\frac{1}{1 + \left|\frac{|\Delta_i(k)|}{R_i} - \frac{|\Delta_0(k)|}{R_0}\right|} \quad i = 1,2,\cdots m.\ k = 2,3...n \tag{6}$$

where $\operatorname{sgn}(\Delta_0(k)\Delta_i(k)) = \begin{cases} 1 & \text{if } \Delta_0(k)\Delta_i(k) \ge 0 \\ -1 & \text{if } \Delta_0(k)\Delta_i(k) < 0 \end{cases}$.

The GRG $\gamma_{0i}$ is given as the following Equation (7) below.

$$\gamma_{0i} = \frac{1}{n-1}\sum_{k=2}^{n}\gamma_{0i}(k) \tag{7}$$

**Liu's Grey Relational Grade Approach.** In Liu's Grey Relational Grade Approach [23], he assumed that the reference series are represented as $X_0 = \{x_0(k) \mid k = 1,2,\cdots,n\}$, and the comparative series are given as the following Equation (8).

$$X_i = \{x_i(k) \mid k = 1,2,\cdots,n\}, \text{ where } i = 1,2,\cdots,m \text{ , and}$$

$$\Delta_i(k) = x_i(k) - x_i(k-1), \quad i = 0,1,2,\cdots m.\ k = 2,3...n$$

$$\text{Let } |s_i| = \left|\sum_{k=1}^{n}\Delta_i(k)\right| \quad i = 0,1,2,\cdots m. \text{ and}$$

$$|s_{ij}| = \left|\sum_{k=1}^{n} s_i(k) - s_j(k)\right| \quad i, j = 0,1,2,\cdots m.\ i \neq j \text{ , then}$$

$$\gamma_{ij} = \frac{1 + |s_i| + |s_j|}{1 + |s_i| + |s_j| + |s_{ij}|} \tag{8}$$

where $\gamma_{ij}$ is the Grey correlation coefficient of series *i* and *j*.

**The Improved Grey Slope Relational Grade Approach.** The improved Grey Slope Relational Grade (GSRG, for short) approach is proposed by Sun [24], he combined current practical models, and developed his approach with multiple influence factors.

In this model, we assume that the reference series are represented as $X_0 = \{x_0(k) \mid k = 1,2,\cdots,n\}$, and the comparative series are given as $X_i = \{x_i(k) \mid k = 1,2,\cdots,n\}, \text{ where } i = 1,2,\cdots,m$,

and $\Delta_i(k) = x_i(k) - x_i(k-1), \quad i = 0,1,2,\cdots m.\ k = 2,3...n$, then the Equation (9) is presented as below.

$$\gamma_{0i}(k) = \operatorname{sgn}(\Delta_0(k)\Delta_i(k)) \Big/ \left\{1 + \frac{1}{2}\left\|\frac{\Delta_0(k)}{\bar{x}_0}\right| - \left|\frac{\Delta_i(k)}{\bar{x}_i}\right\|\right\} + \frac{1}{2}\left(1 - \min\left(\left|\frac{\Delta_0(k)}{\bar{x}_0}\right|, \left|\frac{\Delta_i(k)}{\bar{x}_i}\right|\right) \Big/ \max\left(\left|\frac{\Delta_0(k)}{\bar{x}_0}\right|, \left|\frac{\Delta_i(k)}{\bar{x}_i}\right|\right)\right) \tag{9}$$

According to the above, the model has achieved a certain effect on the application. Most of the figures reflect the improvements based on the similarity between the data series. However, the procedure of these methods is too complex for data analysis. Therefore, the key factors are how to accurately quantify the geometric relationship of series proximity and similarity and improve the simplicity and practicability of grey relational analysis model.

***MGSRA Approach:*** To generate series that show the similarity of the physical characteristics of speed, acceleration, and the original data series of displacement, three sets of data $X_i^0$, $X_i^1$ and $X_i^2$ are generated as follows:

Step 1: The first set of generating procedure data series: $X_i^0 = \{x_i(k) | k = 1,2,\cdots,n\}$ $i = 0,1,2,\cdots,m$ as original data series. The second set of generating procedure data series as below. $X_i^1 = \{x_i^1(k) = x_i(k+1) - x_i(k) | k = 1,2,\cdots,n-1\}$, $i = 0,1,2,\cdots,m$.

Step 2: The third set of generating procedure data series as below. $X_i^2 = \{x_i^2(k) = x_i^1(k+1) - x_i^1(k) | k = 1,2,\cdots,n-2\}$, $i = 0,1,2,\cdots,m$, where $i = 0$ are reference series for each set, others $i = 1,2,\cdots,m$ are compared series for each set. Each set of series will be pre-processing, respectively.

Step 3: Then, the Deng's GRG $\gamma(X_0^0, X_i^0)$, $\gamma(X_0^1, X_i^1)$, and $\gamma(X_0^2, X_i^2)$ are calculated.

Step 4: Calculate that if $\gamma(X_0^t, X_i^t) = 1$ then $\gamma(X_0^p, X_i^p) = 1$, $t < p \leq 2$. Thus, each original series comparative reference series can generate similar physical nature of displacement, velocity, acceleration, and three grey relational grades. By overall considerations, each grey relational grade whichever is greater for the reference series.

Step 5: Set reference series as $\hat{X}_0 = \{\max_i \gamma(X_0^t, X_i^t) | t = 0,1,2\}$ $i = 1,2,3\cdots m$, and the comparative series are listed as $\hat{X}_i = \{\gamma(X_0^t, X_i^t) | t = 0,1,2\}$ $i = 1,2,3\cdots m$.

Step 6: Do a second grey relational grade to get the order of the series grey relational grade. The order considerate data series of the development trend of similarity and proximity. As a result, more objective and practical, and the calculation method than are simple.

□

## 3. Multi-Generating and Second Grey Relational Analysis Approach

Due to Deng [8] only considered the similarity between the two series in the analysis process, he did not considered the serial oscillation and trend issues. Therefore, in the analysis of oscillation and distribution of complex data, the results are often not consistent with the actual serial (i.e. they simply look for each variable by the correlation calculations result between two adjacent data series). It does not consider the difference between the changing rate of each variable, and the changes of the relation. Meanwhile, it does not take full examination the data series, which overlooks the changing rate of the potential information, as well as the changing difference of information and oscillation data when there will be a large deviation. Thus, many GSRG approaches [12, 14, 23, 24] were proposed in order to improve the Deng's GRG approach [8]. For example, the use of the grey relational analysis of B–mode approach

[13] is performed with the concept of similarity and proximity in analyzing the data series trends. In this paper, the data series focus on the rate of changing of potential information. The multi-generating and second grey relational analysis (*MGSRA approach*, for short) procedure is first proposed in [12], and this paper.

Here, there are three examples calculated by Deng's GRG model [8], B-mode GRG [13], Slope GRG [16], Cui's GRG [16] and *MGSRA Approach*. At first, the Example 1 is given as below.

**Example 1.** A comparsion between the data series obtained from the Ref. [12, 16]. Assume that four data series are listed as below.

$$X_0 = \{1\ 1.3\ 2.5\ 2.8\ 3\ 4.6\ 5\ 6.4\},$$

$$X_1 = \{1\ 1.5\ 2\ 2.5\ 3\ 3.5\ 4\ 4.5\},$$

$$X_2 = \{1.2\ 1.53\ 2.85\ 3.18\ 3.4\ 5.16\ 5.6\ 7.14\},$$

$$X_3 = \{1\ 2.5\ 3\ 1\ 5\ 2\ 7\ 6\},$$

where $X_0$ is reference series, and $X_1$, $X_2$ and $X_3$ are comparative series.

The result of the grey relational grade and data in Ref. [12, 16] are compared with *MGSRA Approach* in Table 1.

**Table 1.** Calculated different GRG

| Models | Deng's Model GRG [8] | B-mode GRG [13] | Slope GRG [16] | Cui's GRG [16] | MGSRA |
|---|---|---|---|---|---|
| $\gamma_{01}$ | 0.739 | 0.387 | 0.894 | 0.910 | 0.643 |
| $\gamma_{02}$ | 0.878 | 0.643 | 0.995 | 1 | 1 |
| $\gamma_{03}$ | 0.568 | 0.126 | 0.109 | 0.053 | 0.369 |
| Results (Order) | $\gamma_{02} \succ \gamma_{01} \succ \gamma_{03}$ | $\gamma_{02} \succ \gamma_{01} \succ \gamma_{03}$ | $\gamma_{02} \succ \gamma_{01} \succ \gamma_{03}$ | $\gamma_{02} \succ \gamma_{01} \succ \gamma_{03}$ | $\gamma_{02} \succ \gamma_{01} \succ \gamma_{03}$ |

Note1: $\gamma(X_0, X_i)$ abbreviated as $\gamma_{0i}$

According to the comparison, these approaches are constructed to meet the behavior of the GRGs $\gamma_{02} \succ \gamma_{01} \succ \gamma_{03}$, and analysis the trends of $X_0$ and $X_2$. The *MGSRA Approach* is proved to be fully reflecting its similarity between these series based on the results of GRG.

**Example 2.** The comparsion of the data series selected from the Ref. [12, 23]. Assume that four data series are listed as below.

$$X_0 = \{3\ 4\ 5\ 6\ 9\},$$
$$X_1 = \{1\ 2.5\ 2.5\ 4\ 7\},$$
$$X_2 = \{1\ 3\ 4\ 4\ 8\},$$
$$X_3 = \{1\ 5\ 4\ 3\ 3.5\},$$

where $X_0$ is reference series; $X_1$, $X_2$ and $X_3$ are comparative series.

The results of calculations of GRG and relation order with Liu's calculation models [23] are compiled with *MGSRA Approach* in Table 2.

**Table 2.** The comparison of Dang, Cao, Liu and our model [12] for GRG calculations

| Models | Dang's GRG (1994) [23] | Cao's GRG [23] | Liu's GRG [15] | MGSRA |
|---|---|---|---|---|
| $\gamma_{01}$ | 0.962 | 1 | 0.9615 | 0.7693 |
| $\gamma_{02}$ | 0.880 | 0.880 | 0.8906 | 0.7435 |
| $\gamma_{03}$ | 0.7642 | 0.9759 | 0.7424 | 0.6615 |
| Results (Order) | $\gamma_{01} \succ \gamma_{02} \succ \gamma_{03}$ | $\gamma_{01} \succ \gamma_{03} \succ \gamma_{02}$ | $\gamma_{01} \succ \gamma_{02} \succ \gamma_{03}$ | $\gamma_{01} \succ \gamma_{02} \succ \gamma_{03}$ |

The computing model is constructed in this paper, and obtains the results $\gamma_{01} \succ \gamma_{02} \succ \gamma_{03}$. It meets the characteristics of the GRG, and shows this model is feasibility and effectiveness.

**Example 3.** A comparsion of the data series chosen from the Ref. [12, 24]. Assume that four data series are listed as below.

$$X_0 = \{1\ 2\ 2.5\ 2.5\ 3\ 5\ 6\},$$
$$X_1 = \{1\ 1.8\ 2.3\ 2.4\ 2.8\ 4.8\ 5.8\},$$
$$X_2 = \{1\ 1.8\ 2.3\ 2.2\ 3\ 4.1\ 5.7\},$$
$$X_3 = \{1\ 2.08\ 2.5\ 2.2\ 3\ 4.3\ 5.8\},$$

where $X_0$ is reference series; $X_1$, $X_2$ and $X_3$ are the comparative series.

The results of calculations, in this paper, the GRG and relation order with Sun's calculation are compiled with *MGSRA Approach* in Table 3.

**Table 3.** The comparison of GRG calculations with Sun and our model

| Models | Dang's GRG(1994) [24] | Dang's GRG(2004)[24] | Sun's GRG[24] | MGSRA |
|---|---|---|---|---|
| $\gamma_{01}$ | 0.9742 | 0.9783 | 0.9867 | 0.8927 |
| $\gamma_{02}$ | 0.9295 | 0.9239 | 0.8040 | 0.4057 |
| $\gamma_{03}$ | 0.9277 | 0.9238 | 0.8133 | 0.5822 |
| Results (Order) | $\gamma_{01} \succ \gamma_{02} \succ \gamma_{03}$ | $\gamma_{01} \succ \gamma_{02} \succ \gamma_{03}$ | $\gamma_{01} \succ \gamma_{03} \succ \gamma_{02}$ | $\gamma_{01} \succ \gamma_{03} \succ \gamma_{02}$ |

In fact, after analyzing the results of the GRG, it can be seen that Dang's GRG cannot reflect the true closeness of the series curves. Therefore, the *MGSRA Approach* is constructed to further compare with Sun's model [24] in order to improve the validity of the results. According to the above, the proposed *MGSRA Approach* is an improved method in comparison with the previous grey relational analysis methods [8, 16, 23, 24]

in candidate selection in soft-handover procedure. The existing grey relational analysis methods can be broadly divided into three categories. One stressed that the absolute displacement difference between the reference series and comparative series. Another emphasizes the reference series and comparative series in terms of the relationship of data change rate. The third one is considering the relation between the aforementioned, discussing both the absolute displacement difference between the series and the series itself, and the related change rate. In particular, the last one is a more comprehensive presentation of the geometric similarity between the data series. However, these methods increase the complexity and uncertainty of the parameters of the model [25]. To develop a procedure series that also examines the similarity in the velocity and acceleration of data series, we measure not only the original data, but also two more sets of series. The applications of traditional Deng's grey relational grade only do two times of computing on grey relational grades for comprehensive response. In this paper, we first propose a new grey relational analysis method, the *MGSRA Approach*, to discuss and analyze the problems like requirements on data size, typical distribution pattern, and small amount of calculations.

Finally, the results of simulation show that this procedure has more feasibility and effectiveness than others. The grey relational computing model gains the advantages compared with the previous studies. It is simpler and retains the Deng's grey relational grade and meets the four axioms.

## 4. Approach for Selecting Candidates in Soft-Handover Procedure Using MGSRA

According to Section 3, *MGSRA Approach* is better than GRG models [8, 13, 16, 23, 24]. Therefore, in this paper, *MGSRA Approach* is choice to design a novel soft-Handover procedure in 3th generation or 4th generation mobile communication systems. In this proposed approach, it could be performed to select the candidate target cells by UE instead of eNodeB. However, it adopts an UE's viewpoint to perform the candidate target-cells selecting procedure through the innovative approach which simultaneously considers the following four collected and measured data in 4G communication environment. Assume that four data are measured by UE, which are including down-link reference signal received power (RSRP, denoted $P$), down-link reference signal received quality (RSRQ, denoted $Q$), down-link received signal strength indicator RSSI (RSSI, denoted $W$), idle channel number (ICN, denoted $I$), which are extracted by simulation of cells traffic in real time, in order to reallocate serving channel of serving eNodeB and candidate cells or channels from neighborhood eNodeBs. Adopting the ICN, it is a novel idea in order to achieve the UE's viewpoint. In this paper, ICNs are assumed that it should be periodically by serving cell and neighborhood cells via some common channel, and be collected by UE. Further, the detail of MGSRA candidate target cell selecting for soft-handover procedure is described in Section 4.1. Finally, the application example of MGSRA candidate-selection procedure is illustrated in Section 4.2.

### 4.1. MGSRA candidate target cell selecting for soft-handover procedure

For the fundamental cellular communication model as shown in the Fig.1, MGSRA candidate target cell selecting for soft-handover procedure in the proposed scheme is defined with four factors *P, Q, W, I*, which are presented as the following Equation (10).

$$Cell\ x = (P, Q, W, I) \quad (10)$$

*where x = {A1, A2, A3, B2, B3, C2, D3}, and P* is the reference signal received power, which is the power level of the received signal in the mobile station and including the signals from the current serving cell and neighbor cells; *Q* is the reference signal received quality; *W* is the received signal strength indicator, which is the total received wideband power observed by the UE from all sources; *I* is real-time idle channel numbers.

The approach for selecting candidate target cells in soft-handover procedure using *MGSRA Approach* (*MGSRA Candidate-selecting Procedure*, for short) in the 4th generation mobile communication is described as below.

| ***MGSRA Candidate-selecting Procedure*** | |
|---|---|
| ***Input:*** | *Cell x= (P, Q, W, I), where x = {1, 2, 3, 4, 5, 6, 7}, where the reference signal received power (RSRP, denoted P), reference signal received quality (RSRQ, denoted Q), received signal strength indicator RSSI (Distance, denoted W), idle channel number (ICN, denoted I).* |
| ***Output:*** | The target cell, and candidate cells |
| *Step 1:* | *Sort the reference signal received powers, where all the results of reference signal received powers are sorted are from high to low. The highest one will be chosen to be the next serving cell called target cell, Cell x', according to the rule of existing mobile communication system.* |
| *Step 2:* | *Perform MGSRA Approach, and generate the result of the grey relational grades of the series comparing with the Cell x', which has the highest received power* |
| *Step 3:* | *Sort the result of series via the comparison of grey relational grades.* |
| *Step 4:* | *Choose the candidate cells from the series, where the numbers of candidate cells are depending on the mobile communication system.* |

### 4.2. MGSRA candidate target cell selecting for soft-handover procedure

There are 7 cells defined by the fundamental cellular communication model for the soft-handover decision-making procedure in the 3th or 4th generation mobile communication. *Cell A1* is the serving cell and *Cell C2* and *B3* are the candidate cells. The number of candidate cells are depending on the mobile communication system. In addition, all raw data P, Q, W, except ICN, will be extracted by UE in existing mobile communications. ICNs are assumed that it should be periodically by serving cell and neighborhood cells via some common channel, and be also collected by UE. For easy discussion, each collected value in data set should be normalized firstly in order to

present as the input for the proposed procedure. In this example, each data is assigned the equal weight for the initial simulation in the first round. They are discussed in the following case studies, Case I and Case II. For each case study, the corresponding data set will be performed by *MGSRA Candidate-selecting Procedure*, individually.

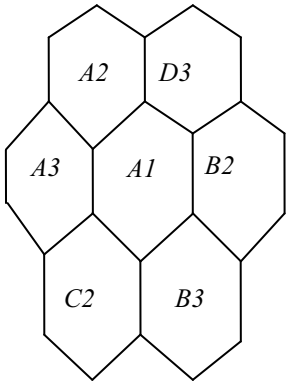



**Fig. 1.** Typical cellular communication model (*Cell A1*: serving cell)

**Case I.** The measured data set collected by a UE for Case I is listed as below.

| | P | Q | W | I | |
|---|---|---|---|---|---|
| Cell A1 = | (6.0 | 5.0 | 7.0 | 4.0) | → the original serving Cell *x* for UE |
| Cell C2 = | (9.0 | 9.0 | 8.0 | 5.0) | |
| Cell B3 = | (9.0 | 9.0 | 8.0 | 4.0) | |
| Cell B2 = | (5.0 | 3.0 | 5.0 | 8.0) | |
| Cell A3 = | (2.0 | 1.0 | 7.0 | 1.0) | |
| Cell D3 = | (1.0 | 1.0 | 4.0 | 3.0) | |
| Cell A2 = | (1.0 | 2.0 | 1.0 | 7.0) | |

To perform *MGSRA Candidate-selecting Procedure***,** the measured data of serving cell and neighborhood cells are inputted as below.

P Q W I
Cell *A1* = (6.0, 5.0, 7.0, 4.0)
Cell *C2* = (9.0, 9.0, 8.0, 5.0)
Cell *B3* = (9.0, 9.0, 8.0, 4.0)
Cell *B2* = (5.0, 3.0, 5.0, 8.0)
Cell *A3* = (2.0, 1.0, 7.0, 1.0)
Cell *D3* = (1.0, 1.0, 4.0, 3.0)
Cell *A2* = (1.0, 2.0, 1.0, 7.0)

Step CI-1: Sort the inputs according to the reference signal received powers. All the reference signal received power are sorted from high to low. Thus, the *Cell C2* is chosen to be the target cell by UE, and *Cell B3*, *A1*, *B2* in descending order for the candidate target cells according to the rule of existing mobile communication system. The sorted results are listed as below.

P Q W I
Cell *C2* = (9.0, 9.0, 8.0, 5.0) → the target *Cell x'* for UE
Cell *B3* = (9.0, 9.0, 8.0, 4.0) → the first candidate target cell for UE
Cell *A1* = (6.0, 5.0, 7.0, 4.0) → the second candidate target cell for UE
Cell *B2* = (5.0, 3.0, 5.0, 8.0) → the third candidate target cell for UE
Cell *A3* = (2.0, 1.0, 7.0, 1.0)
Cell *D3* = (1.0, 1.0, 4.0, 3.0)
Cell *A2* = (1.0, 2.0, 1.0, 7.0)

Step CI-2: After Step CI-2, *MGSRA Approach* is performed**.** The results of the series of grey relational grades comparing with *Cell C2*, the highest received power, are listed as below.

$$\gamma(C2, B3) = 1;$$
$$\gamma(C2, A1) = 0.6589;$$
$$\gamma(C2, B2) = 0.6684;$$
$$\gamma(C2, A3) = 0.4607;$$
$$\gamma(C2, D3) = 0.5122;$$
$$\gamma(C2, A2) = 0.4588.$$

Step CI-3: After calculating the *MGSRA Approach*, the results are listed in Table 4.

**Table 4.** The results of the series of grey relational grades comparing with *Cell C2,* the highest received power

| Grey relational grade | $\gamma(C2, B3)$ | $\gamma(C2, A1)$ | $\gamma(C2, B2)$ | $\gamma(C2, A3)$ | $\gamma(C2, D3)$ | $\gamma(C2, A2)$ |
|---|---|---|---|---|---|---|
| Results | 1.0000 | 0.6589 | 0.6684 | 0.4607 | 0.5122 | 0.4588 |

Step CI-4: The results of candidate series are *Cell B3*, *Cell B2*, and *Cell A1* in descending order. Thus, UE can choose the *Cell B3* and *Cell B2* to be the first candidate cell and the second candidate cell, individually, from Table 5, where the number of candidate cells is chosen based on the mobile communication system. The results in Table 5 are sorted and listed in descending order as below.

P Q W I

Cell *C2*＝（9.0，9.0，8.0，5.0）→ the target *Cell x'* for UE

Cell *B3*＝（9.0，9.0，8.0，4.0）→ the first candidate cell for UE

Cell *B2*＝（5.0，3.0，5.0，8.0）→ the second candidate cell for UE

Cell *A1*＝（6.0，5.0，7.0，4.0）→ the third candidate cell for UE

Cell *D3*＝（1.0，1.0，4.0，3.0）

Cell *A3*＝（2.0，1.0，7.0，1.0）

Cell *A2*＝（1.0，2.0，1.0，7.0）

**Table 5.** The sorted results of the series of grey relational grades

| Grey relational grade | $\gamma(C2, B3)$ | $\gamma(C2, B2)$ | $\gamma(C2, A1)$ | $\gamma(C2, D3)$ | $\gamma(C2, A3)$ | $\gamma(C2, A2)$ |
|---|---|---|---|---|---|---|
| Results | 1.0000 | 0.6684 | 0.6589 | 0.5122 | 0.4607 | 0.4588 |

**Case II.** The measured data set collected by a UE for Case II is listed as below.

P Q W I

Cell *A1*＝（6.0，5.0，7.0，4.0）→ the original serving Cell *x* for UE

Cell *C2*＝（9.0，9.0，8.0，5.0）

Cell *B3*＝（9.0，9.0，8.0，4.0）

Cell *B2*＝（9.0，9.0，5.0，8.0）

Cell *A3*＝（2.0，1.0，7.0，1.0）

Cell *D3*＝（1.0，1.0，4.0，3.0）

Cell *A2*＝（1.0，2.0，1.0，7.0）

To perform *MGSRA Candidate-selecting Procedure*, the measured data of serving cell and neighborhood cells are inputted as below.

P Q W I

Cell *A1*＝（6.0，5.0，7.0，4.0）

Cell *C2*＝（9.0，9.0，8.0，5.0）

Cell *B3*＝（9.0，9.0，8.0，4.0）

Cell *B2*＝（9.0，9.0，5.0，8.0）

Cell *A3*＝（2.0，1.0，7.0，1.0）

Cell *D3*＝（1.0，1.0，4.0，3.0）

Cell *A2*＝（1.0，2.0，1.0，7.0）

Step CII-1: Sort the inputs according to the reference signal received powers. All the reference signal received power are sorted from high to low. Thus, the *Cell C2* is chosen to be the target cell by UE, and *Cell B3, Cell A1,* and *Cell B2* in descending order for the candidate target cells according to the rule of existing mobile communication system. The sorted results are listed as below.

P Q W I

Cell *C2*＝（9.0，9.0，8.0，5.0）→ the target *Cell x'* for UE
Cell *B3*＝（9.0，9.0，8.0，4.0）→ the first candidate target cell for UE
Cell *B2*＝（9.0，9.0，5.0，8.0）→ the second candidate target cell for UE
Cell *A1*＝（6.0，5.0，7.0，4.0）→ the third candidate target cell for UE
Cell *A3*＝（2.0，1.0，7.0，1.0）
Cell *D3*＝（1.0，1.0，4.0，3.0）
Cell *A2*＝（1.0，2.0，1.0，7.0）

Step CII-2: After Step CII-1, *MGSRA Approach* is performed**.** The results of the series of grey relational grades comparing with *Cell C2*, the highest received power, are listed as below.

$$\gamma(C2, B3) = 1;$$
$$\gamma(C2, B2) = 0.4908;$$
$$\gamma(C2, A1) = 0.6133;$$
$$\gamma(C2, A3) = 0.4026;$$
$$\gamma(C2, D3) = 0.4578;$$
$$\gamma(C2, A2) = 0.3984.$$

Step CII-3: After calculating the *MGSRA Approach*, the results are listed in Table 6.

**Table 6.** The result of series of grey relational grades comparing to the cell which has the highest received power

| Grey relational grade | $\gamma(C2, B3)$ | $\gamma(C2, B2)$ | $\gamma(C2, A1)$ | $\gamma(C2, A3)$ | $\gamma(C2, D3)$ | $\gamma(C2, A2)$ |
|---|---|---|---|---|---|---|
| Results | 1.0000 | 0.4908 | 0.6133 | 0.4026 | 0.4578 | 0.3984 |

Step CII-4: The results of candidate series are *Cell B3*, *Cell A1*, and *Cell B2* in descending order. Thus, UE can choose the *Cell B3* to be also the first candidate cell and *Cell A1* to be the second candidate cell from Table 7, where the number of candidate cells is chosen based on the mobile communication system. The results in Table 7 are sorted and listed in descending order as below.

P Q W I

Cell *C2* = (9.0, 9.0, 8.0, 5.0) → the target *Cell x'* for UE

Cell *B3* = (9.0, 9.0, 8.0, 4.0) → the first candidate cell for UE

Cell *A1* = (6.0, 5.0, 7.0, 4.0) → the second candidate cell for UE

Cell *B2* = (9.0, 9.0, 5.0, 8.0) → the third candidate cell for UE

Cell *D3* = (1.0, 1.0, 4.0, 3.0)

Cell *A3* = (2.0, 1.0, 7.0, 1.0)

Cell *A2* = (1.0, 2.0, 1.0, 7.0)

**Table 7.** The sorted results of the series of grey relational grades

| Grey relational grade | $\gamma(C2, B3)$ | $\gamma(C2, A1)$ | $\gamma(C2, B2)$ | $\gamma(C2, D3)$ | $\gamma(C2, A3)$ | $\gamma(C2, A2)$ |
|---|---|---|---|---|---|---|
| Results | 1.0000 | 0.6133 | 0.4908 | 0.4578 | 0.4026 | 0.3984 |

Finally, according to *the* above Case I and Case II, after dealing with them, MGSRA candidate-selection procedure could get the proof that it could accurately and fast find the candidates out through distinct overall measured data.

## 5. Conclusions

The MGSRA Candidate-selecting Procedure could find easily the candidates out, through considering multiple measured factors on the same time instead of the traditional approach only satisfying one of the conditions in RSRP, RSRQ, and RSSI. ICNs are assumed that it should be periodically by serving cell and neighborhood cells via some common channel, and be collected by UE. It is a very special factor that is a novel idea in order to achieve the UE's viewpoint. The proposed procedure could be performed to select the candidate target cells by UE instead of eNodeB. Therefore, it is the first proposed approach could choose the candidate target cells through comparing multiple measured data for candidate-selecting with the target cell. Moreover, it could be easily applied to soft-handover procedure for 3th generation or 4th generation mobile communication systems. In conclusion, the *MGSRA Approach* is proved to be an effective means to apply in selecting candidates in soft-handover procedure for 3th generation or 4th generation mobile communications. With several simulations are validated, the approach can be used to select the candidates in soft-handover procedure, and obtain the best results of feasibility and effectiveness for UE in 4th generation mobile communications.

**Acknowledgements.** This work was supported in part by the National Science Council, Taiwan, Republic of China, under Grant NSC 102-2221-E-468-007, also by Asia University, Taiwan, under Grant 101-asia-28.

## References


1. Dahlman E., Parkvall S., Sköld J., Beming P.: 3G Evolution: HSPA and LTE for Mobile Broadband. (2007)
2. Jeong H.D.J., Lim J., Hyun W., An A.: A Real-time Location-based SNS Smartphone Application for the Disabled Population. Computer Science and Information Systems, Vol. 10, No. 2, 747–765. (2013)
3. Hofestadt, H.: GSM-R: Global System for Mobile Radio Communications for Railways. In Processding of International Conference on Electric Railways in a United Europe, Amsterdam, Netherlands, 111–115. (1995)
4. Barbu G.: E-TRAIN–Broadband Communication with Moving Trains. International Union of Railway. (2008)
5. European Telecommunications Standards Institute (ETSI): Requirements for support of radio resource management. 3GPP TS 36.133 V10.5.0. (2012)
6. Anas M., Calabrese F.D., Mogensen P.E., Rosa C., Pedersen K. I.: Performance Evaluation of Received Signal Strength Based Hard Handover for UTRAN LTE. Department of Electronic Systems, Aalborg University, Nokia Networks. (2007)
7. Wikipedia: History_of_mobile_phones. [Online]. http://en.wikipedia.org/wiki/History_of_mobile_phones. (2012)
8. Deng J.: Method of Grey Theory. In Processding of Wuhan Huazhong of University Science and Technology, 37–138. (2002)
9. Chen H.C., Cahyadi Y., Marsha A.V.: A Grey Prediction Based Hard Handoff Hysteresis Algorithm for 3GPP LTE System. In Processding of International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA 2012), University of Victoria, Victoria, Canada, 590–595. (2012)
10. Chen H.C., Cahyadi Y., Deviani R., Liu T.W., Wen J.H.: Using GM (1, 1) Model to Forecast the Trend of Research in Internet of Things. In Processding of International Conference on Automatic Control and Artificial Intelligence (ACAI 2012), The Institution of Engineering and Technology (IET), Xiamen, China, 2247 – 2250. (2012)
11. Chen H.C., Deviani R., Shih N.Y., Liu C.C., Cahyadi Y.: Application of GM (1, 1) Model for Forecasting Research Trends of Security in Internet of Things. In Processding of the 5th IET International Conference on Ubi-media Computing (U-Media 2012), The Institution of Engineering and Technology (IET), 47-51, Xining, China. (2012)
12. Shih N.Y., Chen H.C.: Multi-Generating Procedure and Second Grey Relational Analysis. In Processding of 2013 Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS 2013), 601-604, Asia University, Taiwan, July 3-5. (2013)
13. Wang Q.Y.: The Grey Relational Analysis of B-Mode. The Journal of Huazhong University of Science and Technology, Vol. 17, 77–82. (1989)
14. Dang Y., Liu S., Liu B., Mi C.: Improvement on Grade of Grey Slope Incidence. Engineering Science, Vol.6, No.3, p23– 26. (2004)
15. Liu S., Guo T. B., Dang Y.: Grey System Theory and Application. Science Publisher: Beijing, 46–63. (2010)
16. Cui L.Z., Liu S.F., Li Z.P., Cui J.: Study on Grey Slope Similarity Incidence and Its Applications. Statistics & Information Forum, Vol. 25, No. 3. (2010)
17. Rezaei et.al.: Grey Prediction Based Handoff Algorithm. World Academy of Science, Engineering and Technology. (2005)
18. Dahlman E., Parkvall S., Sköld J., Beming P.: 3G Evolution: HSPA and LTE for Mobile Broadband. (2007)
19. Deng J.L.: Introduction of grey system. Journal of Grey System. Vol. 1, No. 1, 1–24. (1989)
20. Kuo Y.Y., Yang T., Huang G.W.: The Use of Grey Relational Analysis in Solving Multiple Attribute Decision-Making Problems. Computers & Industrial Engineering, Vol. 55, No. 1, 80–93. (2008)

21. Chan J.W.K., Tong T.K.L.: Multi-Criteria Material Selections And End-Of-Life Product Strategy: Grey Relational Analysis Approach. Materials & Design, Vol. 28, No. 5, 1539–1546. (2007)
22. Shih N.Y., Liu H.C.: Intelligent Estimation of Distinguishing Coefficient in Grey Relational Grade. In Proceedings of Cross-Strait Conference on Information Science and Technology, 88–94. (2008)
23. Liu W., He X.: A New Grey Relational Grade Mode. Statistics and Decision, Vol.14, 160–161. (2011)
24. Sun Y.G., Dang Y.G.: Grey Slope Relational Grade Improved Mode. L. Ilu Nxintan, Vol.8, 12–13. (2007)
25. Yin H.Z., Liu B., Zhang H., Zhang R.: A New Computation Model of Incidence Grade Met Four Axioms of Grey Incidence. Henan Science, Vol. 24, No. 2, 162–165. (2006)

**Neng-Yih Shih** was born in Changhua, Taiwan, in 1959. He received his B.S and M.S. degrees in the Department of Automatic Control Engineering, from Feng Chia University, Taiwan, in 1982 and 1984. He received a Ph.D. degree in Institute of Aeronautics and Astronautics from the Nation Cheng Kung University, Taiwan, in 2001. He is currently an associate professor with the Department of Computer Science and Information Engineering, Asia University, Taiwan. He is also Secretary General of Asia University, a position he has held since Aug. 2011. His main research interests include the application of expert control, networked control system and intelligent systems.

**Hsing-Chung Chen (Jack Chen)** is the corresponding author of this paper. He received the Ph.D. degree in Electronic Engineering from National Chung Cheng University, Taiwan, in 2007. During the years 1991-2007, he had served as a Mobile Communication System Engineer at the Department of Mobile Business Group, Chunghwa Telecom Co., Ltd. From Feb. 2008 to Feb. 2013, he was the Assistant Professor of the Department of Computer Science and Information Engineering at Asia University, Taiwan. Since February 2013–present, he is the Associate Professor at the same University. He is also the Research Consultant of Department of Medical Research at China Medical University Hospital, China Medical University Taichung, Taiwan. Currently, he is interested in Information Security, Cryptography, Role-based Access Control, Computer Networks and Wireless Communications. He was Program Co-Chair of numerous conferences. Dr. Chen was the Editor-in-Chief of Newsletter of TWCERT/CC from July 2012 to June 2013.

*Received: September 30, 2013; Accepted: February 28, 2014.*

CIP – Каталогизација у публикацији
Народна библиотека Србије, Београд

004

COMPUTER Science and Information Systems : the International journal / Editor-in-Chief Mirjana Ivanović. – Vol. 11, No 3 (2014) - . – Novi Sad (Trg D. Obradovića 3): ComSIS Consortium, 2014 - (Belgrade : Sigra star). –30 cm

Polugodišnje. – Tekst na engleskom jeziku

ISSN 1820-0214 = Computer Science and Information Systems
COBISS.SR-ID 112261644

Cover design: V. Štavljanin
Printed by: Sigra star, Belgrade

ComSIS Computer Science and Information Systems
Volume 11, Number 3, Special Issue, August 2014

# Contents

Editorial

## Papers



ISSN: 1820-0214