Computer Science & Information Technology 62

Natarajan Meghanathan
David C. Wyld (Eds)

Computer Science & Information Technology

Eighth International Conference on Networks & Communications
(NETCOM - 2016)
Sydney, Australia, December 23~24, 2016

AIRCC Publishing Corporation

Volume Editors

Natarajan Meghanathan,
Jackson State University, USA
E-mail: nmeghanathan@jsums.edu

David C. Wyld,
Southeastern Louisiana University, USA
E-mail: David.Wyld@selu.edu

ISSN: 2231 - 5403
ISBN: 978-1-921987-60-1
DOI : 10.5121/csit.2016.61501 - 10.5121/csit.2016.61518

Typesetting: Camera-ready by author, data conversion by NnN Net Solutions Private Ltd., Chennai, India

Preface

The Eighth International Conference on Networks & Communications (NETCOM 2016) was held in Sydney, Australia, during December 23~24, 2016. The Eighth International Conference on Network and Communications Security (NCS 2016), The Eighth International Conference on Wireless & Mobile Networks (WiMoNe 2016), The Eighth International Conference on Applications of Graph Theory in Wireless Ad hoc Networks and Sensor Networks (GRAPH-HOC 2016), The Third International Conference on Signal, Image Processing and Multimedia (SPM 2016) and The Third International Conference on Computer Science, Engineering and Information Technology (CSEIT 2016) was collocated with the NETCOM 2016. The conferences attracted many local and international delegates, presenting a balanced mixture of intellect from the East and from the West.

The goal of this conference series is to bring together researchers and practitioners from academia and industry to focus on understanding computer science and information technology and to establish new collaborations in these areas. Authors are invited to contribute to the conference by submitting articles that illustrate research results, projects, survey work and industrial experiences describing significant advances in all areas of computer science and information technology.

The NETCOM-2016, NCS-2016, WiMoNe-2016, GRAPHHOC-2016, SPM-2016, CSEIT-2016 Committees rigorously invited submissions for many months from researchers, scientists, engineers, students and practitioners related to the relevant themes and tracks of the workshop. This effort guaranteed submissions from an unparalleled number of internationally recognized top-level researchers. All the submissions underwent a strenuous peer review process which comprised expert reviewers. These reviewers were selected from a talented pool of Technical Committee members and external reviewers on the basis of their expertise. The papers were then reviewed based on their contributions, technical content, originality and clarity. The entire process, which includes the submission, review and acceptance processes, was done electronically. All these efforts undertaken by the Organizing and Technical Committees led to an exciting, rich and a high quality technical conference program, which featured high-impact presentations for all attendees to enjoy, appreciate and expand their expertise in the latest developments in computer network and communications research.

In closing, NETCOM-2016, NCS-2016, WiMoNe-2016, GRAPHHOC-2016, SPM-2016, CSEIT-2016 brought together researchers, scientists, engineers, students and practitioners to exchange and share their experiences, new ideas and research results in all aspects of the main workshop themes and tracks, and to discuss the practical challenges encountered and the solutions adopted. The book is organized as a collection of papers from the NETCOM-2016, NCS-2016, WiMoNe-2016, GRAPHHOC-2016, SPM-2016, CSEIT-2016.

We would like to thank the General and Program Chairs, organization staff, the members of the Technical Program Committees and external reviewers for their excellent and tireless work. We sincerely wish that all attendees benefited scientifically from the conference and wish them every success in their research. It is the humble wish of the conference organizers that the professional dialogue among the researchers, scientists, engineers, students and educators continues beyond the event and that the friendships and collaborations forged will linger and prosper for many years to come.

Natarajan Meghanathan
David C. Wyld

Organization

General Chair

Natarajan Meghanathan,	Jackson State University,USA
Brajesh Kumar Kaushik,	Indian Institute of Technology - Roorkee, India

Publicity Chair

Jae Kwang Lee,	Hannam University, South Korea

Program Committee Members

Abdul Kadir,	Technical University of Malaysia Malacca, Malaysia
Aloizio,	Aeronautic Institute of Technology, Brasil
Ambresh,	Mangalore University, India
Amol D Mali,	University of Wisconsin, USA
Anamika Ahirwar,	Rajiv Gandhi Technical University, India
Ankit Chaudhary,	Truman State University, USA
Apai,	Universiti Malaysia Perlis, Malaysia
Asmaa Shaker Ashoor,	Babylon University, Iraq
Avadhani P.S,	Andhra University, India
B Srinivasan,	Monash University, Australia
Balasubramanian K,	Lefke European University, Cyprus
Chin-Chih Chang,	Chung Hua University,Taiwan
Dhinaharan Nagamalai,	Wireilla Net Solutions, Australia
Doina Bein,	The Pennsylvania State University,USA
Dong Seong Kim,	Duke University, USA
Doreswamy,	Mangalore University, India
Emilio Jiménez Macías,	University of La Rioja, Spain
Erritali Mohammed,	Sultan Moulay Slimane University, Morocco
Fahimeh Farahnakian,	University of Turku, Finland
Farhat Anwar,	International Islamic University, Malaysia
Ford Lumban Gaol,	University of Indonesia, Indonesia
Franz Ko,	Dongkuk University, South Korea
Genge Bela,	Petru Maior University of Tirgu Mures, Romania
Hariharan S,	J.J.College of Engineering, India
Hossein Jadidoleslamy,	University of Zabol, Iran
Hossein Jadidoleslamy,	University of Zabol, Zabol, Iran
Houcine Hassan,	Univeridad Politecnica de Valencia, Spain
Isa Maleki,	Islamic Azad University, Iran
Islam Atef,	Alexandria University, Egypt
Jae Kwang Lee,	Hannam University, South Korea
Jaime Galán Jiménez,	University of Extremadura, Spain
Jan Zizka,	Mendel University in Brno, Czech Republic
Juan A. Fraire	Universidad Nacional de Córdoba, Argentina
Kannan A,	K.L.N.College of Engineering, India

Kayhan Erciyes,	Izmir University, Turkey
Li Zheng,	University of Bridgeport, USA
Lorena González Manzano,	University Carlos III of Madrid, Spain
Mahdi Mazinani,	IAU Shahreqods, Iran
Mahi Lohi,	University of Westminster, UK
Mohamed Ashik M,	Salalah College of Technology, Oman
Mohamed Fahad AlAjmi,	King Saud University, Saudi Arabia
Mohammad M.Banat,	Jordan University of Science and Technology, Jordan
Mohammed Ghanbari,	University of Essex, United Kingdom
Moses Ekpenyong,	University of Edinburgh, Nigeria
Mujiono Sadikin,	Universitas Mercu Buana, Indonesia
Nabila Labraoui,	University of Tlemcen, Algeria
Nadia Qadri,	University of Essex, United Kingdom
Nagaraju A,	Central University of Rajasthan, India
Nazmus Saquib,	University of Manitoba, Canada
Neda Darvish,	Islamic Azad University, Iran
Nicolas Anciaux,	Researcher at INRIA Paris-Rocquencourt, France
Nikunj Domadiya,	B. H. Gardi College of Engg. and Tech, India
Paramartha Dutta,	Visva Bharati University, India
Pavan Kumar K,	PVP Siddhartha Institute of Technology, India
Pooja jain,	IIIT Kota, India
Rahul Gupta,	Fractal Analytics, India
Rahul Moriwal,	Acropolis Institute of Technology & Research, India
Ruchi Tuli,	Yanbu University College, Kingdom of Saudi Arabia
Saad M.Darwish,	Alexandria University, Egypt
Samarendra Nath Sur,	Sikkim Manipal Institute of Technology, India
Sandhya Magesh,	B.S.Abdur Rahman University, India
Sattar B. Sadkhan,	IT College- University of Babylon, IRAQ
Satyanarayana V Nandury,	Indian Institute of Chemical Technology, India
Selwyn Piramuthu,	University of Florida, Florida
Selwyn Piramuthu,	University of Florida, United States
Sergio Pastrana,	University Carlos III of Madrid, Spain
Seyyed AmirReza Abedini,	Islamic Azad University, Iran
Seyyed Reza Khaze,	Islamic Azad University, Iran
Shahid Siddiqui,	Integral University, India
Shrirang Kulkarni,	K.L.S Gogte Institute of Technology, India
Thatiparti venkata Rajini Kanth,	Sreenidhi Institute of Science and Technology, India
Thuc-Nguyen,	University of Science, Vietnam
Zsolt Polgar,	Technical University of Cluj Napoca, Romania

Technically Sponsored by

Networks & Communications Community (NCC)

Computer Science & Information Technology Community (CSITC)

Soft Computing Community (SCC)

Organized By

Academy & Industry Research Collaboration Center (AIRCC)

TABLE OF CONTENTS

The Eighth International Conference on Networks & Communications (NETCOM - 2016)

The Eighth International Conference on Network and Communications Security (NCS 2016)

The Eighth International Conference on Wireless & Mobile Networks (WiMoNe 2016)

The Eighth International Conference on Applications of Graph Theory in Wireless Ad hoc Networks and Sensor Networks (GRAPH-HOC 2016)

The Third International Conference on Signal, Image Processing and Multimedia (SPM 2016)

The Third International Conference on Computer Science, Engineering and Information Technology (CSEIT 2016)

DESIGN OF A SECURE DISASTER NOTIFICATION SYSTEM USING THE SMARTPHONE BASED BEACON

Jae Pil Lee[1] and Jae Gwang Lee[2] and Jun hyeon Lee[3]
Ki-su Yoon[4] and Jae Kwang Lee[5]

[12345]Department of Computer Engineering,
Han Nam University, Dae-jeon City, Korea
{jplee,jglee,jhlee,ksyoon}@netwk.hnu.kr,
jklee@hnu.kr

ABSTRACT

The number of disaster occurrences around the world based on the climate changes due to the global warming has been indicating an increase. To prevent and cope with such disaster, a number of researches have been actively conducted to combine the user location service as well as the sensor network technology into the expanded IoT to detect the disaster at early stages. However, due to the appearance of the new technologies, the scope of the security threat to the pre-existing system has been expanding. In this thesis, the D-SASS using the beacon to provide the notification service to the disaster-involved region and the safe service to the users is proposed. The LEA Algorithm is applied to the proposed system to design the beacon protocol collected from the smartphone to safely receive the notification information as well as to provide the confidentiality during the data transfer between smartphone and notification server.

KEYWORDS

Disaster, notification, Beacon, Security, Smartphone, LEA, Google Chart

1. INTRODUCTION

According to the data announced by the CRED (Centre for Research on the Epidemiology of Disasters) in which the global disaster risk factors and the number of global disaster occurrences by the year are analyzed in accordance to the temporal/spatial distribution., the number of the disaster occurrences such as earthquake, surge, typhoon, flood and forest fire has been indicating an increase every year [1]. The disaster includes natural disasters (typhoon, flood, drought, tsunami and surge), man-made disasters (fire, collapse, explosion, environmental contamination and accident) and social disasters (energy, communication, traffic and infectious disease). The Centers for Disease Control and Prevention has been continuously developing and providing the manuals/measures on preparation for and management of the disaster/emergency. In addition, the Centers for Disease Control and Prevention also has been conducting cooperative projects by constructing the public health crisis management centers at the universities located in the main regions of the United States. Such cooperative projects provide services to individuals, workers and communities nationwide [2].

Natarajan Meghanathan et al. (Eds) : NETCOM, NCS, WiMoNe, GRAPH-HOC, SPM, CSEIT - 2016
pp. 01– 12, 2016. © CS & IT-CSCP 2016 DOI : 10.5121/csit.2016.61501

Based on the development/supply of the ICT (Information Communication Technology), the disaster communication has been making a transition from the control provided by the preexisting disaster management organizations to the construction of the full-range disaster communication system which allows the people regardless of their region to mutually communicate with others. The WORKPAD Project conducted in the European Union is a case where the state-of-the-art technologies are converged with the disaster communication to consider the safety of the field management team, and the Emergency 3.0 Project conducted in Australia is a case where the private/government-based cooperation is used to distribute the disaster information in real-time[3][4]. In Korea, the disaster field is faced with the limited management of the crisis related to the man-made disasters occurring due to the accident death rate and safety ignorance relatively high in comparison to the rapid economic development. To make progress in the disaster/safety areas, the following 4 strategies are being promoted: the construction of the public safety infrastructure, the construction of the natural disaster infrastructure, the connection/use of the private data and the exchange/expansion of the information for the citizen-participated services [5].

The IoT (Internet of Things) is a technology to which various companies and academic circles have been paying their global attention. Through this technology, the users may connect all devices including smartphone and resource-limited sensor to the internet. In addition, the IoT-based devices may be connected with one another to collect, process, exchange and share information. According to Gartner, more than 26 billion devices will be mutually connected by the year 2020, and such connection will create diverse innovations and business opportunities [6]. The smartphone is one of the popular high-performance devices mostly used to actualize the IoT. Such smartphone is a medium suitably used for communicating with the surrounding sensors, immediately applying the information collected through its own sensor to its services and transferring the information collected through the network to the necessary locations. Through the application of the IoT-based communication and sensor network technology, the importance of the system capable of providing the real-time disaster information to the smart mobile devices has been magnified.

The sophistication of the recent cyber threats has been causing social confusion and has been threatening the national security as well. The scope of its use has been expanding into causing financial damages to individuals. Accordingly, an issue has been raised on the information leakage as well as the security. In addition, the sophistication of the malignant codes and hacking technologies has been constructing a structure where the maliciously acquired contents are easily distributed. Since the BLE (Bluetooth Low Energy) based beacon provides a prompt communicative connection to the smartphone without requiring the pairing process, the communication can be conducted based on a small data transfer volume. In addition, the strength is that any smartphone capable of receiving messages is capable of receiving this message. Accordingly, there is no need to protect the data exposed from the smartphone. It is necessary to come up with a security technology which can be used to prevent the abuse of personal information contained in the smartphone, protect the privacy of the disaster notification service users and create a safe use environment.

In this thesis, to conduct a research on combining the user location-based service and the sensor network technology into the expanded IoT technology in order to cope with the home/overseas disaster-related, The D-SASS using the BLE-based beacon to provide the notification service to the disaster-involved region and a safe service to the users is proposed. The LEA (Lightweight Encryption Algorithm) is applied to the proposed system to design the beacon protocol collected by the smartphone to safely receive the notification information as well as to provide the confidentiality during the data transfer between smartphone and notification server. In addition, to monitor the status of the users in the disaster-involved region, the Google Chart is used to visualize the status of the people who received the notification in the disaster-involved region as well as the status of the people in the disaster-involved region on the web. This thesis is organized as follows. In Chapter 2, the precedent researches on the disaster management information system are examined. In Chapter 3, the system for collecting/analyzing the disaster notification system is designed and actualized. Lastly, the conclusion and the future researches to be conducted are proposed.

2. RELATED STUDIES

2.1. Big Data Disaster Prediction Service

The need to manage the disaster through the use of IT technologies is being expanded in order to detect at early stages and minimize the damages caused by enlargement, concentration and globalization of the disasters such as natural disaster and environmental contamination. Some of the IT technologies used to manage the disaster are disaster management robot, disaster safety wireless communications network, CCTV-based monitoring service, smartphone-based forecast/ notification service, computer-based disaster prediction and homeworking through the construction of cyber offices. Such technologies are being actively developed/applied at home and overseas [7]. The IT technology-based safety system can be used to prevent and promptly react to the damages caused by the disaster, and the intellectual image recognition technologies such as CCTV can be used to safely prevent the national level disaster.

In the precedent research [8], the damages caused by the natural disaster were measured to be restored, and the smartphone-based damage measurement standard work process was developed to develop a system which can be used to measure the damages caused by the disaster in the involved field through the use of the smartphone in order to input such measured data into the NDMS (National Disaster Management System. Through such development, the work process was decreased by 56% in comparison to the pre-existing work process. In Korea, a new government operation paradigm known as the Government 3.0 is proposed to provide the nation-customized service through positive disclosure, sharing, communication and cooperation among the departments. In addition to the attention paid by the private companies to the Big Data, a national level strategy is being established as well. The disaster management has been making a transition from the government-based management to the sensing model of disaster issues used for connecting/analysing the public/social data to sense and cope with the home/overseas issues and changes [9].

2.2. Smartphone-based Sensor Information Collection Service

Due to the changing patterns of the disaster management based on the supply of the smartphone, the need to develop diverse mobile apps featuring communication and interaction in the disaster situation has been expending [10]. The beacon is a BLE-based precision location information system and is highly evaluated as a short distance data communication technology. The beacon can be signified as a transmitter using the 2.4GHz bandwidth radio frequency serving as the ISM (Industrial Scientific and Medical) bandwidth to periodically create/distribute signals. In addition, it uses the RSSI (Received Signal Strength Indicator) to mutually interchange the data with the smart mobile to measure the location. It is impossible to confirm the location of the smartphone user through the GPS Signal. However, the Beacon can be used to confirm the precise location of the smartphone user within approximately 5cm distance error, and such beacon can be installed/used indoor/outdoor [11].

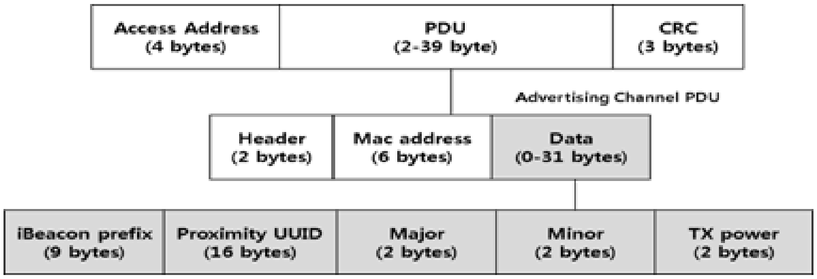


Figure 1. iBeacon Advertisement Packet Structure

The size of the data defined in the iBeacon [12] is 31bytes and the iBeacon prefix is 9bytes. The configuration is as shown in Figure 2. In general, UUID (Universally Unique Identifier) is the unique ID of the product and is constructed based on its own UUID depending on the service. The Major ID and Minor ID are respectively 2bytes and the setting range is from 0 to 65535. The Tx Power represents the RSSI value measures by the beacon at 1m distance. In this thesis, the combination of MAC/UUID/Major/Minor is used as the authentication factor among the authentication servers. In the precedent research [13], due to the expanded importance/value of the Big Data in the IoT environment, a number of home/overseas researchers have been proposing the convergence service through the Big Data analysis. The researches on how to most accurately/promptly collect information from rapidly changing spaces have been conducted. In such researches, the concept of the ‘majority living in the region’ instead of the ‘minority of experts’ is applied and the real-time information is accurately provided based on the information collected from the majority living in the region.

In addition, the sensor information from the natural disaster situation and the sensor information from the region/area where the smart mobile device users participate are collected to conduct a research on the crowd sensing and beacon information collection to create new knowledge. However, a designing for protecting the sensor information is not added during the disaster information collection. Accordingly, it is necessary to create an atmosphere where the personal information of the users can be protected and safely used to create new information.

2.3. Smartphone-based Security Service

The security intelligence field defined by the Gartner [14] Group has been receiving attention as the main alternative, and the technologies for processing/analysing the Big Data are being used to process/analyse diverse types of the long-term accumulated Big Data. The cyber threats appearing from 2010 to the present include insider threats and entering of malicious codes through normal network services. The internal network can be infiltrated at any time through diverse routes/methods. Accordingly, the internal network behaviour analysis technology used for collecting/analyzing the diverse system log information and as well as the dynamic behaviour information occurring in the internal network has been attracting the attention [15]. Since the android platform involves the Java-based programs that can be easily reversed through the app reverse engineering, app pirating/plagiarizing have been occurring frequently [16].

In the precedent research [17], to stop the production/distribution of the illegal/malicious apps through pirating/plagiarizing/repackaging the codes from the android apps, the code obfuscation techniques used for protecting the software programs by modifying the codes so that it is difficult to conduct a counterattack is considered into using the strong Birthmarking to propose a technique detecting/identifying the program pirating through comparing the similarities between the features of the involved programs. The android apps are distributed in the APK (Android

Application Package) and the byte code-level execution file DEX (Dalvik Executable) is included in such apps. In the precedent research, the security is focused on the smart mobile apps featuring strong obfuscation and efficient/reliable anti-pirating..

In the precedent research [18], the block cipher LEA (Lightweight Encryption Algorithm) is an algorithm used for encrypting the 128-bit data block. The 128/192/256-bit keys can be used. The round function of the LEA only consists of the 32-bit ARX (Addition, Rotation, XOR)-based arithmetic operations and therefore is promptly processed in the universal 32-bit software platforms supporting such arithmetic operations. In addition, the arrangement of the ARX-based arithmetic operations within the round function not only sufficiently guarantees the safety, but also features the lightweight actualization through excluding the use of the S-box.

3. DESIGN OF SECURITY NOTIFICATION SYSTEM

In this thesis, the user location-based service and the BLE-based beacon are used to provide the notification service to the disaster-involved region, and the D-SASS is proposed to provide the prompt/safe service to the service users. This system is provided to the disaster information service users in the wireless communication environment and is provided to the smart device users.

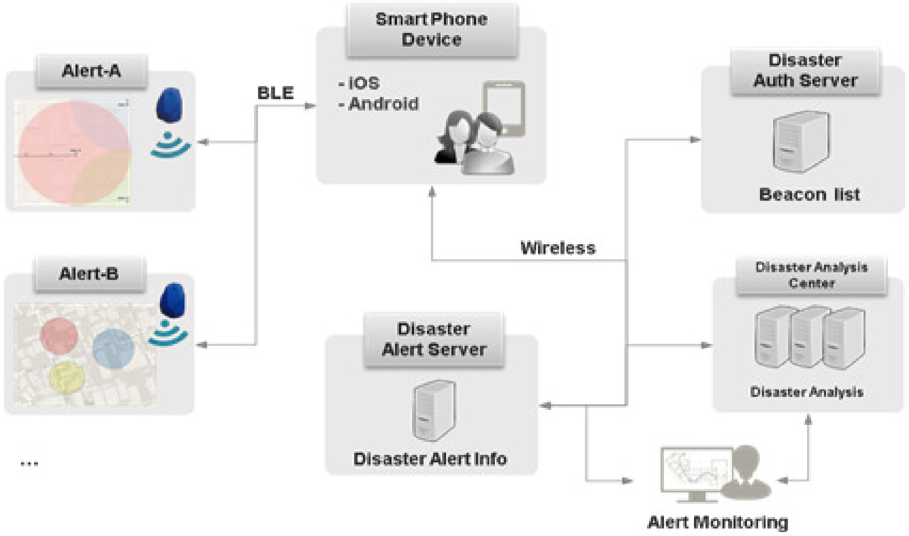


Figure 2. Configuration of Disaster Security Notification System

The proposed system proposes its scope consisting of the surrounding beacon sensor collected by the disaster notification service users within the smartphone, the user GPS information, the disaster notification service, the beacon authentication server and the analysis center. To protect the data collected by the service users during the disaster information collection, the LEA cipher algorithm is used to design the security system for protecting the disaster data collected by the users.

Figure 2 is the configuration of the disaster security alarm system for the DISU (Disaster Information Service Users). The configuration of the security alarm system consists of the smart mobile device collecting the disaster information and displaying the alarm, the disaster notification server providing the disaster information notification/visualization service, the disaster authentication server providing the disaster information beacon and the smartphone information authentication service, and the disaster analysis center server detecting/determining the disaster situation. In this thesis, to provide the safe service to the disaster information service users, the authentication procedures between smartphone and disaster alarm server is designed, the secret key is produced by combining beacon information and user membership information and the produced secret key is used for encrypting/decrypting the personal information data of the disaster information service users within the smartphone.

3.1. Design of D-SASS Encryption Protocol

In this thesis, the beacon packet structure is partially used to product the secret key used for the LEA encryption algorithm in order to design the safe security alarm system providing the notification Beacon (authenticated) information received from the notification server to the smartphone.

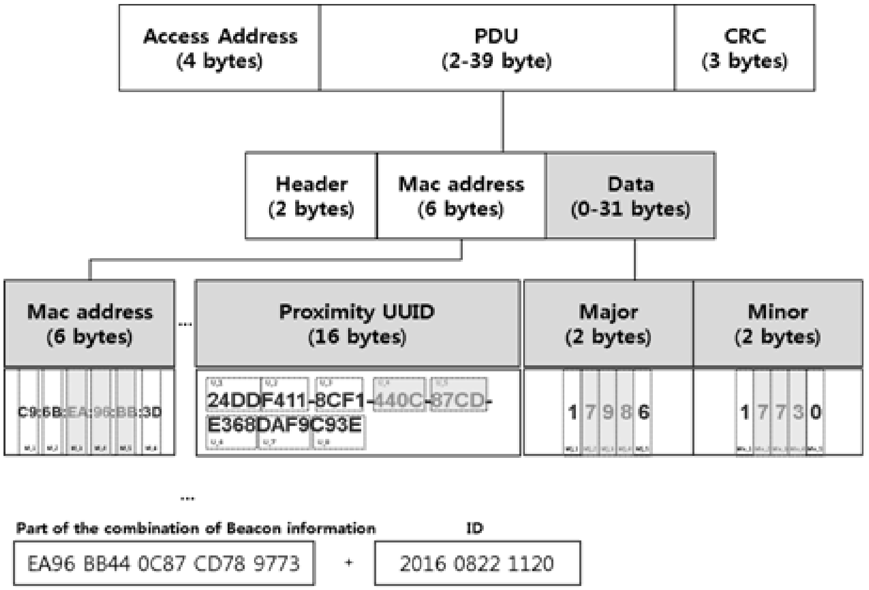


Figure 3. Design of Beacon Encryption Protocol

As shown in Figure 3, in accordance to the IEEE 802.11 Standard [19], the data frame is set, the 4 types of information (MAC, UUID, Major, Minor) are converted into the character string format, the B_Pinfo is combined with the DISU ID, and the secret key is produced. The LEA [20] is used to protect the data of DISU during the transfer of the notification information created in the disaster-involved region. The LEA is a 128-bit block cipher algorithm developed to provide the confidentiality in the high speed environments such as Big Data and Cloud and the lightweight environments such as mobile device. The LEA is included in the target algorithms validated through the CMVP (Cryptographic Module Validation Program) in June 2015[21].

The design of the LEA algorithm is as follows. To use the LEA algorithm, the secret key used for encryption/decryption and the information (IV: Initial Vector) used in the CBC (Cipher-Block Chaining) mode are used. Then the 16-bit secret key of the disaster notification system as well as the initial value is used to encrypt the 128-bit plain text of the disaster notification information.

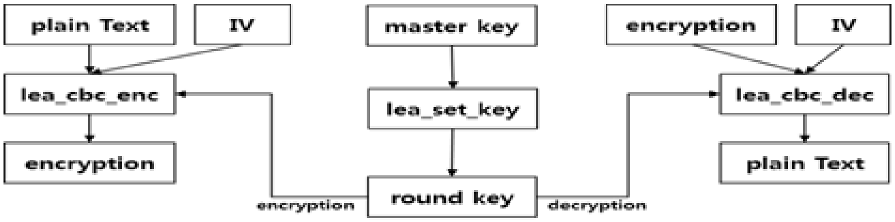


Figure 4. encryption and decryption of LEA Process

3.2. Process of D-SASS

Figure 5 shows the safe notification authentication process among DISU smartphone, notification server, authentication server and analysis center. The overall system consists of registration stage, authentication stage (TYPE_A) and service stage (TYPE_B). The overall system consists of DBS (Disaster Beacon Sensor), SUS (Service User Smartphone), DASS (Disaster notification Service Server), DAS (Disaster Authentication Server) and DACS (Disaster Analysis Center Server). The scope proposed in this thesis is Figure 5 and the security alarm process is as follows.

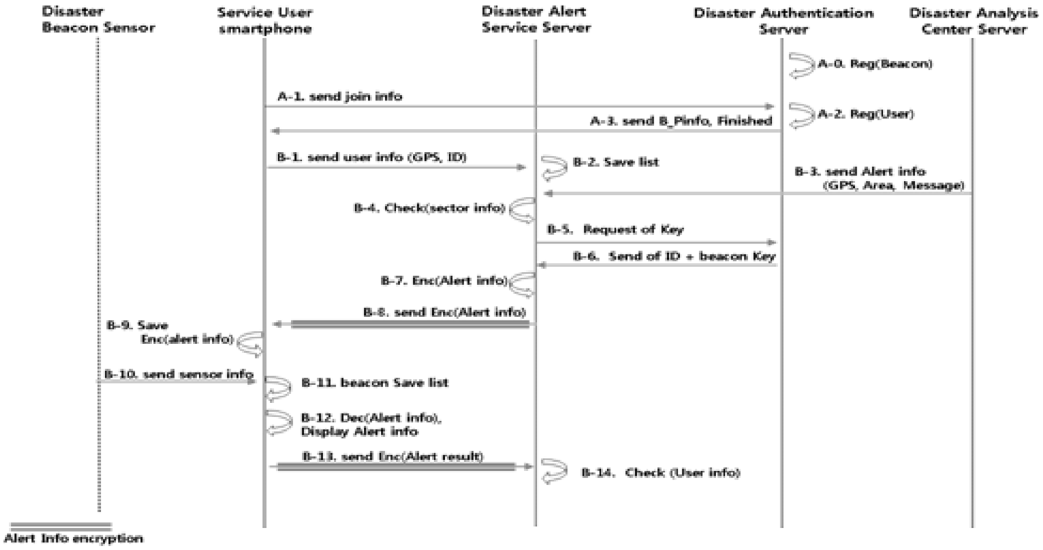


Figure 5. Disaster Security Notification Authentication Protocol Design

The disaster security alarm authentication protocol is processed in 3 steps. Step 1: A-0 is the step for registering the DAS prior to installing the beacon at the involved regions for providing the disaster service. Step 2: A-1, A-2 and A-3 are the steps for conducting the registration. Initially, the app is downloaded into the smartphone of the DISU to process the disaster service registration. For the registration, the ID/PW are issued and the registration information is transferred to the DAS to use the membership information as well as the ID/B_Pinfo as the encrypted key. Then the B_Pinfo is transferred to the SUS. Step 3: B-1 to B-14 are the steps for providing the service.

In Step B-1, the DISU transfers the user GPS information as well as the user ID information to the DASS at a constant interval. In Step B-2, the data transferred in the SUS step is received by the DASS, and the ID/GPS information of the DISU are stored in the database according to the time sequence. In Step B-3, the GPS/Beacon/Disaster Type/Message of the involved region is transferred to the DASS during the provision of the disaster notification from the DACS.

In Step B-4, the data transferred from the DACS is received by the DASS. After the data is received, the location information of the DISU located in the disaster-involved region is analysed to count the number of users in the involved region. In Step B-5, the information required for creating the secret key for encrypting/decrypting the LEA algorithm is requested from the DASS to the DAS in order to protect the disaster notification message. In Step B-6, the B_Pinfo from the advertisement packet structure of the beacon registered in the DAS is combined with the ID information of the DISU to create the secret key. The created secret key is then transferred from the DAS to the DASS.

In Step B-7, the secret key transferred from the DAS to the DASS is transferred, and the LEA algorithm as well as the secret key is used to conduct the encrypted arithmetic operations in order to encrypt the notification information (GPS/Beacon Information/Disaster Type/Message) related to the disaster-involved region. In Step B-8, the encrypted notification information is provided from the DASS to the users in the disaster-involved region through the SUS. In Step B-9, the encrypted notification information is received from the DASS to the SUS and stored in the smartphone.

In Step B-10, the information of the beacon installed in the disaster-involved region is provided to the smartphone of the DISU. In Step B-11, the beacon information received by the DISU is stored in the smartphone. In Step B-12, the ID of the DISU and the B_Pinfo from the received beacon information are extracted and used to create the secret key. In addition, the secret key created based on the encrypted notification information is used as the decryption key for the LEA algorithm to conduct the arithmetic operations required for processing the decryption.

In Step B-13, after the sound/message notification is received from the DASS to the DISU through the smartphone, the current location of the DISU as well as the beacon information is transferred to the DASS. In Step B-14, DASS provides the visualized data and measures the situation of the disaster-involved region based on the information received from the SUS of the DISU to measure the current situation of the DISU in the disaster-involved region.

3.3. Table Information of Development Environment

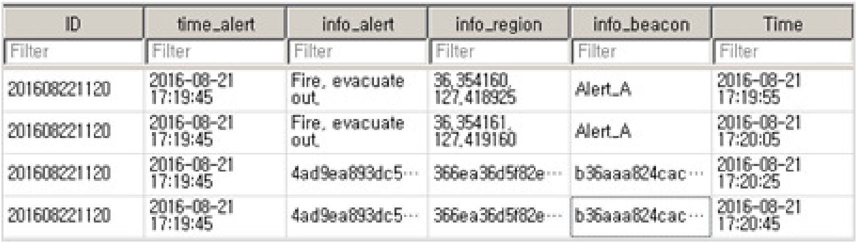
In this thesis, as shown in Table 1, the environment for testing the disaster security alarm system is constructed to apply the encryption between notification server and smartphone in order to design the safe disaster notification system using the smartphone-based beacon.

Table 1. Development Environment

Division	Item	Specification
Beacon	RECO	Bluetooth 4.0
	iBeacon	
NOTIFICATION Server	OS	Windows 7
	Apache	Version 2.2.14.
	PHP	Version 5.2.12
	MYSQL	Version 5.6.31
Smartphone	Galaxy Note4	Android version 6.0.1
Develop Server	OS	Windows 7
	Language	Java, C
	H/W	Intel Xeon, 16GB DDR3
	DB	SQLite
	Tool	Android Studio 1.5.1

The encrypted sections of the notification information are Step B-8 and Step B-13. The design needs to be set so that the disaster notification information received from the DACS to the DASS is stored in the database. As shown in Figure 6, the field of the disaster notification information table needs to be created. The received data as well as the notification time, notification information, region information, beacon information and user information is stored in the DACS.

The DASS uses the user ID value and the B_Pinfo value to create the secret key and uses the LEA algorithm to encrypt and store the data within the field. The stored data is transferred to the SUS in the encrypted format shown in Step B-8 and saved in the SQLite Database.

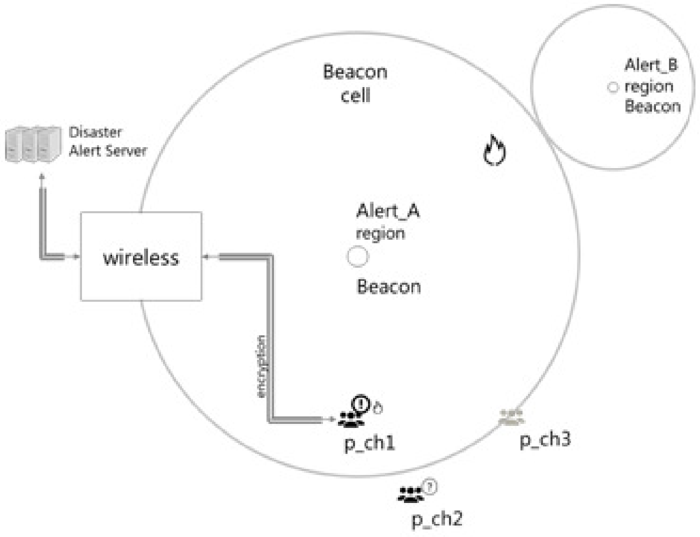


ID	time_alert	info_alert	info_region	info_beacon	Time
Filter	Filter	Filter	Filter	Filter	Filter
201608221120	2016-08-21 17:19:45	Fire, evacuate out.	36.354160, 127.418925	Alert_A	2016-08-21 17:19:55
201608221120	2016-08-21 17:19:45	Fire, evacuate out.	36.354161, 127.419160	Alert_A	2016-08-21 17:20:05
201608221120	2016-08-21 17:19:45	4ad9ea893dc5···	366ea36d5f82e···	b36aaa824cac···	2016-08-21 17:20:25
201608221120	2016-08-21 17:19:45	4ad9ea893dc5···	366ea36d5f82e···	b36aaa824cac···	2016-08-21 17:20:45

Figure 6. Design of Disaster-encrypted Notification Information DB

3.4. Scenarios of Disaster Notification system

The up of Figure 6 is the Beacon notification scenario model included in the disaster notification system for the DISU. If the DISU conducts movement within the Beacon-installed region, the GPS/ID is transferred to the DASS at a constant interval. If the disaster notification occurs in the wireless environment, the users are divided into 3 channels: the users included in the involved region (p_ch1), the users excluded from the involved region (p_ch2) and the users who escaped the involved region (p_ch3) 3.



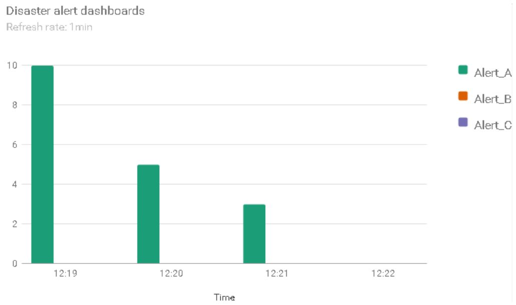


Figure 7. up: Safe Disaster Beacon notification Scenario, down: notification Dashboards

The information of the Beacon installed at the disaster-involved region is provided to the smartphone of the DISU. Only the Beacon information registered in the DAS can be collected and the Beacon information non-registered in the DAS serves as the filter during the collection. The p_ch1 users use the received Beacon information to produce the secret key and decrypt the encrypted notification information to receive the disaster notification sound/message through the smartphone. The down of Figure 6 shows a screen of the disaster notification information transferred to the 10 users included in the Beacon cell of the notification a included in the disaster-involved region as well as the current information of the users transferred from the SUS to the DAS. Then the situation of the p_ch2/p_ch3 within the disaster-involved region is displayed through the use of the Google Chart based on the received information to monitor the situation of the disaster-involved region.

4. CONCLUSIONS

To cope with the frequently occurring home/overseas disaster-related accident, various researches are being conducted to combine the user location-based service and the sensor network technology into the expanded IoT technology in order to detect the disaster at early stages. The preparation for and management of the disaster are considered essential for stabilizing and continuously developing the society. The need to develop a system capable of promptly/efficiently collecting/analysing the risk regions during the disaster occurrence has been expanding. However, the expanded scope of the IoT infrastructure increased the scope of the malicious actions applicable to the disaster system. In addition, an issue is being raised on the security due to the increased damages caused by the random exposure/leakage of the collected personal information.

In this thesis, the D-SASS (Disaster Secure Alarm Service System) using the BLE (Bluetooth Low Energy) based Beacon to provide the notification service to the disaster-involved region and the prompt/safe service to the service users is proposed. The LEA encryption algorithm is applied to the proposed system to design the secret key based on the Beacon protocol information collected from the smartphone to safely receive the notification information of the disaster service users as well as to provide the confidentiality during the data transfer between smartphone and notification server. In addition, to monitor the status of the users in the disaster-involved region, the Google Chart is used to visualize the status of the people who received the notification in the disaster-involved region as well as the status of the people in the disaster-involved region on the web.

It is estimated that the scope of the security threats which may occur to the IoT system from the collection stage to the authentication stage among the Beacon/smartphone/alarm server would be decreased and the damages to the personal information would be prevented. Based on the future disaster big data information, the communication protocol for transferring the real-time disaster notification is to be designed.

ACKNOWLEDGEMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2014R1A1A2055522).

REFERENCES

[1] Guha-Sapir D & Hoyois Ph. , Below. R. Annual Disaster Statistical Review 2013: The Numbers and Trends. Brussels: CRED; 2014.", Centers for Disease Control and Prevention, 2014.

[2] Centers for Disease Control and Prevention, "Emergency Preparedness and Response", Available on http://emergency.cdc.gov/planning/index.asp, 2015.

[3] Ministry of Government Administration and Home Affairs, "Disaster Safety Wireless Network Major Requirements", Vol. 2011, No. 76, pp1-5., 2011.

[4] W.S. Jun, "Disaster-Responsive IT Technology", ETRI, 2013 Electronics and Telecommunications Trends, pp145-153, 2013.

[5] National Information Society Agency, "The new ICT Convergence Strategy Information Security Policies of the disaster areas", Vol. 3, 2014.

[6] Gartner, "Gartner Says the Internet of Things Installed Base Will Grow to 26 Billion Units By 2020", http://www.gartner.com/newsroom/id/2636073, 2013.

[7] Chun, H.W., Electronics and Telecommunications Trends, "Disaster Prevention Information Technology", 2013.

[8] National Disaster Management Institute, "Development of Technology on Linked with NDMS for Disaster Damage Investigation using Smartphone", 2012.

[9] Choi, S. H. & Bae, B.G., The Sensing Model of Disaster Issues from Social Big data, Journal of KIISE: Computer Systems and Theory, Vol.20, No.05, 2014.

[10] Sung, S. J., "How can we use mobile apps for disaster communications in Taiwan: Problems and possible practice," 8th Asia-Pacific Regional ITS Conference. 2011.

[11] ITWORLD, IDG Korea, "Apple's Beacon of position sensing technology operating principle", Available on http://www.itworld.co.kr/slideshow/85994, 2014.

[12] estimote, "Beacon Tech Overview", Available on http://developer.estimote.com/iBeacon/, 2012.

[13] Lee, J. P., International Journal of Applied Engineering Research, "Design of Disaster Information Collection System Base on In-memory in Crowd Sensing Environments", Vol. 5., No.1, pp217-218., 2015.

[14] Gartner, "Security and Risk Management Summit 2014", 2014.

[15] Kim, I. G., ETRI, "Big data analytics technology and cyber security", 2014.

[16] C. Davies, 95% Android game piracy experience highlights app theft challenge, Retrieved May, 15, 2013, Available on http://www.slashgear.com/95-android-game-piracy-experience-highlights-app-theft-challenge-15282064, 2013.

[17] Km, D.J., Android App Birthmarking Technique Resilient to Code Obfuscation, Journal of Korean Institute of Communications and Information Sciences, Vol.40, No.04, pp700-708, 2014.

[18] Hong, D.J., et al., International Workshop on Information Security Applications. Springer International Publishing, "LEA: A 128-Bit Block Cipher for Fast Encryption on Common Processors." 2013.

[19] IEEE Standard for Information technology, "802.11n-2009 - IEEE Standard for Information technology", Available on, https://standards.ieee.org/findstds/standard/802.11n-2009.html, 2009.

[20] KISA (Korea Internet & Security Agency), "Lightweight Encryption Algorithm", Available on https://seed.kisa.or.kr/iwt/ko/sup/EgovLeaInfo.do, 2013.

[21] NIST, "Lightweight Cryptography", http://www.nist.gov/itl/csd/ct/lwc-project.cfm, 2015.

AUTHORS

Lee Jae Pil

Obtained his Bachelor of engineering from Joongbu University, Master of Science degrees in computer science from Hannam University, South Korea. He is submitted his master's thesis on the title of Security framework of big data distributed processing environment using Hadoop. His current doctoral student, research area is in Network, Security, and Mobile Technologies.

Lee Jae Gwang

Obtained his Bachelor of engineering and Master of Science degrees in computer science from Hannam University, South Korea. He is submitted his master's thesis on the title of CCTV Mobile Monitoring System using Kinect with Linux HA. His current research area is in Sensor Network, IoT Security, and Beacon.

Lee Jun Hyeon

Obtained his Bachelor of engineering from Hannam University, South Korea. He has studied the indoor positioning algorithms to master's thesis theme. His current research interests include vulnerability analysis and the indoor positioning technology.

Yoon Gi Su

Obtained his Bachelor of Science from Chungnam University, South Korea. He is currently studying the trends and issues of Big Data with the theme of the master's thesis. His current research interests are Bluetooth Low Energy, Internet Of Things, and security.

Lee Jae Kwang

Obtained his Bachelor of Science, Master of Science and Doctorate degrees in computer science from Kangwoon University, South Korea. Prof. Jae Kwang Lee submitted his master's thesis on the title of Study on Using Text Editor Addressing Mapping Structure and Ph.D. thesis on the title of Information protection protocol in local area networks. His current research area is in Network, Security Technologies. At present he is working as a dean of research and professor of computer science in Hannam University.

SECURITY FOR SOFTWARE-DEFINED (CLOUD, SDN AND NFV) INFRASTRUCTURES – ISSUES AND CHALLENGES

Sara Farahmandian and Doan B Hoang

Faculty of Engineering and Information Technology, School of Computing and Communications, University of Technology Sydney, Australia Sydney
sarah.farahmandian@student.uts.edu.au
Doan.Hoang@uts.edu.au

ABSTRACT

Cloud computing has transformed a large portion of the IT industry through its ability to provision infrastructure resources – computing, networking, storage, and software - as services. Software-Defined Networking (SDN) has transformed the physical underlying network infrastructure into programmable and virtualized networks. Network Functions Virtualization (NFV) has transformed physical telecommunication infrastructures and network functions into virtualised network functions and services. Cloud, SDN and NFV technologies and their associated software-defined infrastructures all rely on the virtualization technology to provision their virtual resources and offer them as services to users. These new technologies and infrastructures invariably bring with them traditional vulnerabilities and introduce new technology-specific security risks. In this paper, we discuss extensively cloud-, SDN-, and NFV-specific security challenges as well as approaches for addressing integrated infrastructural issues where cloud, SDN, and NFV all play their integral parts.

KEYWORDS

Cloud computing, SDN, NFV, virtualization, security challenges, software-defined security, multi-tenancy

1. INTRODUCTION

With a huge increase in the acceptance and usage of cloud computing, the majority of IT services are now being deployed and operated in cloud environments. Cloud computing is a large scalable environment which consists of a large number of physical hosts and virtual machines (VMs) operating and communicating over the cloud network. Each physical server or host may serve as a host to multiple virtual machines by virtue of virtualisation. Since cloud computing supports a multi-tenant environment where each tenant has its own networking requirements based on its clients' demands, one of the challenges of a cloud network is to adapt its network resources dynamically in order to support scalability and maintain real-time configuration while virtual networks are provisioned and migrated dynamically on-demand or virtual machines move from one domain to another. Providing a dynamic and automatic virtual network for cloud multi-tenant

Natarajan Meghanathan et al. (Eds) : NETCOM, NCS, WiMoNe, GRAPH-HOC, SPM, CSEIT - 2016
pp. 13– 24, 2016. © CS & IT-CSCP 2016 DOI : 10.5121/csit.2016.61502

infrastructure is a significant challenge for future of networking architecture. In telecommunication networks, Software-Defined Network (SDN) and Network Function Virtualization (NFV) are two effective technologies with great impact in this area. SDN is based on the separation of the network control from the data forwarding functions, allowing the controller to directly program the underlying infrastructure and present it as a high level, network-functionality abstraction to applications and network services [1]. NFV offers a new approach to design, deploy and manage networking services. It decouples the network functions, such as firewalls, intrusion detection, etc. from proprietary hardware appliances so they can be implemented in software and deployed wherever and whenever needed [2]. Although these new technologies and their associated software-defined infrastructures (Cloud, SDN, and NFV) can solve existing limitations in providing cost effective, on-demand IT services and elastic but scalable network architectures/network services in software-defined, virtualized, multi-tenancy environments, they also present many critical challenges related to both traditional and technology-specific security. This paper presents major security challenges in cloud, SDN and NFV; and solutions for infrastructural security issues. The paper discusses the need for a software-defined security technology for handling software-defined integrated infrastructures and systems. Specifically, in an integrated infrastructure platform such as a data centre or a telecom cloud, where cloud, SDN and NFV functionalities are integrated, the paper discusses the compound security challenges and suggest possible solutions using virtualization technology.

The paper is organized as follows. Section 2 provides essential definitions and characteristics of cloud, SDN, NFV and virtualization technologies. Section 3 presents issues and challenges these technologies are facing. Section 4 presents software-defined security solutions for cloud and SDN. Sections 5 discusses security issues and a software-defined security solution for an integrated infrastructure platform with virtualization technology. The conclusion is in section 6.

2. CLOUD, SDN, NFV AND VIRTUALISATION TECHNOLOGIES

In this section we explain the essential characteristics of cloud computing, SDN, NFV, and Virtualization.

2.1 Cloud Computing

Cloud computing has evolved into a key structure for IT industries for providing users on-demand services. Cloud architecture enables users to access cloud services over the Internet at any time regardless of their location through application software like web browsers. Cloud computing resources such as virtual servers, virtual storage, virtual networks and virtual services, are made available using virtualization technologies. The National Institute of standard and technology (NIST) recently offered an explanation for defining the cloud computing. In this definition, Cloud computing is a computing model that enables omnipresent, convenient and on-demand network access to a shared pool of configurable computing resources such as networks, storages, servers, applications, and services. Cloud computing offers three service models known as Software as a service (SaaS), Platform as a service (PaaS), and Infrastructure as a Service (IaaS) [3].

SaaS model enables users to access their services through a web application but without the ability to control the network infrastructure and operating systems. PaaS model provides a platform for software developers to use application development languages and tools such as java, .net, python and etc. for creating, compiling, designing, running, deploying, and testing their own software applications. IaaS is a form of cloud computing which provides access to computing resources in a virtualized environment. Virtualization is deployed to pool all underlying physical resources together and offer them as virtual resources on-demand and elastically in the form of

IaaS, PaaS, and SaaS [4]. OpenStack is a major open source cloud computing platform that orchestrates and manages shared storage, compute, and network resources using multiple hypervisors based on a set of applications and open-source. OpenStack is used as a cloud framework for creating public and private clouds.

2.2 Network Functions Virtualization (NFV)

Network Function Virtualization (NFV) is proposed recently aiming to virtualize an entire class of network component functions using virtualization technologies. NFV enables network functions to be realized and executed as software instances in a VM on a single or multiple hosts instead of customized hardware appliances. NFV offers a new means for creating, deploying and managing networking services. Network Function Virtualization can be applied to both data and control planes in fixed or mobile infrastructures. NFV provides telecommunication operators the ability to combine numerous different types of network equipment into high volume switches, servers, and storage inside data centres, network nodes, and end user premises. NFV implements network functions using software virtualization methods and performs them on top of underlying hardware equipment. These software-based virtual functions can be installed and deployed flexibly and strategically based on tenants' requirement without the need for new hardware equipment. A hypervisor is responsible for controlling network functions within a supporting NFV infrastructure. NFV technology helps cloud tenants to avoid vendor lock-in problem by allowing them to use multiple virtual appliances from different vendors while using different hardware platforms and/or hypervisors.

In today's market, NFV concentrates on providing four categories of software-based virtual network functions known as Virtual Switching, Virtualized Network Appliance (security functions such as IDS, Firewalls, and etc.), Virtualized Network Services (load balancers, network monitoring tools, traffic analysis tools), and Virtualized Applications (any available application in the network environment) [5]. Enable dynamic deployment of NFV within a networking platform is a big challenge. Traffic of a network function (NF) must be isolated at multiple levels - services, virtual networks and tenants' levels - and hence, a comprehensive controller is required to provide strict multilevel isolation within the NFV Infrastructure. The combination of SDN and NFV can solve these challenges in both dynamic network infrastructure and functionality of an integrated cloud-network environment.

2.3 Software-Defined Systems

Software Defined System (SDSys) is conceived to address control and management challenges which exist in cloud computing. SDSys is a concept that provides an abstraction of actual hardware at different layers based on software components. This type of abstraction enables system administrators to create a centralized decision-making system to handle and monitor all control and management decisions instead of having a decentralized system where each component only manages itself [6]. Among all SDSys subsystems, SDN is the most well-known.

2.4 Software-Defined Network (SDN)

Software-Defined Networking is developed as a technology to remove the current black box network infrastructure restrictions. This is done through the separation of the decision-making functions from the data forwarding functions, allowing them to evolve separately into a centralized and programmable control plane and a simple and high-performance data plane operation respectively [7]. According to [7], Software-Defined Network architecture consists of three layers known as Data Plane, Control Plane, and Application Plane. SDN devices are all placed at Data plane layer. SDN controller (or group of controllers) is located at the control plane layer, and applications and network services are on the application plane layer. SDN devices simply forward packets according to instructions programmed by the SDN controller. An SDN application, gaining network capability abstraction from the controller, has the ability to determine traffic streams and routes on the network devices to fulfil the requirements responding to user's dynamic requests [7, 8].

The main responsibility of the SDN controller is to program and centrally control SDN devices forwarding behaviours with the support of a comprehensive information database of all underlying network infrastructure operations. The SDN controller uses interfaces for communicating with other layers. To communicate with the data/infrastructure layer, a Southbound Application Interface (API) Interface is used for programming and configuring network devices. To communicate with the application layer a Northbound Interface is provided for the interaction between the SDN controller and applications. East/West Interfaces are for information exchange between multiple or federated controllers. The OpenFlow protocol has been developed and widely adopted as one of the southbound interfaces between SDN controllers and SDN switches. OpenFlow uses a secure channel for message transmission over the transport layer security (TLS) connection.

2.5 Virtualization

Virtualization is the technology that simulates the interface to a physical object by multiplexing, aggregation, or emulation. With multiplexing, it creates multiple objects from one instance of a physical object. With aggregation, it creates one virtual object from multiple physical objects. With emulation, it constructs a virtual object from a different type of physical object.

Virtualization is critical to cloud computing, SDN and NFV as it allows abstraction of the underlying resources for sharing with other tenants, isolating of users in the same cloud/network, and isolation of services and functions running on the same hardware. It also plays an important role in the development and management of services offered by a provider. Virtualization is often introduced as a software abstraction layer placed between operating systems and the underlying hardware (computing, network, and storage) in the form of a hypervisor. In cloud data centres since the hypervisor manages the hardware resources, multiple virtual machines each with its own operating system and applications and network services, can run in parallel within a single hardware device [9]. Virtual technology thus allows multi-tenancy, isolate workloads, enhances server utilization and provides elastic and scalable resources/services to its users.

Virtualization technology has been deployed by enterprises in data centres storage virtualization (NAS, SAN, database), OS virtualization (VMware, Xen), software or application virtualization (Apache Tomcat, JBoss, Oracle App Server, Web Sphere), and Network Virtualization [10]. Virtualization technology enables each cloud tenant to perform its own services, applications, operating systems, and even network configuration in a logical environment without any consideration about physical underlying infrastructure [11]. The technology enables Network Functions Virtualization (NFV) and Software-Defined Network (SDN) the ability to create a scalable, dynamic, and automated programmable virtual network functions and virtual network infrastructures in integrated cloud platform such as telecom clouds.

3. Security Issues and Challenges in Cloud, SDN, and NFV

3.1 Cloud computing security challenges

As a cloud has become a large-scale and complex infrastructural environment, it becomes more vulnerable to both traditional and new security threats related to its structure and elements. NIST declares security, portability, and interoperability as main obstacles for adopting to cloud environment completely. Some of traditional security issues found in the cloud infrastructure are data access control (illegal access to confidential data), loss and data leakage, trust, isolation. Cloud-specific security issues include insecure interfaces and APIs, malicious insider, account or service hijacking, virtualization security, and service interruption. We discuss these critical and significant security challenges below.

Insecure interfaces and APIs. Cloud providers deliver services to their customers through software interfaces mostly integrated with the web application layer. The stability of cloud components is dependent upon the security level of these APIs within the cloud infrastructure. Insecure cloud APIs can cause various threats related to confidentiality, availability, integrity, and accountability. These API functions and web applications share a number of vulnerabilities which may result in high level security problems. Consequences of any malfunction in APIs may allow malicious codes to be imported inside the cloud and expose user confidential data. Although strong authentication methods, proper access controls, and encryption methods may solve some of the above problems, still, there are serious gaps especially related to the inability of massive auditing and logs. Any APIs that will interact with sensitive data within cloud infrastructure must be protected with a secure channel such as SSL/TLS.

Malicious insider. This type of threats is one of the most serious cloud-specific security challenges according to the Cloud Security Alliance (CSA) cloud security threat list. It happens when an employee of cloud service providers (CSPs) abuses his/her level of access to gain confidential information of cloud customers for any nefarious purposes. The worst case is when a malicious system administrator has access to client resources hosted on virtual machines and data stores. So detecting such indirect accesses to client data is one of challenging tasks in cloud infrastructure.

Account or service hijacking. It is a kind of identity theft that aims to deceive end-users to obtain their sensitive data. If an attacker gains control of a user account it can snoop on all customer's activities, manipulate and steal their data, or redirect the customer into inappropriate sites. This kind of threats can be accomplished through phishing email, faux pop-up windows, spoofed emails, buffer overflow attacks which result in the loss of control of the user's account.

Virtualization security. Since virtualization is a crucial technology in cloud infrastructure, any vulnerability can place the whole system in a high-security breach. For example, any error and vulnerability inside the hypervisor can allow an attacker to launch VMs attacks (shutting down VMs) or monitor others VMs and their shared resources. A compromised VM can inform an attacker of the underlying network operation for exploitation of existing network vulnerabilities. It also enables an adversary to compromise the hypervisor and achieve control over the whole system. Local users and malicious codes can bypass security boundaries or even gain privileges to cause damages to the infrastructure and its users through vulnerabilities found in virtualization software.

Service interruption. It is a vital security issue in cloud computing since everything in the cloud is defined as service. Service interruption is placed in the category of threats related to the availability of cloud services. DDoS attack is usually attempted against Internet services with large population of users and it is more so against cloud as a centre of high number of cloud services and users. These attacks may render services and computing resources unavailable. A DDoS attack may occur when an attacker gains access to tenant's VMs credentials due their vulnerabilities.

3.2 NFV Security Challenges

Most critical security challenges in NFV are related to network function generator/hypervisor, security of virtual functions, performance isolation, communication and functional/service interfaces, multiple administrative isolation, and secure crash of virtual network functions.

Hypervisor security. The main security issue in virtualized environments and especially NFV is related to hypervisor vulnerabilities. A hypervisor creates VMs inside the infrastructure and has the ability to monitor each VM's operating system. According the European Telecommunications Standards Institute (ESTI), this feature introduces high security risk to NFV in terms of Confidentially, Integrity and Availability (CIA). It may allow an attacker to view, inject, or modify operational state information connected with NFV in direct/indirect method and as a result the attacker is able to read/write contents of resources such as memory, storage, and other components of NFV. Hypervisor hijacking is a type of attacks that allow an adversary to take control of a hypervisor and access all VMs created by that particular hypervisor, or other less insecure hypervisors inside the infrastructure. In the worst case it may even introduce misconfigurations in SDN controllers when integrated with NFV technology. Furthermore, existing errors or bugs inside a virtual function or a hypervisor may allow an attacker to compromise other virtualized network functions for more serious attacks.

Virtual network function security. Virtual network functions encounter attacks common to those on physical network functions such as sniffing, denial of service, and spoofing. Insider attacks are possible on virtual network functions when a malicious administrator, who has a specific access right, gains access through other virtual functions within the infrastructure. Insider attacks can modify data in network equipment and introduce unauthorized configuration of network functions. In a public deployment of NFV it is possible for a malicious third-party or remote client to gain access through the network to control the VNFs. A malicious or compromised virtual network function inside the NFV infrastructure can monitor activities of other virtual functions or even send fraudulent instructions through the hypervisor to disrupt their operations.

Performance Isolation. Lack of inappropriate isolation among virtual functions can cause data leakage similar to the way a VM can access through another VM data (VM-to-VM attack). Performance isolation is one of many specific security concerns in NFV infrastructure. A proper virtualization technology has to isolate VMs from one another to ensure that crashes, hangs, loops, or compromises in one VM do not affect others, however, VMs isolation is difficult to achieve due to variable usage of resources and workloads among them. According to the ESTI, network and I/O partitioning and shared core partitioning are two major issues in performance isolation. Isolating network workload from other functions is a difficult task since it can be placed over various distributed network resources and can be dynamically changed at any point in time, particularly when numerous virtual functions in the NFV Infrastructure (NFVI) share resources [12]. Lack of complete isolation can be exploited by an adversary to gain information about a compromised victim. Insufficient isolation mechanism may allow cross virtual network side-channel attacks that threaten VNFs hosted in a NFV shared infrastructure. It is possible that a side-channel attack can bypass compulsory access controls to violate resource isolation.

Communication and functional/service interfaces. New security threats associated with new interfaces present other critical challenges related to interconnectivity between NFV end-to-end components, such as communication between VNF components, communication between VNF and VNF manager, communication between VNF and NFVI, and communication between VNFs. NFV encompasses different types of network and security functions, so defining standard interfaces for different security functions is one of the security challenges in a virtualized network infrastructure. Each tenant may have different security services with different user authentication methods, privilege control schemes, and network configurations. So the way a network function communicates with one another and other tenants' functions through a standard interface is a huge challenge in NFV technology. Currently, there is no standard communication interfaces in NFV technology.

Multiple administrative isolation. It is an NFV security challenge related to the existence of multiple administrative domains in the same platform. Multiple administrator domains imply different administrator privilege domains for network, hypervisor, storage, compute, NFV orchestration, VFNM (Virtual Function Network Manager), and network services running in the platform. Requirements for an administrative role for each of the above domains are different and involve various levels of policies. Security is even more critical when there are virtualization infrastructure administrator roles with higher privileges than the administrator of existing virtualized function within the NFVI.

Secure crash of virtual network functions. Components crash in any infrastructure and system can cause security problems and in virtual environment the impact is more severe. According to the ETSI, a crash of any virtualized function within NFVI can bring about critical security issues which allow attackers to gain access to information through existing insecure data on that particular component [12]. It is so critical that a VNF component should be reinstalled securely after a crash. It should be noted that many important components in the NFV framework might be at high risk states during a crash; these include VNF component instances, network and storage resources attached to virtual network functions. Availability of services is also will be affected due to a function crash [12].

3.3 SDN Security Challenges

As with other new technologies, SDN suffers from both existing security threats in traditional networks and new challenges due to SDN architecture. Since SDN uses virtualization technology to virtualize networks (VNs), it inherits traditional security problems related to the virtualization of virtual machines as well as new security issues related to the virtualization of network hypervisors and their isolation. It also suffers threats such as Dos/DDoS attack, with higher impact because of the centralized architecture of SDN control. SDN introduces new and critical security challenges due to its architecture, including security of SDN controllers, forwarding plane security issues, unauthorized access, routing policy collision, fraudulent flow rules insertion or tampering in switching level, insecure interfaces, and system level SDN security challenges.

SDN controllers. Since SDN controller is a core element in the SDN architecture; if it is compromised the whole system is placed in a high risk of failure. The majority of security challenges related to SDN controller are around the vulnerabilities at the controller plane where an attacker can get hold of the control function to compromise integrity, confidentiality, and availability of SDN [13]. Since SDN decouples the data plane from the control plane, it is the responsibility of centralized controller to deal with all incoming network flows. As a consequence, the controller itself is a key bottleneck and is the target for various attacks such as flooding and DDoS attacks. An SDN controller can be implemented in a virtual or physical server with associated resources. An attacker can launch a kind of resource consumption attack on the controller to render it unavailable in response to flow rules coming from underlying switches and force it to respond extremely slowly to packet-in events or sending packet-out messages. A DoS/DDoS attack is one of the most serious security threats against SDN controller when an attacker endlessly sends IP packets with different headers to the controller to put it in the nonresponsive state.

Forwarding plane security issues. There are two specific security challenges in the forwarding plane of SDN architecture. The first and most critical issue is related to identifying genuine flow rules from malicious or fake rules within the infrastructure where the SDN controller is responsible for all decision making functions. A compromised controller can simply transmit false flow rules within underlying virtual network elements. The second security challenge is that it in vulnerable to saturation attack [14] due to the limited storage capacity for flow rule entries in flow tables of SDN OpenFlow switches.

Unauthorized access. A critical security challenge in SDN is related to unauthorized access in an SDN architecture- unauthorized access through the SDN controller or unauthorized access through the applications- where a large number of third-party applications operate. One of the serious security breaches in SDN is when an authorized SDN component accesses SDN services or controller without having the appropriate level of access and modifies network data or reprograms the SDN controller components [15].

Routing policy collision. Policy collision is another specific security challenge in SDN architecture when various vendors and third party applications using different configurations and programming models. This is critical since a malicious component can delete, insert, or modify existing and predefined policies of flows inside the SDN controller. Separate servers or application with different policy rules may result in policy conflict with each other.

Fraudulent flow rules insertion or tampering in switching level. A compromised or malicious application can generate fraud flow rules while communicating with the controller. An attacker can inject fake flow rules through the switches by exploiting vulnerabilities of southbound interfaces. It is possible for attacker to tamper with network information by modifying flows in flow tables. These malicious flow rules can cause network to behave abnormally. For instance, [16] introduced an attack in which an attacker generates forged link layer discovery protocol (LLDP) packets through an OpenFlow network to create vulnerabilities on internal links between two switches. An adversary can also insert malicious flow rules by monitoring the traffic from OpenFlow Switches.

Insecure interfaces. Another critical security challenge in SDN infrastructure is related to insecure Application Programming Interfaces (APIs): Northbound, Southbound, and East and West Interfaces. This security issue is critical since all communications between the SDN controller and the application layer, the underlying forwarding layer, or even the communication between multiple controllers, go through these interfaces. For instance, vulnerabilities and the lack of standard protocol in northbound interface may enable attackers to interfere with the operation of both the application and the controller and send malicious request through the controller or network elements or even generate flooding attack with purpose of disrupting its operation. An adversary is also capable of sending a large number of requests through the northbound interface to occupy the interface bandwidth. In a multi-domain multi-controller environment, controller's communication goes through the East/West APIs. These SDN controllers may be from different vendors and do not have a common secure channel between them. Message among them may be sniffed by an attacker through vulnerabilities of East-West APIs and sensitive information may be exposed.

System level SDN security challenges. A specific SDN system level security concerns auditing processes. As it is essential to keep comprehensive state information of network devices in the infrastructure to prevent unauthorized access, providing an auditing and accountability mechanism in SDN is a critical security challenges [15].

4. RECENT SOFTWARE-DEFINED SECURITY SOLUTIONS FOR CLOUD AND SDN

In this section we discuss and tabulate a number of software-defined security solutions for SDN and cloud infrastructures. As SDN and NFV are relative new technologies, infrastructures based on them are still being adopted and developed, security issues are being explored and discovered. Currently, only a limited number of solutions exists. Most of them adopt the logically centralized control paradigm of SDN in building software-defined security solutions. Several efforts are described below.

[17] proposed a Software Defined Security Architecture (SDSA) that has the ability to separate security controls from security executions, improves scalability and security of systems and decreases the costs of software developments. The authors provided two structures (Physical and Logical) for the architecture to allow both business logical providers and security developers to only work within their scope of expertise without concern about the design and implementation of security structures or development of business logic programs.

[18] proposed a framework for protecting network resources via SDN-based security services using an Interface to Network Security Functions (I2NSF). The aim was to create a self-governed protection system against network attacks, capable of providing rapid responses to new threats.

In [19], a comprehensive security architecture was proposed to deliver a range of security services including enforcing mandatory network policies, packet data scan detection, transforming network policies into flow entries, authentication, and authorization for solving security challenges related to policy enforcement and attack detection for SDN architecture.

[20] proposed an architecture for enhancing network security using network monitor and SDN control as separated functions. The OrchSec architecture adopts the separation principle of SDN by decoupling of monitoring and control functions. This allows flexible and more comprehensive and intelligent control over security functionality and activities and also reduces overhead on SDN controller.

The table 1 provide a summary of other recent efforts in providing software-defined security solutions for cloud and SDN.

5. SECURITY ISSUES AND CHALLENGES IN AN INTEGRATED CLOUD-SDN-NFV INFRASTRUCTURE PLATFORM

Cloud computing demonstrated how best computing and storage resources can be virtualised and provisioned on demand and offered as IT services. More importantly, its effective orchestration of services offers an excellent model for resources and service management. SDN and NFV demonstrated most effective way network resources and services (network infrastructures, network functions, and connectivity services) can be created and managed. Cloud needs SDN and NFV to be integrated seamlessly to be able to offer truly any resource as a service. SDN and NFV need to include cloud management infrastructure to offer network services and functionality. For example, existing telecommunications network infrastructures and service models are too rigid and they have to evolve into a form of telecom cloud to be able to offer emerging and flexible services to its customers. An integrated software-defined infrastructure that seamlessly integrates cloud, SDN and NFV will certainly create a powerful service model that incorporates all the best features of these technologies.

Two major issues concerning cloud, SDN, NFV and the integrated software-defined infrastructures are the security of the virtualization technology itself and the complexity of the virtualized interconnecting infrastructure. Cloud and SDN networks are facing an increasing complexity of emerging social networks, applications and services and their associated security problems. The whole range of problems include scalability of cloud networks, the complexity of the way network function communicates to each other, the lack of a centralized infrastructure control component, policy enforcement, dynamic workloads, multi-tenancy, isolation of tenants, services, resources (virtual networks, virtual machines, virtual storage). SDN and NFV allow tenants to share the underlying physical network to create their own virtual networks, network functions and services with their policy in a cloud environment. Integrating cloud, SDN and NFV into a software-defined infrastructure provides a truly scalable, dynamic, and automatic programmable platform for creating *everything as a service* on demand.

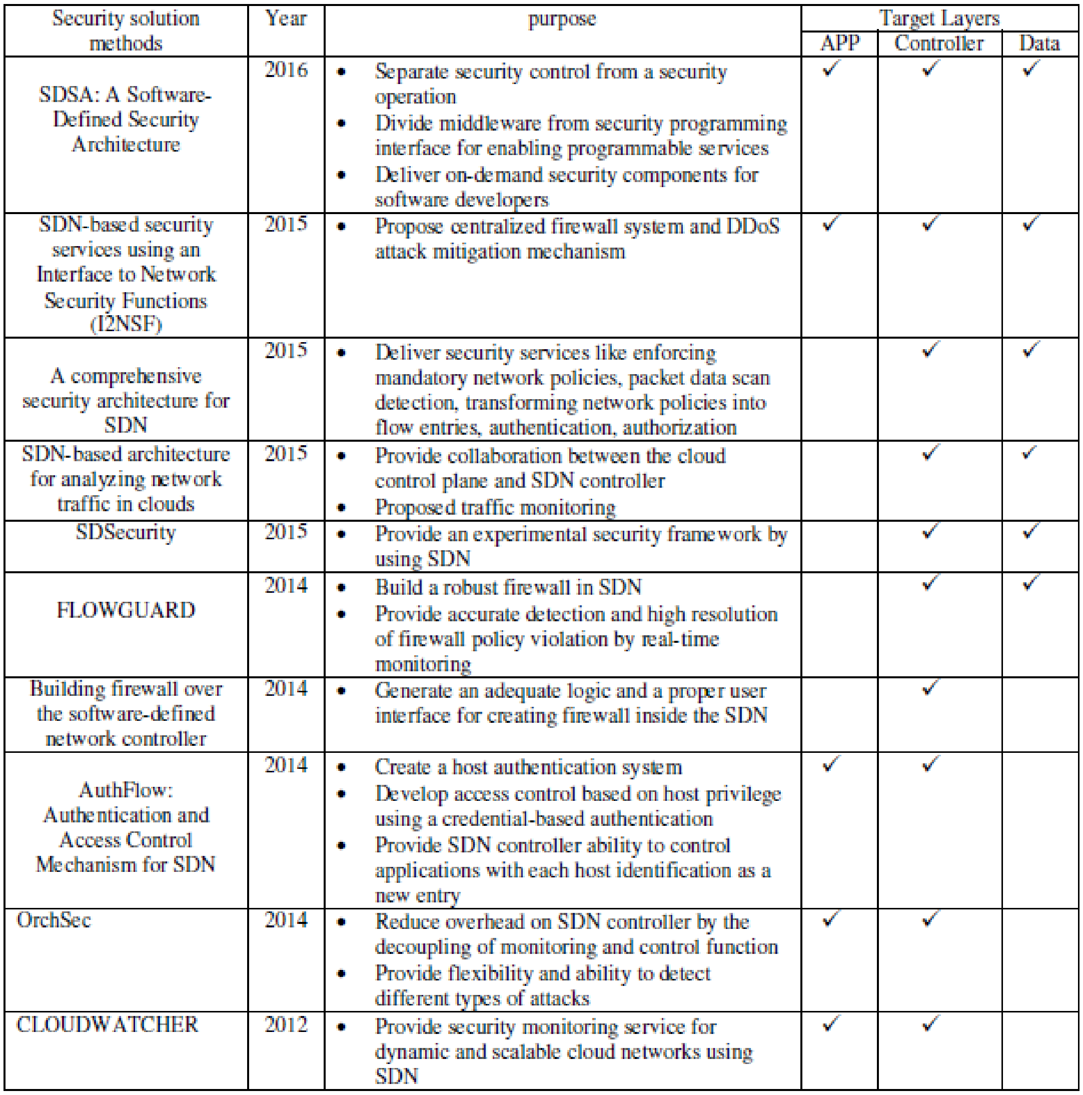
All these infrastructures rely on virtualization as the core technology. Virtualization is pervasive in almost all components of the service infrastructures: virtual machines, virtual networks, virtual storage, virtual network functions, and virtual services. Virtualization, however, brings with it new security challenges in the way virtual elements are created and maintained. For the security of the infrastructure, all virtual elements have to be secure for their whole lifecycle; their creators (hypervisors) must be trusted and secure; appropriate isolation among servers, among services, and among tenants must be preserved.

Clearly, although integration of cloud, SDN, and NFV into a truly service infrastructure provides is beneficial to both service providers and service users, the complexity of security of each technology, of virtual components, of individual infrastructures present a major obstacle for a comprehensive integration. One important aspect of virtualization is that it introduces boundaries that are invisible to traditional security mechanisms at various levels. In order to deal with this integrated software-defined infrastructure, one should use the very virtualization technology to provide security of the overall infrastructure; one should deploy the logically centralized paradigm of SDN and NFV to separate security control from functionality of security network functions. We suggest Software-Defined Security (SDSec) in that spirit to create a centralized security infrastructure for the cloud-SDN-NFV infrastructure platform. SDSec provides a centralized security controller over the infrastructure. The SDSec controller will possess the ability to create its own flexible interconnecting infrastructure for connecting its security function elements. It will have the ability to program and manage its security function elements autonomously. Security function elements are both virtual and physical: networks, and security functions. However, there are many open questions on how best to secure a software-defined integrated infrastructure related to all the security issues and challenged discussed in previous sections.

6. CONCLUSION

Cloud computing has been most effective in orchestrating and provisioning IT resources and offer them as on-demand services. SDN and NFV are most effective in provisioning network infrastructures and network services. Seamlessly integrated, these provide a most powerful software-defined infrastructure to provision *everything as a service*. The main obstacle is the security of the underlying virtualization technologies and their virtualized resources. This paper discussed at length specific security issues and challenges concerning cloud, SDN, and NFV. The paper discussed the need for a software-defined security technology and software-defined control paradigm to handle software-defined integrated infrastructures and systems.

Table 1. Proposed Security solution for SDN



Security solution methods	Year	purpose	Target Layers		
			APP	Controller	Data
SDSA: A Software-Defined Security Architecture	2016	• Separate security control from a security operation • Divide middleware from security programming interface for enabling programmable services • Deliver on-demand security components for software developers	✓	✓	✓
SDN-based security services using an Interface to Network Security Functions (I2NSF)	2015	• Propose centralized firewall system and DDoS attack mitigation mechanism	✓	✓	✓
A comprehensive security architecture for SDN	2015	• Deliver security services like enforcing mandatory network policies, packet data scan detection, transforming network policies into flow entries, authentication, authorization		✓	✓
SDN-based architecture for analyzing network traffic in clouds	2015	• Provide collaboration between the cloud control plane and SDN controller • Proposed traffic monitoring		✓	✓
SDSecurity	2015	• Provide an experimental security framework by using SDN		✓	✓
FLOWGUARD	2014	• Build a robust firewall in SDN • Provide accurate detection and high resolution of firewall policy violation by real-time monitoring		✓	✓
Building firewall over the software-defined network controller	2014	• Generate an adequate logic and a proper user interface for creating firewall inside the SDN		✓	
AuthFlow: Authentication and Access Control Mechanism for SDN	2014	• Create a host authentication system • Develop access control based on host privilege using a credential-based authentication • Provide SDN controller ability to control applications with each host identification as a new entry	✓	✓	
OrchSec	2014	• Reduce overhead on SDN controller by the decoupling of monitoring and control function • Provide flexibility and ability to detect different types of attacks	✓	✓	
CLOUDWATCHER	2012	• Provide security monitoring service for dynamic and scalable cloud networks using SDN	✓	✓	

REFERENCES

[1] A. Manzalini and N. Crespi, "SDN and NFV for Network Cloud Computing: A Universal Operating System for SD Infrastructures," in Network Cloud Computing and Applications (NCCA), 2015 IEEE Fourth Symposium on, 2015, pp. 1-6.

[2] L. R. Battula, "Network security function virtualization (NSFV) towards cloud computing with NFV over Openflow infrastructure: Challenges and novel approaches," in Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on, 2014, pp. 1622-1628.

[3] C. Rong, S. T. Nguyen, and M. G. Jaatun, "Beyond lightning: A survey on security challenges in cloud computing," Computers & Electrical Engineering, vol. 39, pp. 47-54, 2013.

[4] M. Yang and H. Zhou, "New Solution for Isolation of Multi-tenant in cloud computing," 2015.

[5] R. Mijumbi, J. Serrat, J.-L. Gorricho, N. Bouten, F. De Turck, and R. Boutaba, "Network function virtualization: State-of-the-art and research challenges," IEEE Communications Surveys & Tutorials, vol. 18, pp. 236-262, 2015.

[6] Y. Jararweh, M. Al-Ayyoub, E. Benkhelifa, M. Vouk, and A. Rindos, "Software defined cloud: Survey, system and evaluation," Future Generation Computer Systems, vol. 58, pp. 56-74, 2016.

[7] D. Hoang, "Software Defined Networking–Shaping up for the next disruptive step?," Australian Journal of Telecommunications and the Digital Economy, vol. 3, 2015.

[8] K. Govindarajan, K. C. Meng, and H. Ong, "A literature review on Software-Defined Networking (SDN) research topics, challenges and solutions," in Advanced Computing (ICoAC), 2013 Fifth International Conference on, 2013, pp. 293-299.

[9] J. Sahoo, S. Mohapatra, and R. Lath, "Virtualization: A survey on concepts, taxonomy and associated security issues," in Computer and Network Technology (ICCNT), 2010 Second International Conference on, 2010, pp. 222-226.

[10] Y. Xing and Y. Zhan, "Virtualization and cloud computing," in Future Wireless Networks and Information Systems, ed: Springer, 2012, pp. 305-312.

[11] C.-J. Chung, "SDN-based Proactive Defense Mechanism in a Cloud System," ARIZONA STATE UNIVERSITY, 2015.

[12] N. F. V. NFV, "Draft ETSI GS NFV-SEC 001 V0. 2.1 (2014-06)," 2013.

[13] A. Feghali, R. Kilany, and M. Chamoun, "SDN security problems and solutions analysis," in Protocol Engineering (ICPE) and International Conference on New Technologies of Distributed Systems (NTDS), 2015 International Conference on, 2015, pp. 1-5.

[14] I. Ahmad, S. Namal, M. Ylianttila, and A. Gurtov, "Security in software defined networks: A survey," IEEE Communications Surveys & Tutorials, vol. 17, pp. 2317-2346, 2015.

[15] S. Scott-Hayward, S. Natarajan, and S. Sezer, "A survey of security in software defined networks," IEEE Communications Surveys & Tutorials, vol. 18, pp. 623-654, 2015.

[16] S. Hong, L. Xu, H. Wang, and G. Gu, "Poisoning Network Visibility in Software-Defined Networks: New Attacks and Countermeasures," in NDSS, 2015.

[17] L. Yanbing, L. Xingyu, J. Yi, and X. Yunpeng, "SDSA: A framework of a software-defined security architecture," China Communications, vol. 13, pp. 178-188, 2016.

[18] J. Kim, M. Daghmehchi Firoozjaei, J. P. Jeong, H. Kim, and J.-S. Park, "SDN-based security services using interface to network security functions," in Information and Communication Technology Convergence (ICTC), 2015 International Conference on, 2015, pp. 526-529.

[19] Z. Hu, M. Wang, X. Yan, Y. Yin, and Z. Luo, "A comprehensive security architecture for SDN," in Intelligence in Next Generation Networks (ICIN), 2015 18th International Conference on, 2015, pp. 30-37.

[20] A. Zaalouk, R. Khondoker, R. Marx, and K. Bayarou, "Orchsec: An orchestrator-based architecture for enhancing network-security using network monitoring and sdn control functions," in 2014 IEEE Network Operations and Management Symposium (NOMS), 2014, pp. 1-9.

RETURN ORIENTED OBFUSCATION

Vivek Balachandran[1], Sabu Emmanuel[2] and Ng Wee Keong[3]

[1]Singapore Institute of Technology, Singapore
vivek.b@singaporetech.edu.sg
[2]Kuwait University, Kuwait
sabu@cs.ku.edu.kw
[3]Nanyang Technological University, Singapore
wkn@pmail.ntu.edu.sg

ABSTRACT

Software reverse engineering is an active threat against software programs. One of the popular techniques used to make software reverse engineering harder is obfuscation. Among various control flow obfuscations methods proposed in the last decade there is a lack of inter-functional control flow obfuscation techniques. In this paper we propose an inter-functional control flow obfuscation by manipulating return instructions. In our proposed method each function is split into different units, with each unit ending with a return instruction. The linear order in which functions appear in the program is obscured by shuffling these units there by creating an inter-functional control flow obfuscation. Experimental results show that the algorithm performs well against automated reverse engineering attacks.

KEYWORDS

Software protection, Code obfuscation, Reverse engineering

1. INTRODUCTION

To develop high quality software, engineers use various software analysing tools to detect vulnerabilities and loopholes in the program thereby facilitating them with an environment to improve their software. However, software analysing tools are double-edged swords that can be used to reverse engineer the software for malicious intents like intellectual property theft or finding vulnerabilities to exploit. Tools and books on reverse engineering are readily available for download on various Internet websites [1, 2].

A major factor that makes it harder to prevent software reverse engineering is that the attacker is a user and has all the power of a user to control the software and its running environment. One of the ways to provide some security to the distributed program is to incorporate a security mechanism embedded within the program. Software obfuscation is one such effective mechanism that hinders the process of software reverse engineering. Obfuscation is the process of translating a software into a semantically equivalent obscure form, so that it is harder to understand the logic of the program. Obfuscation can be applied to an entire program or partly to a section of the program, like watermarked code [4]. Low performance overhead compared to other techniques like encryption, is one of the desirable properties of obfuscation [5]

Software obfuscation can be applied to a program at different stages of its compilation. Source code obfuscation refers to the application of obfuscation on the source code of the program [6-8]. Similarly binary level obfuscation refers to applying obfuscation algorithms on compiled binary

Natarajan Meghanathan et al. (Eds) : NETCOM, NCS, WiMoNe, GRAPH-HOC, SPM, CSEIT - 2016
pp. 25– 36, 2016. © CS & IT-CSCP 2016 DOI : 10.5121/csit.2016.61503

programs [10]. Obfuscation can be applied on various intermediate levels such as on bytecode representation in the case of Android applications [19] or Java programs.

Most of the modern reverse engineering tools, like IDAPro [1], are capable of constructing the control flow graph of a program by converting a binary program to its equivalent assembly representation.

A control flow graph shows the basic block [16] of instructions as vertices and the possible control flow directions as edges, which enables an attacker to follow the program logic and find possible points to attack. Thwarting the disassembly process, by not allowing the reverse engineering tools to determine the correct program representation will result in an erroneous assembly program generation, thereby making program analysis harder. This is the basic idea of most of the binary obfuscation algorithms. In the past years, many binary obfuscation algorithms have been designed to fool the reverse engineering tools. Signal based obfuscation [11], control flow flattening [18], self-modification based obfuscation [14], double process obfuscation [12], instruction embedding [13], are some of the binary level obfuscation algorithms.

One of the limitations of all these obfuscation techniques is that they are all trying to obfuscate the instructions within a function. So, even though the obfuscation does a good job in obscuring the program, the functions remain intact. A reverse engineering tool will still be able to find the number of functions in the program and will be able to differentiate the instructions of one function from the other.

In this paper we discuss an obfuscation technique where, we shuffle code fragments from different functions disturbing the linear order of functions in the program. The reverse engineering tool will identify more functions than the original program and each function will be a small code fragment of the original function.

The paper is organized as follows. The proposed algorithm is explained in section 2. Section 3 discusses about the implementation details of the obfuscation method. In section 4, we analyses the overhead created by our obfuscation on the program performance. Performance evaluation of our obfuscated algorithm is discussed in section 5. The paper concludes with section 6.

2. PROPOSED METHOD

In this section, our new obfuscation method against software reverse engineering is discussed. Our obfuscation algorithm takes an assembly program as input and split the functions in the program and shuffles them, while maintaining the semantics of the program. The assembly representation generated by any assembler maintains a functional structure of the program i.e., the functions in the program are spatially arranged one after another. Each function starts with the standard set of instructions, to set the stack, and ends with a return instruction(s). When a reverse engineering tool disassembles a binary program to an assembly representation, it is thus capable of identifying functions and could segregate them into different functional units. This can help the reverse engineer, better analyse the program or creating a function call graph.

The basic idea of our technique is to disturb this normal representation of the program. In our technique, a function is split into various code fragments by inserting return instruction at the end of each code fragment. Each of these code fragments are then shuffled between functions, giving a inter-functional mix as shown in Fig. 1. One of the advantages of this method is that the linear arrangement of functions (one after another) is obscured. One of the challenges in implementing such a technique is to maintain the semantics of the program. In a normal program, a return instruction is used to return the control flow from a *callee* function to a *caller* function. Adding new return instructions could thus affect the behaviour of a function. In this section, we explain in detail about inserting the return instructions into functions while maintaining the semantics of the program.

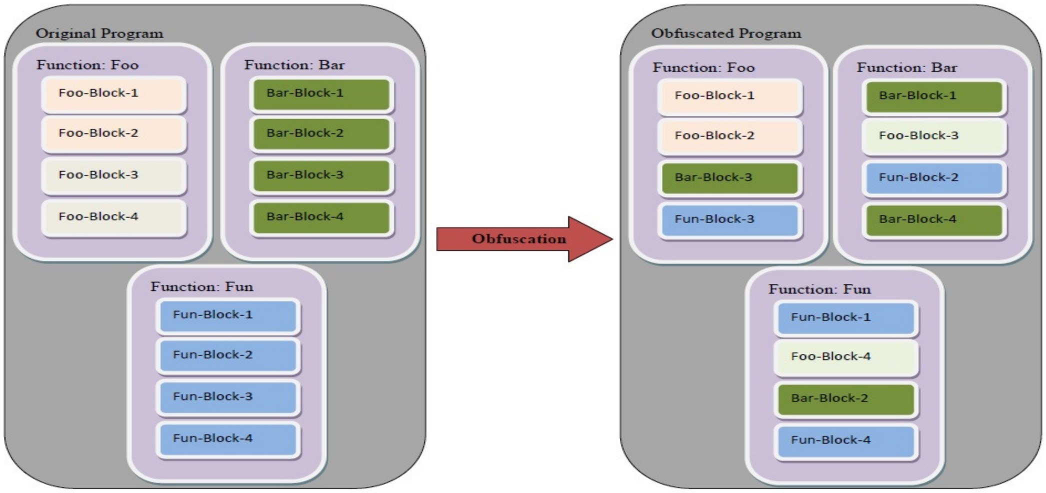
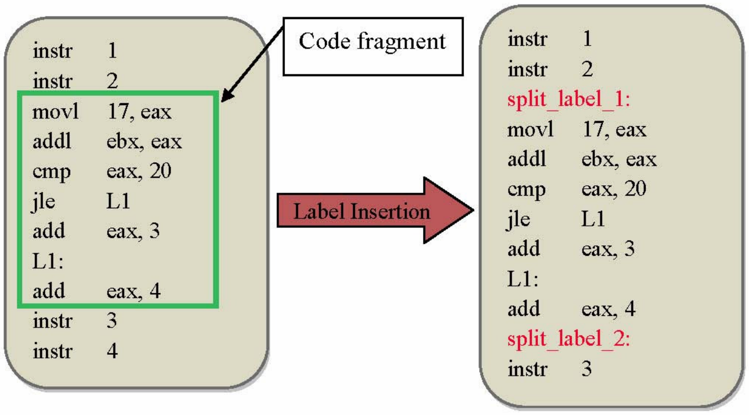


Figure 1 Overview of the algorithm

2.1. Splitting the function

The first step of our algorithm is splitting the function into different segments. The input assembly program is scanned for finding all the functions in the program. Once the functions are identified, each function is split into different code fragments. The obfuscator has the option to specify the number of splits in the function. In the default mode the obfuscator splits each function into four code fragments. While splitting the function into code fragments, our implementation put a constraint that the code fragment should contain at least five instructions.

For each function, the line numbers at which the function has to be split are identified and a randomly generated unique label is inserted. The unique label refers to the entry point of a code fragment. Fig. 2, shows the insertion of the labels to split the function into different segments.
In the example shown, two labels are inserted at the beginning of the code fragments.


```
instr    1
instr    2
movl     17, eax
addl     ebx, eax
cmp      eax, 20
jle      L1
add      eax, 3
L1:
add      eax, 4
instr    3
instr    4
```

```
instr    1
instr    2
split_label_1:
movl     17, eax
addl     ebx, eax
cmp      eax, 20
jle      L1
add      eax, 3
L1:
add      eax, 4
split_label_2:
instr    3
```

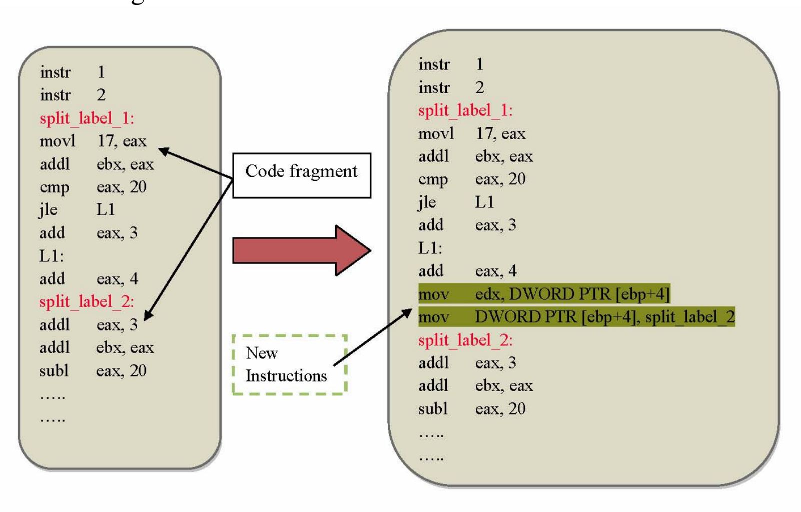
Figure 2 Inserting labels at the splits

2.2. Pushing the return address to the stack

We insert a *ret* (return) instruction at the end of each of the code fragment. This will make a disassembler think that each code fragment is a separate function. The *ret* instruction takes the return address from the stack and transfers the execution control to the particular address location. Thus, to maintain semantics, the current return address should be saved in the stack for later use and the address location of the next code fragment should be pushed into the stack as the new return address.

In our obfuscation algorithm, at the end of each code fragment two assembly instructions are added before inserting the *ret* instruction. One instruction stores the original return address in *ebp + 4* to a register that has not been used. It is followed by another instruction which stores the address of the next code fragment to the stack location *ebp + 4*.

In the example shown in Fig. 2, register edx is used to store the current return address in the stack. The address location *split_label_2* is then stored in the stack location *ebp + 4*. These two instructions can be stored anywhere between *split_label_1* and *split_label_2* and not necessarily at the end of the code fragment.



```
instr   1
instr   2
split_label_1:
movl    17, eax
addl    ebx, eax
cmp     eax, 20
jle     L1
add     eax, 3
L1:
add     eax, 4
split_label_2:
addl    eax, 3
addl    ebx, eax
subl    eax, 20
.....
.....
```

```
instr   1
instr   2
split_label_1:
movl    17, eax
addl    ebx, eax
cmp     eax, 20
jle     L1
add     eax, 3
L1:
add     eax, 4
mov     edx, DWORD PTR [ebp+4]
mov     DWORD PTR [ebp+4], split_label_2
split_label_2:
addl    eax, 3
addl    ebx, eax
subl    eax, 20
.....
.....
```

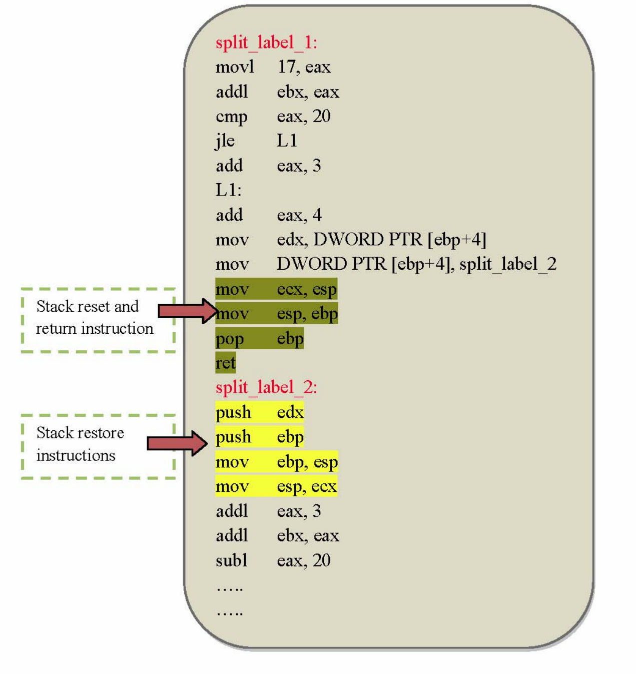
Figure 3 Inserting return address in the stack

2.3. Inserting return instruction

The next step in our technique is to insert the ret instructions in each of the code fragments. Like a standard return instruction the stack pointer and base pointer are reset using the two instructions, *move sp, ebp* and *pop ebp* which is then followed by the ret instruction. We add an extra instruction to store the stack pointer value to a free register. In the example shown in Fig. 4, the instruction is *mov ecx, esp*, is used to store the value of stack pointer to the register *ecx*.

We add an extra instruction to store the stack pointer value to a free register. In the example shown in Fig. 4, the instruction is *mov ecx, esp*, is used to store the value of stack pointer to the register *ecx.* The reason for this instruction is that we cannot reset the stack pointer value as the function is not completely returning to its caller function and it is needed in the following code fragments that will get executed.

With the address location *split_label_2* in stack and return address, the control, flows from the first code fragment to *split_label_2*, the beginning of the second code fragment when the *ret* instruction gets executed.



```
split_label_1:
movl    17, eax
addl    ebx, eax
cmp     eax, 20
jle     L1
add     eax, 3
L1:
add     eax, 4
mov     edx, DWORD PTR [ebp+4]
mov     DWORD PTR [ebp+4], split_label_2
mov     ecx, esp
mov     esp, ebp
pop     ebp
ret
split_label_2:
push    edx
push    ebp
mov     ebp, esp
mov     esp, ecx
addl    eax, 3
addl    ebx, eax
subl    eax, 20
…..
…..
```

Stack reset and return instruction
Stack restore instructions

Figure 4 Inserting return instruction

2.4. Restoring the stack

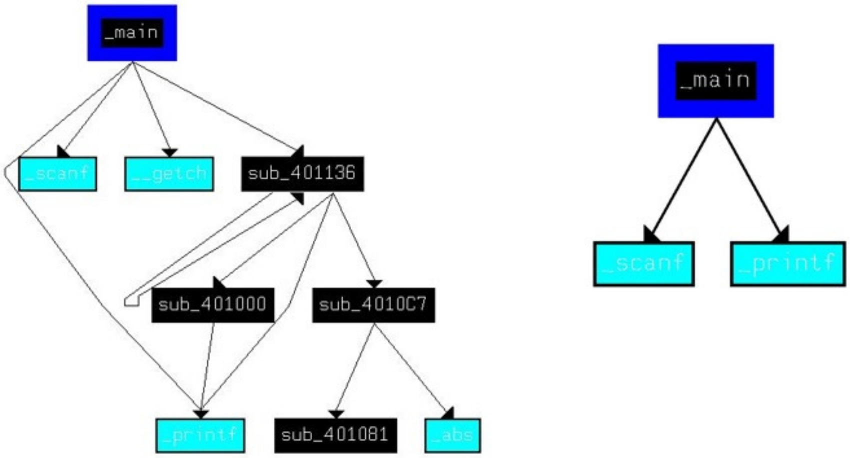
During the execution, after *ret* instruction from one code fragment is executed; the program execution control reaches the next code fragment. Since the stack pointer was reset during the return instruction, the stack has to be restored at the beginning of the new code fragment.

The first address that has to be restored in the stack is the original return address, which is stored in the register *edx*. By pushing the register *edx,* we can restore the original return address in the stack. Instructions *push ebp*, and *mov ebp, esp* restores the *ebp* register. The original stack pointer value is stored in ecx register as shown in the example in Fig. 4. Instruction *mov esp, ecx*, restores the original stack pointer value.

2.5. Shuffling the code fragments

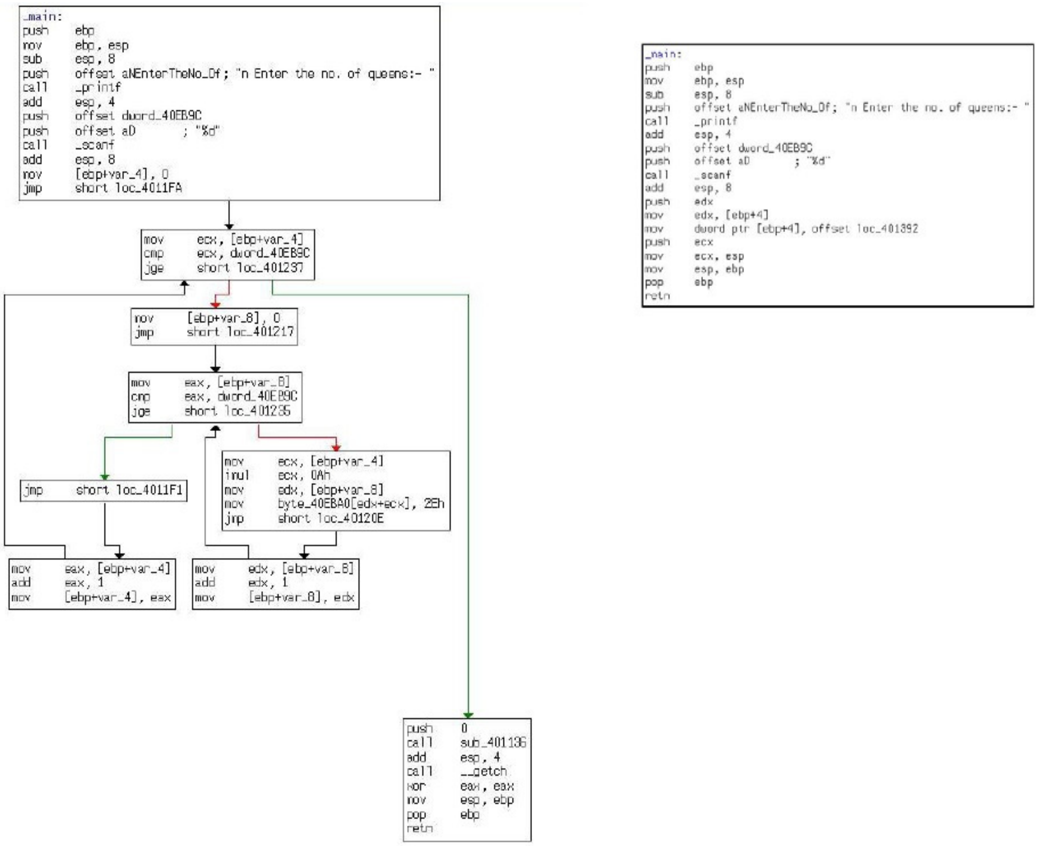
The obfuscation algorithm treats each code fragment as a separate function unit and shuffles them randomly. The linear order of the function representation is disturbed and code fragments from different functions will be interleaved together. This helps in inter-functional control flow obfuscation.

Fig. 5 shows the function call graph of *nqueens* program generated by IDAPro, before and after obfuscation. The obfuscation has clearly confused the IDAPro that it is unable to generate the function calls from *main* after obfuscation. Fig. 6 shows the disassembled *main* function of the program before and after obfuscation. It is clear from the figure that the control flow of the function is completely obscured by the obfuscation.



Before Obfuscation After Obfuscation

Figure 5 Function call graph of nqueens program



```
_main:
push    ebp
mov     ebp, esp
sub     esp, 8
push    offset aNEnterTheNo_Of; "n Enter the no. of queens:- "
call    _printf
add     esp, 4
push    offset dword_40EB9C
push    offset aD       ; "%d"
call    _scanf
add     esp, 8
mov     [ebp+var_4], 0
jmp     short loc_4011FA
```

```
mov     ecx, [ebp+var_4]
cmp     ecx, dword_40EB9C
jge     short loc_401237
```

```
mov     [ebp+var_8], 0
jmp     short loc_401217
```

```
mov     eax, [ebp+var_8]
cmp     eax, dword_40EB9C
jge     short loc_401235
```

```
jmp     short loc_4011F1
```

```
mov     ecx, [ebp+var_4]
imul    ecx, 0Ah
mov     edx, [ebp+var_8]
mov     byte_40EBA0[edx+ecx], 2Eh
jmp     short loc_40120E
```

```
mov     eax, [ebp+var_4]
add     eax, 1
mov     [ebp+var_4], eax
```

```
mov     edx, [ebp+var_8]
add     edx, 1
mov     [ebp+var_8], edx
```

```
push    0
call    sub_401136
add     esp, 4
call    __getch
xor     eax, eax
mov     esp, ebp
pop     ebp
retn
```

```
_main:
push    ebp
mov     ebp, esp
sub     esp, 8
push    offset aNEnterTheNo_Of; "n Enter the no. of queens:- "
call    _printf
add     esp, 4
push    offset dword_40EB9C
push    offset aD       ; "%d"
call    _scanf
add     esp, 8
push    edx
mov     edx, [ebp+4]
mov     dword ptr [ebp+4], offset loc_401392
push    ecx
mov     ecx, esp
mov     esp, ebp
pop     ebp
retn
```

Before Obfuscation After Obfuscation

Figure 6 Main function of nqueens program

3. IMPLEMENTATION

The proposed obfuscation method is implemented in python programming language. Our implementation expects assembly level representation of the program to be obfuscated. We have implemented the obfuscation algorithm for Microsoft Windows XP and Ubuntu Linux 13.04 operating systems. Our implementation accepts Microsoft Visual Studio 10.0 generated assembly programs and assembly program generated by gcc 4.7.3 as input for obfuscation. Our algorithm generates the obfuscated assembly program which is assembled using the corresponding assembly program to generate obfuscated binary program.

The python code copies the input assembly program to a buffer. It analyses the buffer to find the functions and the start and end of the functions. The number of code fragments for each function is calculated according to the size of the function. The instructions to modify the stack for return address and restoring the stack pointer are added to the beginning and end of each code fragments. The line numbers of beginning and end of each code fragments are changed due to the insertion of instructions. The code fragments are given numbers in sequential order and are represented in a data structure with the number, starting line in the buffer and ending line in the buffer. A simple shuffling algorithm is used to shuffle the code fragments as shown in the following psuedocode. The code fragments are then stored into a new file in the shuffled order to generate the obfuscated assembly file.

```
Shuffle_code_fragments (Code_Fragment_list [])
L = Length (Code_Fragment_list)
        while  (L > 1)
                R = Random (1, L-1)
                Temp = Code_Fragment_List [R]
                Code_Fragment_List [R] = Code_Fragment_List [L]
                Code_Fragment_List [L] = Temp
                L = L -1
Return Code_Fragment_List
```

4. PERFORMANCE EVALUATION

In this section we perform experimental evaluation of our algorithm against reverse engineering. We measure the efficacy of our algorithm by measuring the potency against IDAPro [1]. Instruction disassembly error which calculates the number of instructions that the reverse engineering tool is unable to disassemble properly gives the potency of the obfuscation algorithm. In this section, we also analyse the space and time overhead caused by the obfuscation. The increase in space and time at different levels of obfuscation is analysed. We used the test programs from the lcc 4.2 [17] compiler source as input test programs for our obfuscation algorithm.

4.1. Instruction disassembly error

The potency of the obfuscation algorithm against reverse engineering tool, is measured by the error in the disassembly of assembly instructions. IDAPro [1] was used to disassemble the obfuscated test programs. We measured the total number of instructions in the original program and the instructions recognized by IDAPro [1] after reverse engineering the obfuscated program. Confusion factor is then calculated as the ratio of their differences, as defined by the following equation,

$$CF_{instr} = |T_{total} - T_{disasm}| / T_{total}$$

The total number of instruction addresses before obfuscation is represented by T_{total}. The number of instruction addresses recognized by IDAPro [1] after disassembling the obfuscated binary program is represented by T_{disasm}.

Table 1 shows the confusion factor while disassembling obfuscated test pro-grams by IDAPro [1]. The table shows varies levels of splitting the program. The first column represented by zero splits is the original program and all the instructions are reverse engineered successfully. Column 2 represents the obfuscated program, where every function is split into two and the mean disassembly error is 55.9% when functions are split into two. The instruction disassembly error increases as the splitting of the program increases. The splitting of a program saturates after a while. For instance, the program fields has the same instruction disassembly error for 16 splits and 32 splits. This is because the program is split to the maximum possible split by 16 splits and the program cannot be further split down.

Mean instruction disassembly error of 85.16% is obtained at level 8, where each function is split into 8 code fragments.

Table 1. Instruction Disassembly Error

Splits **Prog**	**0**	**2**	**4**	**8**	**16**	**32**	**64**	**128**
8q	283	125	65	42	25	25	25	25
array	341	150	78	51	31	20	20	20
cf	184	81	42	28	17	11	11	11
cq	13786	6066	2068	2068	2068	2068	2068	2068
cvt	674	297	155	101	61	40	27	14
fields	339	149	78	51	20	20	20	20
incr	392	172	90	24	24	24	24	24
init	415	183	95	62	37	17	17	17
limits	162	71	37	24	12	12	12	12
sort	506	223	116	76	46	46	46	46
spill	433	191	100	65	39	26	26	26
struct	505	222	116	76	45	36	30	30
wf1	597	263	137	90	54	24	24	24
CF_{instr}		55.9%	82.9%	85.1%	86.6%	87.2%	87.3%	87.4%

4.2. Space overhead

The insertions of the new instructions have significant effect on the size of the program. If a function is split into 2 code fragments, then 10 new instructions are added into the program and 20 instructions are added if the function is split into 3 code fragments and so on. Let the number of instructions in the program be N_{before} and the original program is split into $n+1$ code fragments. The total number of instructions in the obfuscated program will be,

$$N_{after} = N_{before} + 10n$$

In the worst case, the entire program is split into code fragments with 5 instructions. Let the program is split into *n+1* code fragments, with each code fragment having four instructions, then the total number of instructions in the program before obfuscation is,

$$N_{before} = 5(n + 1)$$

After the obfuscation, 10 instructions are added per code fragment and the total number of instructions in the program after obfuscation is,

$$N_{after} = 5(n + 1) + 10n$$
$$N_{after} = 3N_{before} - 10$$

So, in the worst case there are three times more instructions in the obfuscated program than the original program. We can see in the experimental evaluation that this upper bound is held. $Space_{ovh}$ defines the increase in the size of the program.

$$Space_{ovh} = Space_{after} / Space_{before}$$

In Table 2, we show how the program size increases as the program is obfuscated. The size of the program increases as the number of splits increase. In the worst case, the size increases to 2.2 times the original size. But on an average, the program size increases by 1.57 times the original size with 128 splits, which is less than the theoretical upper bound.

Table 2 Space Overhead

Splits **Prog**	0	2	4	8	16	32	64	128
8q	8743	8779	10256	13586	17843	17843	17843	17843
array	8678	8714	11324	14321	18545	18923	18923	18923
cf	8750	8786	9957	12985	13876	14178	14178	14178
cq	63599	75898	85699	85699	85699	85699	85699	85699
cvt	12983	13019	13091	13235	13523	14099	15251	21651
fields	8728	8764	12633	13589	13589	13589	13589	13589
incr	8594	8630	9987	12371	12371	12371	12371	12371
init	9134	9170	9242	9386	9674	9782	9782	9782
limits	8522	8558	8630	8774	12062	13458	13458	13458
sort	8856	8882	11954	12098	14350	14350	14350	14350
spill	8863	8899	10971	12115	16403	18799	18799	18799
struct	8843	8879	8951	9095	9383	14055	17127	17127
Wf1	13179	13215	13287	17521	18754	21875	21875	21875
Mean	177462	190193	215982	234775	256072	269021	273245	279645
$Space_{ovh}$	1	1.07	1.21	1.32	1.44	1.51	1.53	1.57

4.2. Time overhead

Obfuscation does have an effect on the execution time. The execution time of the program will increase, because of the execution of the additional instructions. We know that the size of the program increases by 3 times in the worst case. The input parameter of the time complexity thus increases by 3 times. If the original time complexity was *T(n),* then the new time complexity will be *3T(n)* in the worst case. The time complexity of the obfuscated program is thus between *T(n)* and *3T(n).*

Time$_{before}$ refers to the time taken by the program to execute without obfuscation and Time$_{after}$ is the time taken by the obfuscated program to execute. We evaluate the effect of obfuscation on execution speed with Time$_{ovh}$ defined as,

$$Time_{ovh} = Time_{after} / Time_{before}$$

Table 3, shows the time overhead caused by our obfuscation on various binary programs. In the worst case the time overhead is 2.36 times in the case of *cvt* with 128 levels of obfuscation. On an average the worst case time overhead is 1.86 which is lower than the upper bound of *3T(N).*

Table 3 Time Overhead

Splits **Prog**	**0**	**2**	**4**	**8**	**16**	**32**	**64**	**128**
8q	1300	1395	1534	1975	2589	2589	2589	2589
cf	2777	3276	3588	4215	5014	5581	5581	5581
cvt	1077	1121	1154	1178	1223	1498	1892	2548
fields	1089	1258	1685	1987	2406	2406	2406	2406
incr	1271	1301	1354	1537	1537	1537	1537	1537
spill	1174	1256	1325	1456	1658	2219	2219	2219
struct	1342	1398	1469	1566	1689	1726	1754	1754
wf1	1245	1322	1391	1499	1785	1997	2015	2015
array	1011	1250	1347	1678	2001	2022	2500	2694
cq	1025	1119	1243	1567	1874	2198	2198	2198
init	1198	1245	1376	1461	1653	1985	1985	1985
sort	1037	1134	1256	1370	1523	1523	1523	1523
Mean	15546	17075	18722	21489	24952	27281	28199	29049
Space$_{ovh}$	1	1.09	1.20	1.38	1.60	1.75	1.84	1.86

5. CONCLUSIONS

In this paper we proposed an obfuscation algorithm to perform inter functional obfuscation. Our method slices each function in the program into separate code fragments. Each fragment ends with a return instruction and starts with stack allocation instructions, thereby appearing itself like a function. The return instruction transfers the control flow to the next code fragment instead of returning to a caller function. The code fragments are shuffled disturbing the linear order of the functions. Unlike other control flow obfuscation, our method adds more control flow instructions (return instruction) to increase the control flow obscurity instead of removing the control flow instructions. The experimental results show that obfuscating with 8 splits provides a good obfuscation without too much overhead on the space and time requirements of the program. Experimental analysis shows that that our method has an instruction disassembly error of 85.1 % with 8 levels of splitting. An average time overhead of 1.38 and space overhead of 1.32 are observed while obfuscating with 8 splits.

REFERENCES

[1] "Data Rescue," Available at: http://www.hex-rays.com/ [Last accessed: October 14, 2016]

[2] J. Miecznikowski and L. Hendren, "Decompiling java using staged encapsulation," in Reverse Engineering, 2001, pp. 368-374.

[3] W. Thompson, A. Yasinsac, and J. McDonald, "Semantic encryption transformation scheme," in International Workshop on Security in Parallel and Distributed Systems, San Francisco, CA, 2004.

[4] J. Hamilton and S. Danicic, A survey of static software watermarking, in World Congress on Internet Security, pp.100-107, 2011.Lecture Notes in Computer Science: Authors' Instructions 13

[5] B. Anckaert, M. Madou, B. De Sutter, B. De Bus, K. De Bosschere, and B. Preneel. Program obfuscation: a quantitative approach, in ACM Workshop on Quality of Protection, ACM New York, NY, USA, 2007, pp. 15-20.

[6] W. F. Zhu, "Concepts and techniques in software watermarking and obfuscation," Ph.D. Thesis, University of Auckland, 2007.

[7] M. Sosonkin, G. Naumovich, and N. Memon, "Obfuscation of design intent in object oriented applications," in ACM workshop on Digital Rights Management, 2003, pp.142-153.

[8] C. Collberg and J. Nagra. "Surreptitious Software: Obfuscation, Watermarking, and Tamperproofing for Software Protection (1st ed.)," Addison-Wesley Professional, 2009.

[9] C. Collberg, C. Thomborson, and D. Low, "Manufacturing cheap, resilient, and stealthy opaque constructs," in ACM Symposium on Principles of Programming Languages, 1998, vol. 25, pp. 184-196.

[10] V. Balachandran and S. Emmanuel,Software Protection with Obfuscation and Encryption, Information Security Practices and Experiences Conference, ISPEC, volume 7863 of Lecture Notes in Computer Science, pp 309-320. Springer, 2013.

[11] I. V. Popov, S. K. Debray, and G. R. Andrews, "Binary obfuscation using signals," in USENIX Security Symposium, 2007, pp. 1-16.

[12] J. Ge and S. Chaudhari, "Control flow based obfuscation," in ACM Digital Rights Management, 2005.

[13] C. LeDoux, M. Sharkey, B. Primeaux, and C. Miles "Instruction Embedding for Improved Obfuscation," in Proceedings of the 50th Annual Southeast Regional Conference (ACM-SE '12),pp.130-135, 2012.

[14] V. Balachandran and S. Emmanuel, "Software code obfuscation by hiding control flow information in stack," in IEEE Workshop on Information Forensics and Security, 2011.

[15] V. Balachandran and S. Emmanuel,"Potent and Stealthy Control Flow Obfuscation by Stack Based Self-Modifying Code," in IEEE Transactions on Information Forensics and Security, vol.8, no.4, pp.669-681, April 2013.

[16] "Basic Block," Available at: http://gcc.gnu.org/onlinedocs/gccint/Basic-Blocks.html [Last accessed: October 14, 2016]

[17] C. W. Fraser, "lcc, A Retargetable Compiler for ANSI C," Available at: https://sites.google.com/site/lccretargetablecompiler/downloads [Last accessed: October 14, 2016]

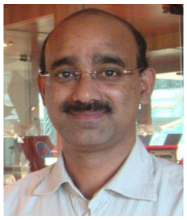
[18] L. Shan and S. Emmanuel, "Mobile agent protection with self-modifying code," in Journal of Signal Processing Systems, vol. 65, pp. 105-116, 2010.

[19] V. Balachandran, Sufatrio, D.J.J. Tan, and V.L.L. Thing. "Control Flow Obfuscation for Android Applications." in Computers and Security, vol.61,pp.72-93, Aug 2016.

AUTHORS

Dr. Vivek Balachandran holds a PhD in Computer Engineering from the Nanyang Technological University, Singapore, where he worked with the Center for Strategic Infocomm Technologies and Temasek Laboratories. After graduating he worked as a Research Scientist with Institute for Infocomm Research, A*STAR, Singapore, where he worked on mobile security and forensics. He has published numerous articles on software security. He is currently a Lecturer at the Singapore Institute of Technology, Singapore. His research interests includes software security, mobile forensics, program analysis, and compiler optimization.

Dr. Wee Keong Ng is Associate Chair (Research) of the School of Computer Science & Engineering, Nanyang Technological University, Singapore. He received his Ph.D. from the University of Michigan at Ann Arbor, USA. His research areas are secure data analytics, secure data storage, data monetization, and data security, and has published more than 200 technical papers in these areas. Dr. Ng has served in program/organizing committees of international conferences. In recent years, he is General Co-chair of the International Conference on Information and Communications Security 2016; Jury Member of the Second Dutch Cyber Security Research Award in March 2016; Senior PC Member of the 20th, 19th, 18th, 17th Pacific-Asia Conference on Knowledge Discovery and Data Mining. Dr. Ng has advised and graduated more than 20 Ph.D. students and 20 Master students



Dr. Sabu Emmanuel received the B.E. degree in electronics and communication engineering from Regional Engineering College, Durgapur, India, in 1988, the M.E. degree in electrical communication engineering from the Indian Institute of Science, Bangalore, in 1998, and the Ph.D. degree in computer science from the National University of Singapore in 2002. He is currently an associate professor with Kuwait University. Previously, he was an assistant professor in the School of Computer Engineering, Nanyang Technological University (NTU), Singapore. He began his career as a research and development engineer and then moved on to the academic profession. He has taught engineering students of Mangalore University, National University of Singapore, and NTU. His current research interests are in multimedia and software security and surveillance media processing. He has been a reviewer for ACM Multimedia Systems. He is a member of the Technical Program Committee of several conferences. Dr. Emmanuel has been a reviewer for IEEE Transactions on Circuits and Systems for Video Technology, IEEE Transactions on Multimedia, and IEEE Transactions on Information Forensics and Security.

RITA SECURE COMMUNICATION PROTOCOL: APPLICATION TO SCADA

Fadi Obeid[1] and Philippe Dhaussy[2]

[1]PhD. at Ensta Bretagne, Brest, France
fadi.obeid@ensta-bretagne.org
[2]HDr. at Ensta Bretagne, Brest, France
philippe.dhaussy@ensta-bretagne.fr

ABSTRACT

Supervisory control and data acquisition (SCADA) systems have their own constrains and specifications. These systems control many of our critical industrial infrastructures, yet they are hardly secured. The biggest problem in securing these systems is the lack of cryptography support especially that most SCADA systems work in real-time which is not compatible with most cryptography algorithms. Additionally, a SCADA network may include a huge amount of embedded devices with little computational powers which adds to the cost of any security improvement. In this paper we present a new approach that would secure SCADA communications by coding information without the need of the complex cryptography algorithms. The reconfigurable information transmitter agent (RITA) protocol that we present does not need the already installed devices to be modified nor replaced, it only needs to add costless electrical chips to these devices. This approach can also be used to secure any type of communication that respects the protocol's constraints.

KEYWORDS

Information Security, Network and Communication Security, SCADA Networks, Cryptography

1. INTRODUCTION

Information Security (*InfoSec)* is the act of protecting a set of information against unauthorized entities. A complete protection means that the information can be created, manipulated, and red by authorized entities only. The measurements to ensure InfoSec depend a lot on the state of the information, whether the information is stocked in a data base, being processed, or being communicated between entities. In this article we are interested in the measurements taken to secure the communication of information.

Securing a communication is needed when the channel being used for communication is considered insecure as in the case of Internet communication, wireless communication and others. In most cases, InfoSec relies on cryptography algorithms to ensure a secure communication. The complexity of a cryptographic algorithm requires additional power, time and space. Although good cryptography algorithms exist, few are considered secured today. An additional problem with cryptographic algorithms is that many were considered to be secure until attacks and analysis proved they are not, which means that what is considered secure today may be insecure tomorrow.

Natarajan Meghanathan et al. (Eds) : NETCOM, NCS, WiMoNe, GRAPH-HOC, SPM, CSEIT - 2016
pp. 37– 48, 2016. © CS & IT-CSCP 2016 DOI : 10.5121/csit.2016.61504

SCADA systems are considered secured by isolation, still, they can be attacked from the inside. Also, due to the augmented connectivity to the outside, it is wise to consider effective security measurements before being able to have authorized outside access.

While security protocols are implemented in many systems, most of SCADA systems are still unsecured. Most companies that rely on SCADA systems do not consider securing these systems because of the expected high costs. This high cost is the consequence of cryptography use which also breaks the real-time constraint of SCADA systems.

Our proposal is to replace cryptography with a measurement that is expected to have a satisfying security level with a very low cost (power, time, and space). Our approach does not need for the already installed system to be replaced nor upgraded which means that the SCADA system would be available during the shift from unsecured to secured.

In this article, we will present the concept of our proposal while unfolding the first and most basic version of our protocol.

2. SCADA

SCADA systems can be found in modern industrial facilities such as water pipes, power plants, oil refineries, chemical factories and nuclear facilities. These systems use coded signals over communicating channels to monitor and control numerous devices on multiple and distant sites.

Unlike standard networks, most of the SCADA nodes are special purpose embedded computing devices with limited capacities such as remote terminal units (RTUs) and programmable logic controllers (PLCs). These nodes exchange data (exp. temperature is *x*, water level is *y*, etc.) and commands (exp. turn off water) between each others and with the supervisory system. The supervisory system can also build statistics about the system and how it is being used based on the received data.

Figure 1 presents a general SCADA network and its communication to a second SCADA, a local network, and the Internet.

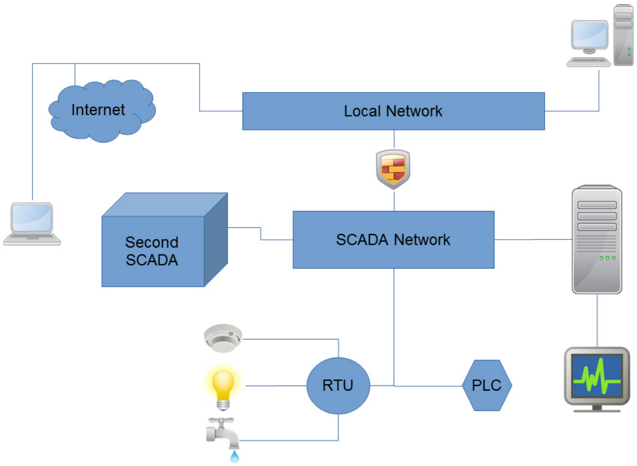


Figure 1. General SCADA network presentation

In addition to using special purpose embedded computing devices, other SCADA properties also affect their security as well, we are mainly interested by the following [1]:

- Non-stop availability: Devices are required to work non-stop for years, exp. traffic lights.

- Geolocation: Nodes can be very sparse and geographically extensive, exp. water pipelines.

- Hard conditions: Nodes may exist in hard physical conditions, exp. chemical factories.

- Performance: Devices must have a hard real-time constraint, exp. microchips industries.

Many security problems are caused by the properties mentioned above: The use of special purpose devices with limited input/output choices causes communicated messages to be easily predictable. Cryptography is hardly supported since performance would be dramatically reduced. Renewing and updating devices would be very expensive, this is caused by the geolocation of the devices and the availability constraint. The geolocation constraint also allows possible unexpected links to the outside reducing system security. And finally, the geolocation and hard conditions of devices discourage temper resistance. Any approach to secure SCADA systems should consider those properties and specifications.

SCADA properties are not the only aspects causing security problems, many of the choices (made by SCADA manufacturers and users) that characterize current SCADA systems also reduce SCADA security [1]:

- Using open standards which grant attackers more knowledge of the system.

- Using COTS (commercial off-the-shelf) hardware and software (which lacks of security).

- Using fail-safe constraints increases safety while decreasing security.

- Protocols vulnerabilities, whether conceptual or caused by implementation errors.

3. RELATED WORK

Many efforts were put to secure SCADA communication [2,3,4,5]. In the technical and research world, these efforts led to solutions which would insure a high level of security in SCADA. In practice, the proposed solutions are expensive and their requirements are not met in the SCADA networks. Since none of these solutions is proven to be perfect, no one would take the risk and pay the elevated price, which is in some cases changing their whole SCADA system. The imperfection of these solutions would mean constant updates and upgrades that SCADA managers would not risk. For these reasons, we think that any SCADA security solution that desires to pass from the research world to practice should have perfection properties. Also a good security solution for SCADA systems would not need to replace the already installed materials.

The only known perfectly secure cryptosystem is the Vernam cipher, also called the one-time pad. Gilbert Vernam patented this invention in the USA in July 1919 [6]. A few years later, a variation of the one-time pad was patented in Germany by Siemens and Halske [7]. The one-time pad is based on a list of shared keys that can only be used once. If implemented in SCADA systems, the list of shared keys need to be updated constantly which is very time consuming in most cases (exp. water pipes).

In 1949, Claude Shannon proved that the one-time pad is indeed unbreakable and that any unbreakable system must have the same essential characteristics as the one-time pad [8]. The most essential characteristic of the one time-pad is the use of a different key for each encryption.

To the best of our knowledge, there are no other cryptosystems that were considered unbreakable since the one-time pad.

4. PROTOCOL

In this section we will explain the principles of the protocol and its characteristics. We will also demonstrate the communication schema, and finally we show some details of the needed security measurements for the protocol to work effectively.

4.1. Principles of the Protocol

The most important principle of our protocol is the secret sharing between two entities that we call security boxes. These security boxes are considered twins, our security relies on the possibility for these twins to share and maintain a secret which is similar to a symmetric key in the case of cryptographic algorithms.

The shared secret is a table of randomly pre-initialized values, along with a secret algorithm with predefined random operations. To make sure the performance is almost intact, the chosen operations are simple binary compositions (exp. binary XOR) and substitutions. Also, most of the operations take place after sending/receiving a message and not before. That way, the message is sent and processed with almost no security related latency.

The 2-security boxes are initialized together before being placed each on the entry/outlet of any type of device that needs its communication secured. A security box can be an embedded device or an algorithm implemented on another already existent device such as a computer program, a smart phone application, etc. Any communication between the 2 devices would go through both security boxes, the first security box translates the communicated message to a matching secret message and the second security box would reveal the match for the received secret. After sending or receiving a secret message, a security box would change the secret table.

Figure 2 represents a synchronous version of the communication using our security protocol. The twins *Agent A* and *Agent B* communicate synchronous messages. In our latest version of the protocol asynchronous communication is possible. Each time a security box sends or receives a message it changes the used value in the table using function *f*. Function *f*'s output depends not only on the input but on the current state of the security box (ex. whole table, number of communicated messages, etc.). This dependency exists for security reasons so outputs would not be redundant or have a pattern.

4.2. Characteristics

The security boxes transform a plain message to a secret message/code and vice versa to ensure that messages in the channel are unintelligible to any eavesdropper that may be analysing channel communication. Only intelligible messages are accepted by the security boxes.

The security boxes can be used as a middle-ware between 2 devices, the translation from plain to secret message and vice versa can be different from one side to the other to make sure both devices send and receive messages they understand.

The security boxes can also be used as a middle-ware between the devices and the communication channel, no matter what form of message the device produces. The security box transforms a produced message into a secret message that respects the protocol used in the communication channel without the need to change or adjust the secured device itself. The only modifications are the ones we do to the security box itself.

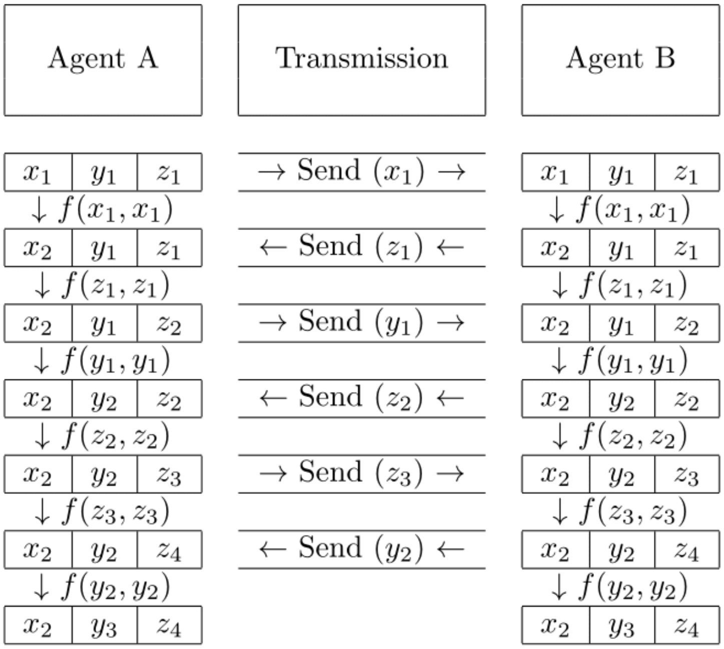


Figure 2. Simple Synchronous Method

A very important aspect of the security boxes is that they do not require any changes from the devices being secured. It is up to the security box to adjust itself depending on the secured devices. To do so, the security device can either be a general box which would require configuration depending on the secured devices. This would increase the flexibility of the box while reducing its performance and security level. The second method is to have boxes specially designed depending on the requirements of the secured devices. Although the second method provides less flexibility, it insures maximum performance and security. While the second method seems extreme with its need to recreate a security box depending on the requirements, it is feasible since the part that would change in the security box is very small and easy to modify.

4.3. Communication Schema

Figure 3 describes the communication schema using our protocol. *A* and *B* are communicating in a synchronous fashion. A creates a clear message *clm* and sends it to *A.mySecurityBox* which codes it into a matching coded message *com* and forwards it to the communication channel before updating the secret table using the contents of com and other variables. *B.mySecurityBox* is waiting for this message, it receives *com*, generates *clm* and sends it to *B* before updating its own version of the secret table using the same variables and operations. Finally *B* would answer by creating a new *clm* (response) and sending it to *A* in the same fashion.

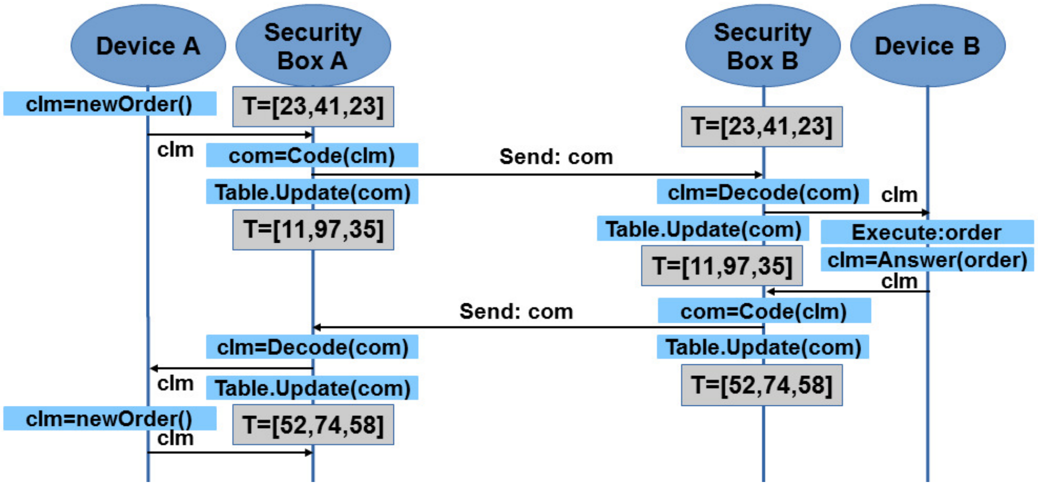


Figure 3. Communication Schema

4.4. Function Necessities and Security Measurements

The size of the secret table depends on the requirements of the secured devices and the different possible signals. For example, to control a lamp with *3* possible commands and *3* possible responses, we need a table of *3* values (for the most basic implementation of our mechanism).

The cryptosystem needs to insure confusion and diffusion [8]. To make sure that the minimum required confusion and diffusion is offered we use substitution boxes. It is also considered important that the transformation of the secret table is irreversible. This would reduce the possibilities of future analytical attacks.

It is clear that the values in the table should be different one from the other at any point of time or the same value would have multiple meanings which results in an ambiguous message.

A list of unacceptable/undesired values contains values considered to decrease the security of our system such as *zero* in addition to any value with Hamming weight equal to *1* or even *2*. these values are never used in the secret tables.

Getting a substitution box output should be normalized regarding time, power, etc. If we have a substitution box of the form *X = numberOfRows* and *Y = numberOfColumns*, and we are searching for the output *OUT(row=x,col=y)* then we should also search for a phantom output *OUT(row=X-x,col=Y-y)* without actually using it.
Any list/table search should consider the same approach. Looking for the value *x* in list *T* should have *phantom = T[l-i]* where *l* is the length of *T* and *i* is the index of *x* in *T.*

Although most phantom outputs are used after sending/receiving, they still affect the performance and power consumption of our approach. Therefore, phantom outputs should only be used when side channel attacks are considered as threats to the system.

5. Discussions

In this section we will analyse the robustness of our approach and show an effective way of implementing it on already installed systems.

5.1. Robustness Analysis

Consider the following: *A* and *B* communicating using our protocol, they use a table of 3 values. *A* sends a message with value X_1 to *B* and replaces X_1 with X_2, X_1 is intercepted by an attacker. From the attacker's (let us call him *C*) point of view, X_1 has no signification other than a strange signal being sent from *A* to *B*, since *C* cannot understand the meaning of this signal then the confidentiality requirement is respected.

The integrity requirement is also respected since if *C* tries to change X_1, or to invent a message and send it to *B*, it has a negligible chance of succeeding. The success probability is actually $a/(2^b – c)$ where a is the number of possible values (*3* in our example), b is the number of used bits *32*, and c is the number of unacceptable values *1+32* if we only refuse values with hamming weight equal to *0* and *1*. $3 / (2^{32} – 33) = 6 * 10^{-10}$ which is *6* times lesser then the probability of gaining the jackpot Mega Millions multi-state lottery in the United States.

If *C* tries to redirect X_1 to *A*, A would not accept it since X_1 was replaced by X_2 and does not have any meaning to *A* any longer. The only thing *C* can do is to interrupt messages from *A* to *B* and *B* to *A*. Interrupted messages cannot be replaced, therefore, if *A* and *B* use a time constraint (the system knows something wrong if no message is received for *t* seconds) then both devices would know there is an undergoing attack or a connection problem.

5.2. Installation Method

Let us consider devices *A* and *B* are already functioning in our system, we wish to secure the communication between these 2 devices. If we do not want to break the communication between the devices we proceed as follows:

First we add the security box to *A* by switching the cables connecting *A* to the security box, and adding a cable between *A* and its security box. The security box would simply forward messages from *A* without securing them. Then we add the security box to *B*, once installed the security box would send a notification to *A*'s security box to start securing messages. After the secured communication is well established, unsecured messages would stop being accepted.

While a security box is being installed on a machine, this machine will not be able to send and receive messages only for the instance of cables switching.
In addition to the security boxes, we use port boxes (Figure 4). The current ports would simply route the messages to the correct channel. The ports may also be used to scramble the messages which would add to the system security while slightly affecting performance.

The structure of a secret message would contain the following information: *ID* of the sender (*portID_A*, *boxID_A*, and *deviceID_A*), ID of the receiver (*portID_B*, *boxID_B*, and *deviceID_B*), sequences of the message (*box_seq* and *device_seq*), and finally the coded data itself. Additional information can be added to the message if needed. Some information can also be reduced if it is found heavy for a system. For example, we can remove *boxID_A* since it is *boxID_B*'s twin and the only one who could have sent the readable message. In some cases, a box and a device can have the same ID and messages sequences which would also reduce the size of the secret message.

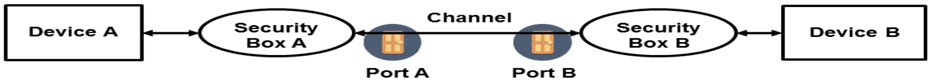


Figure 4. Using Port Boxes

6. General Structure

The connection between a device and a security box should not be exposed to other devices, entities, or the outside since it is unsecured. The connection between a port and a security box should be a single line or messages would be routed wrongly, messages passing this connection are already secured. The communication between ports can pass through any type of channels (wireless, Internet, etc.), messages are secured during this communication.

Our protocol can be used to support communication between multiple devices (Figure 5). Consider the following security boxes twins: *A* and *AA*, *B* and *BB*, *C* and *CC*. Devices *1* and *2* communicate with devices *4* and *5* through *A* and *AA*. Device *5* communicates also with device *3* using *2* twins: *B,BB* and *C,CC*. Device *3* and device *5* cannot communicate with each others.

If a message is sent from Device *3* to device *5* passes by the first twin *B,BB* and a second message passes by the second twin *C,CC*, there is a chance for the second message to arrive before the first one, therefore it is unsafe to have devices communicate with each others using multiple security boxes unless both devices *3* and *5* have the capacity of readjusting the order of received messages.

Finally, we have Port *1* communicating with ports *2* and *3*.

In conclusion we have the following:

- A security box can secure multiple applications/devices.
- A port can be connected to multiple security boxes and communicate with multiple ports.
- A device can communicate with multiple devices using the same security box.
- A device can communicate with a device using multiple security boxes if both devices have a measurement that keeps track of the correct messages order.

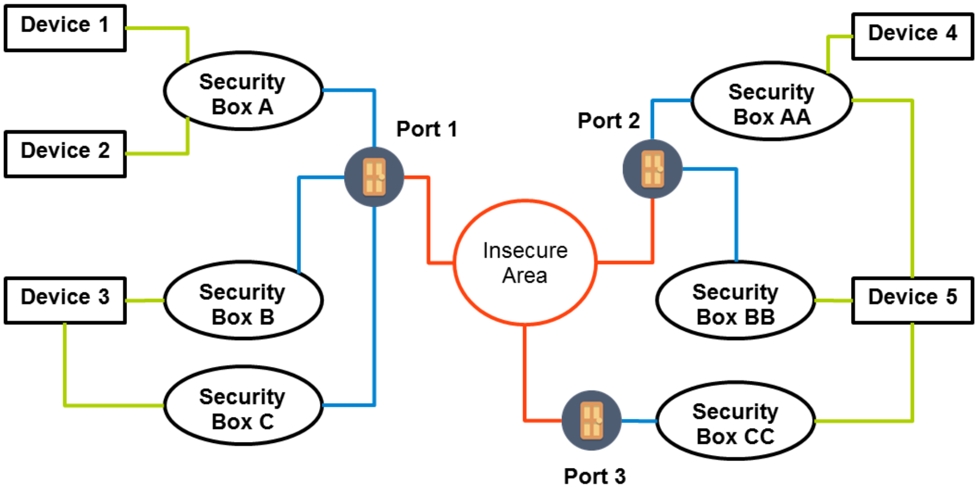


Figure 5. General Structure

7. Implementation

In this section we show an example of how the protocol can be used, and exhibit our simulation.

7.1. Example

Let us consider the following example that do not mention the use of port boxes (Figure 6).

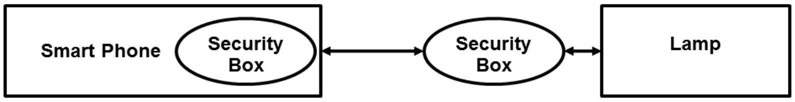


Figure 6. Lamp Control Example

We have a smart phone with many applications, one of which would control a lamp through a wireless communication channel. We need to make sure the communication between the lamp and the application is secured from any attack. The attacks may take advantage of the wireless insecurity or an insecure application installed on the smart phone. We create 2 security boxes: one would be an embedded device that is installed on the lamp's input/output, the second security box would be an algorithm on the smart phone and controls the input/output of the lamp control application.

Any communicated message between the lamp and the application would be translated into a secret message that has no meaning to any outsider (attacker) which guarantees confidentiality. Changing this message would result in a meaningless message which insures the integrity of the communication.

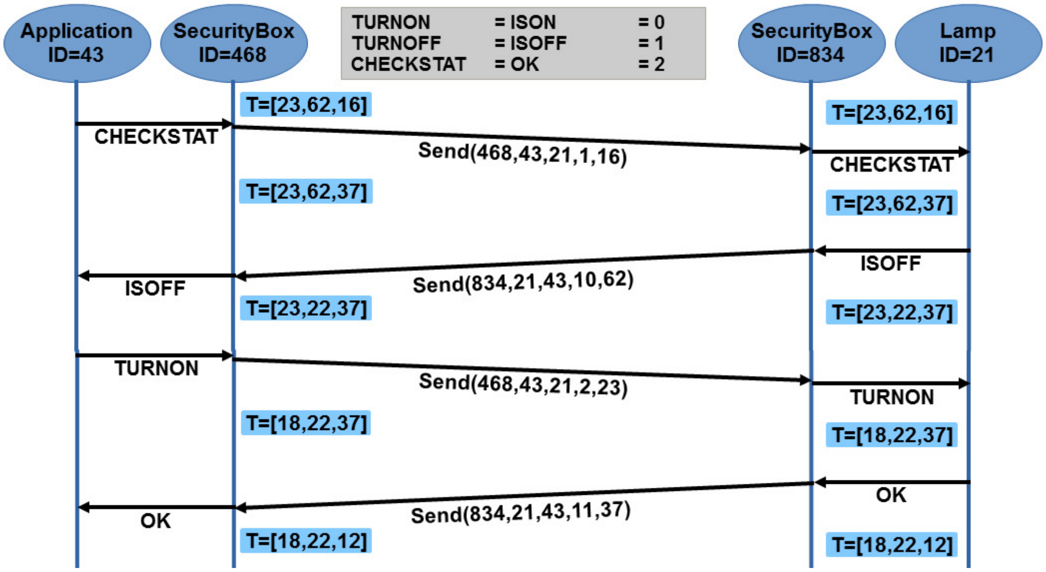


Figure 7. Simple Synchronous Method Example

In this example (Figure 7), we are using a table of *3* items, having *3* possible indexes means *3* different signals (information) sent and *3* different signals received. In the following we have: *TURNON = ISON = 0*, *TURNOFF = ISOFF = 1*, *STATUSCHECK = OK = 2*.

The first and most basic method consists on changing a value after using it. *Agent_A (ID=468)* and *Agent_B (ID=834)* start with a shared secret table *T* and a secret function *f*. *A (ID=43)* starts by sending a *CHECKSTAT* order, which is *index = 2*. Instead of actually sending *2*, *Agent_A* would send *T[2]* which is *16*. After sending *16*, *Agent_A* would change *T[2]* using *f* to obtain *37*. Since *37* replaced *16*, it would have the same index in the table *T*, and since the index is the actual indicator of the meaning of a message, then *37* would have the same meaning as *16* but with a different visible value. In other words, the next time *A* sends a *CHECKSTAT* order, or receives an *OK* response, the visible value in the message would be *37*.

7.2. Simulation

To test our protocol, we created a simulation of the lamp example using a python script on an Intel *i5* CPU *(2.60GHz*4)* with *4GB* of ram. We used the last version of the protocol which includes the following:

- All data are in 16-bits (table values, substitution-boxes values, coded messages, etc.)
- Asynchronous communication between twin security boxes.
- Many additional security measurements to make sure the security requirements are fulfilled.
- An improved version of the *f* function responsible for changing the secret table.
- Dynamic substitution boxes.
- An improved and secured use of messages sequencing.
- Possibility to send and receive unanticipated messages instead of expected signals only. This option reduces the performance.
- A resend option to be able to resend unreceived or erroneous messages.

During our simulations we were able to test different architectures with multiple devices, security boxes, and ports. Simulations were successful and showed no direct patterns on millions of communicated messages.

With all the security measurements implemented, we were able to communicate *4000* messages per second for each twins.

8. CONCLUSION

The SCADA community is looking for a security protocol that has a low cost and does not need constant upgrades. The protocol should also respect the constraints of SCADA such as the low computational power and the real-time environment.

Our security protocol can be a solution to SCADA security since on one side it has a good level of security based on simple operations and on the other side it does not require for the already installed system to change nor to stop working.

The most important security concerns (confidentiality, integrity, and availability) are taken into consideration. The simplicity of the used functions would outplay cryptographic algorithms in the matter of performance especially in the case of embedded devices.

No direct patterns were found, still, we have to continue analysing our protocol for indirect patterns.

We have already advanced in the design of the protocol to be able to communicate normal messages instead of a pre-set of values. We have also added a number of security measurements that consider potential attacks on the basic version of the protocol.

We may find uses of our protocol outside of the SCADA community if the requirements are met and the constraints are respected. For the moment, we focus our work on SCADA systems.

We are currently tuning our simulation to achieve the best possible performance.

We will soon implement our latest version of the protocol on a SCADA platform that we acquire. We have also considered collecting the communicated packages and diffuse them for white hat analysis.

We consider doing some security analysis to test the robustness of our protocol against specific attacks such as correlation and differential power analysis.

We are working with other partners that wish to implement our security protocol in their latest project which aims to put sensors on the repeaters of submarine communication lines to observe and forward information to a research centre.

ACKNOWLEDGEMENTS

The authors would like to thank research engineer Ciprian Teodorov for his encouragements, reviews, and valuable comments. The authors would also like to thank all colleagues that influenced this work.

FUNDING

Funded by the French Government Defence procurement and technology agency - Information department (DGA-MI) and the urban community “Brest Métropole” (BMO).

REFERENCES

[1] Igure, V.M., Laughter, S.A. and Williams, R.D., 2006. Security issues in SCADA networks. Computers & Security, 25(7), pp.498-506.

[2] Pollet, J., 2002, November. Developing a solid SCADA security strategy. In Sensors for Industry Conference, 2002. 2nd ISA/IEEE (pp. 148-156). IEEE.

[3] Bowen, C.L., Buennemeyer, T.K. and Thomas, R.W., 2005, June. Next generation SCADA security: best practices and client puzzles. In Proceedings from the Sixth Annual IEEE SMC Information Assurance Workshop (pp. 426-427). IEEE.

[4] Chandia, R., Gonzalez, J., Kilpatrick, T., Papa, M. and Shenoi, S., 2007, March. Security strategies for SCADA networks. In International Conference on Critical Infrastructure Protection (pp. 117-131). Springer US.

[5] Nicholson, A., Webber, S., Dyer, S., Patel, T. and Janicke, H., 2012. SCADA security in the light of Cyber-Warfare. Computers & Security, 31(4), pp.418-436.

[6] Vernam, G.S., 1919. Secret signaling system. U.S. Patent 1,310,719.

[7] Verfahren, 1923. Device and circuit for News About averaging in cipher. Google Patents DE 371,087. Available: https://www.google.com/patents/DE371087C?cl=en

[8] Shannon, C.E., 1949. Communication theory of secrecy systems. Bell system technical journal, 28(4), pp.656-715.

AUTHORS

Fadi Obeid is a PhD student at ENSTA Bretagne. He is currently researching security solutions for SCADA systems. He received his masters degree in information security and cryptology from the university of Limoges in 2013. He is mainly interested in side channel analysis, security properties verification using model checking, and unbreakable security protocols.

Philippe Dhaussy is a professor at CNRS Lab-STICC within ENSTA Bretagne. His expertise and his research interests include model-driven software engineering, formal validation for real time systems and embedded software design. He has an engineer degree in computer science from ISEN (French Institute of Electronics and Computer Science) in 1978 and received his PhD in 1994 at Telecom Bretagne (France) and his HDR in 2014. From 1980 to 1991, he had been software engineer and technical coordinator in consulting companies (Atlantide group), mainly in real-time system developments. He joined ENSTA-Bretagne in 1996, as professor. He has over 60 publications in the areas of software engineering and computer science. He has been co-supervisor for five PhD students, has been and is involved in several research projects as work package coordinator.

EXPLORING CRITICAL SUCCESS FACTORS FOR CYBERSECURITY IN BHUTAN'S GOVERNMENT ORGANIZATIONS

Pema Choejey, David Murray and Chun Che Fung

School of Engineering and IT, Murdoch University, Perth City, Australia
P.Choejey|D.Murray|L.Fung@murdoch.edu.au

ABSTRACT

This paper presents the results of open-ended survey exploring the critical success factors for cybersecurity implementation in government organisations in Bhutan. Successful implementation of cybersecurity depends on a thorough understanding of cyber threats and challenges to the organisational information assets. It also depends on identification of a responsible, dedicated personnel to lead and direct cybersecurity initiatives. Furthermore, it is important to know the critical areas of cybersecurity activities for management to target, prioritise and execute. Understanding of what key things need to be done right by the responsible agency and its leader, at a particular time and in particular context, can lead to better decision making and resource optimisation including skills and knowledge. The survey findings indicate that, among other factors, awareness and training, policy and standards, and adequate financing and budgetary commitment to cybersecurity projects are three most important success factors. Channelling an organisation's limited resources to these few factors is expected to enhance cybersecurity posture and its management. The research outcome has implications to both government and private organizations in Bhutan.

KEYWORDS

Cybersecurity, Critical Success Factors, Top Management, Awareness and Training

1. INTRODUCTION

Cybersecurity is a global issue that affects both developed and developing countries. Bhutan, which introduced the Internet only in 1999, is facing its own sets of cyber problems. The recent online financial scam, based on the fake email letter that was supposedly sent from the Royal Audit Authority, caused the Bank of Bhutan to transfer 16 million (in Bhutanese currency) to three different accounts in India, Malaysia and Thailand [1]. This cyber incident clearly shows that Bhutan is not immune to cyber threats. Private and government websites have been defaced [2-4] and networks and systems were made inaccessible due to rampant malware and physical disruptions [5].

In just over a decade, the Internet subscriber rate of Bhutan increased from less than 1% in 2004 to 34.3% in 2013. Similarly, the mobile subscriber rate increased from 37% in 2004 to 74.3% in 2013. The Internet and mobile services are now accessible in all 20 *dzongkhags* (or Districts) and 205 *Geogs* (or Village blocks) [6] By 2014, there were more than 80,000 Facebook and Social Networking sites users, which is 10% of the country's 750,000 people [7]

According to the 11th Five Year Plan of 2013, Bhutan's main ICT focus areas are to: i) implement Government-to-Citizens (G2C) services to improve the efficiency and quality of service delivery

Natarajan Meghanathan et al. (Eds) : NETCOM, NCS, WiMoNe, GRAPH-HOC, SPM, CSEIT - 2016
pp. 49– 61, 2016. © CS & IT-CSCP 2016 DOI : 10.5121/csit.2016.61505

to citizens (e.g., online tax filing and birth registration) by improving accessibility, optimizing human resources and reducing service delivery time, ii) establish a government data centre to improve systems reliability, accessibility and resiliency, and iii) consolidate and integrate the wide area network in the capital, which connects all central ministries, and local area networks in the regions for smooth functioning of many services offered online. In addition, the government intends to explore the potential of mobile technology services including implementation of financial payment systems [8-10].

As described earlier, government ICT agenda suggests that Bhutan's dependency on ICT and the Internet is growing and becoming more sophisticated. In other words, it means that its cyber landscape is constantly changing and becoming unpredictable as more people, government, devices, systems and networks become interconnected.

However, aside from the studies in [11, 12], there is no indication of how the government in Bhutan will manage cybersecurity. Clearly, there is a gap of knowledge and understanding of what cyber threats Bhutan is currently facing, who is responsible to lead cybersecurity initiatives and what are the critical success factors that government need to focus upon to make their cyber program a success.

Considering that Bhutan is a developing country, hugely dependent on foreign aid from development partners and international organizations, utilization of limited resources for the wrong strategic goals and objectives may become complete waste of national efforts. Therefore, it is important for the government, policy makers and practitioners to understand and realize what critical things need to done in a specific situation, at a particular time, to make implementation of every national program a success. An understanding of the success factors for cybersecurity is crucial for Bhutan's government, as it has neither material capacity nor human resources to tackle the emerging cybersecurity challenges.

One of the approaches to identity the critical success factors for the organizations is to use the Critical Success Factors (CSFs) method. According to [13, 14], CSFs are defined as *"the limited number of areas in which satisfactory results will ensure successful competitive performance for the individual, department or organisation. CSFs are the few key areas where "things must go right" for the business to flourish and for the manager's goals to be attained. CSFs are the particular areas of major importance to a particular manager, in a particular division, at a particular point in time."*

The key areas are the activities [15]:

- *in which favourable results are necessary to achieve goals.*
- *where things must go right for the organisation to flourish.*
- *that should receive constant attention from management.*

Unlike other approaches, the central idea to CSF method is to focus on "individual managers", by extension to organisations and individuals, and to identify their "information needs". CSF is also unique as it takes into consideration the fact that "information needs vary from manager to manager and that these needs change with time" [13] and by extension with change in environment (e.g., technology). Thus, CSF method is a flexible and dynamic tool that can be used to assess and identify the key areas of activities that are necessary for ensuring the success and performance of a company or an organisation.

While the standard approach of CSFs is to conduct a face-to-face interviews or group discussions with key people in the organisation, this study uses open-ended survey questions to gauge what respondents think and believe would be the critical success factors for implementing cybersecurity in government organisations. The survey approach provides an advantage of having more respondents, anonymity and openness to respond to survey questions.

In the survey, the study asked four open-ended questions to the participants:

- *Please list three of the greatest threats to information resources in your organisation?*
- *Who do you perceive as being responsible for information security in your organisation?*
- *Please list issues that you think are inhibiting cybersecurity effectiveness in your organisation?*
- *Please list things that you think would be critical success factors for implementation of cybersecurity?*

Complete understanding of current cybersecurity situation and context is important. Therefore, the purpose of the study is soliciting knowledge and information on what challenges government organisations are currently facing, who respondents think should be make responsible for cybersecurity and what critical areas the management and its leaders should focus upon to achieve organisational cybersecurity objectives. However, this paper describes only the analysis and findings of the survey responses related to critical success factors for effective cybersecurity implementation.

The paper is organized as follows. Section I introduces Bhutan's cybersecurity situation and the purpose of the study; Section II describes cybersecurity related studies done in Bhutan, Section III presents the research methods and materials; Section IV describes the data analysis and results; Section V provides brief description of study limitations followed by conclusion in Section VI.

2. LITERATURE REVIEW

Because the Internet in general and cybersecurity in particular are fairly new concepts or phenomena, cybersecurity related studies done in Bhutan is far and few.

An E-Readiness study [16] was conducted in 2003 to assess Bhutan's readiness to embrace and participate in the network economy and information society. The purpose of the study was to assess maturity levels in network, human, infrastructure and legal capacity. Country's maturity level below certain threshold in any of these elements is considered as not ready. Knowing the state of ICT development also provide directions where government need to focus and prioritize its national efforts to improve the level of readiness. However, readiness in cybersecurity nor challenges facing Bhutan has been studied.

One of the common mechanisms to counter cybersecurity challenges, especially cyber incidents, is to establish the Computer Incident Response Team (CIRT) [17]. In order to understand how developing countries are managing and responding to cyber incidents, the International Telecommunication Union (ITU) conducted assessment of CIRT covering India, Bhutan, Bangladesh and India [18]. The main objective of the study was to understand cybersecurity challenges facing these countries, to document measures taken to respond to these challenges and to assess their capabilities to coordinate, respond and share information related to cyber incidents. However, this study was limited to cyber incident management capabilities. It has not assessed other security domains such cyber policy, organizational security and personnel security. Nor it has assessed which of security factors developing countries should implement to achieve maximal security benefits.

Another study assessing Bhutan's cybersecurity capability and maturity was conducted by the Global Cyber Security Capacity Centre and the World Bank [19]. The study measured maturity levels in five dimensions: i) policy and strategy, ii) culture and society, iii) education, training and skills, iv) law and regulation, and v) organization, standards and technology. The maturity levels in each dimension were assessed based on five stages: start-up, formative, established, strategic and dynamic. The study findings suggest that Bhutan is at the start-up level of maturity, meaning that Bhutan neither has a capacity nor has undertaken concrete actions with respect to some factors in each dimension. While the study provides an understanding of cybersecurity in Bhutan from the national perspectives, it does not, however, provide specific insights and understanding of how government organizations have implemented cybersecurity activities. Further, their research method is based on group discussion and analysis of available documents.

In [20], a PKI based security framework was proposed for e-government platforms in Bhutan. The framework was derived from PKI solutions and best practices implemented in India, Korea and Taiwan. Even though this study addresses security gaps for e-government platforms, the study is specific to the use of cryptography technologies as solution to the e-government security issues. Moreover, they study used SWOT (Strengths, Weaknesses, Opportunities and Threats) method along with analysis of relevant policy documents.

Recently, an overview of cybersecurity challenges facing Bhutan was presented in [11]. Based on the analysis of available government reports and printed media, common cyber threats and challenges (e.g., hacking and phishing) facing Bhutan were identified and documented. This study was based on a desk audit research method and content analysis, which largely involves reviewing, collation and synthesis of information from secondary sources.

Another recent study related to cybersecurity management was the assessment of cybersecurity practices in the context of e-government implementation [12]. The study surveyed 280 potential respondents to assess the implementation of cybersecurity practices such as cyber policy, risk management, and training and awareness. The study suggests that in most government organizations there is very limited and/or complete lack of cybersecurity policy, risk management, awareness and incident management implementation. It also indicates that many organizations have either suffered from or been affected by cybersecurity threats such as hacking, malware and phishing scams. While the study recommends implementation of both managerial and technological solutions, it does not say which are the few key things government should decide and take action to achieve maximum benefits from security investments.

3. METHODS AND MATERIALS

3.1. Sample and Procedure

A formal approval was sought from the Secretary of the Ministry of Information and Communications (MoIC), Bhutan to provide the contact list of ICT professionals working in various government organisations. Contact addresses of ICT professionals were, then, obtained from the Department of IT and Telecom under the ministry. Emails with a link to the survey were sent to the 280 potential respondents. A follow-up e-mail was sent after one month to improve the survey response rate.

3.2. Instrument

An online survey questionnaire was used to collect data for this study. Survey Monkey was used to design and develop the survey questionnaire. Information related to objectives, confidentiality and consent to participate were included in the survey. The survey also has the option for withdrawal in the case that respondents changed their mind midway through the survey. The survey involved 280 participants. They were asked an open-ended question to list at least 3 critical success factors for cybersecurity program in government organisations. Prior to the actual survey, the questionnaire was pre-tested with 10 senior ICT professionals who were studying abroad in different countries. Further, the survey instrument was reviewed and approved by the Murdoch Ethics Committee to ensure its appropriateness to the research and that the risk factors to the participants were duly considered, especially their privacy and confidentiality.

4. RESULTS

4.1. Response Rate

Electronic mail invitations were sent to potential survey participants to participate in the online survey study. Of 280 respondents, 157 of them responded to the survey. That means that the response rate was about 56% (157/280). However, not all participants who responded to the survey answered all the survey questions. There were only 109 respondents who fully completed the questionnaire. Therefore, the completion rate of the responses was about 69% (109/157).

4.2. Demographic Characteristics

The demographic data is shown in Table 1. Survey participants can be characterised as mostly young with their age ranging from 25 to 34. Most of the participants have a bachelor degree closely followed by diploma and master degree. Their expertise and speciality is mostly in the field of Information Technology, Computer Science and Computer Applications. In terms of gender, more than 68% of participants were male while female participants constituted about 31% of survey responses.

Table 1. Demographic characteristics of survey respondents

Variable		***Frequency***	**Response (%)**
Gender	Male	75	68.81
	Female	34	31.19
Age	45 and over	4	3.67
	35-44	26	23.85
	25-34	72	66.06
	24 and under	7	6.42
Qualification	Certificate	3	2.75
	Diploma	30	27.52
	Bachelor	53	48.62
	Master	23	21.10
	PhD	0	0.00
Specialisation	Computer Science	30	27.52
	Information Technology	53	48.62
	Computer Applications	22	20.18
	Computer Engineering	2	1.83
	Electronics and Communications	1	0.92

	Electrical Engineering	1	0.92
Job Function	Network/System Administrator	26	23.85
	Application/Database Administrator	15	13.76
	IT/Network/Information Systems Security	21	19.27
	IT/MIS/Technical Management	21	19.27
	Web Master/Manager	4	3.67
	Software Programmer/Designer/Developer	11	10.09
	Desktop/Technical Support	11	10.09
Work Experience	Less than 5	29	26.61
	Between 5 and 10	53	48.62
	More than 10	27	24.77

4.3. Analysis

4.3.1 Data Pre-processing

The responses to open-ended questions were analysed using NVivo software. Prior to importing the data into the NVivo program, responses were pre-processed to ensure that non-response items or partially completed responses were removed. Responses were also processed to ensure that words and phrases were correctly spelled and formatted. For example, budget top management is separated as budget and top management or budget, top management. This process improved the quality and accuracy of the data. In addition, responses were categorized into codable texts and classifiable texts. Coding can be performed only on codable texts while classifiable texts can be used for answering multiple questions or to perform demographic comparisons as male versus female.

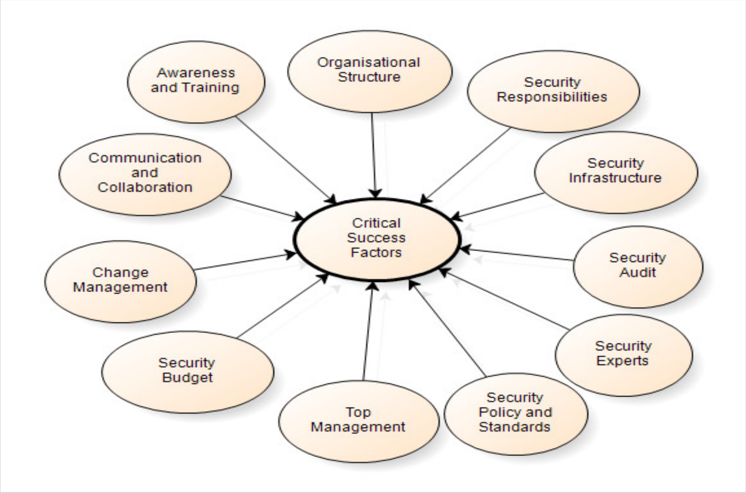


Figure 1. Themes coded from qualitative data

4.3.2. Coding Themes

The coding of qualitative data was performed using the In Vivo Coding method [21]. This method is used to code themes emerging from the codable texts of responses. In other words, it allows texts to be coded using words and phrases found in the qualitative data. For example, as question 4 is related to success factors for cybersecurity implementation in Bhutan, this question is broadly coded as Critical Success Factors under which further sub-themes are categorized. Within this broad category, sub-themes such as awareness and training, security policy and standards, and top management can be categorized. Within the sub-category, for example, training and awareness, there are sub-sub-themes such as seminars, workshops, advocacy, training, etc. These sub-sub-themes constitute or aggregate into abstract concept of training and awareness, which further can be abstracted as one to critical success factors for effective cybersecurity implementation. The resulting coded themes from the qualitative data is shown in Figure 1.

Table 2. Critical success factors for cybersecurity.

Critical Success Factors	***Frequency***	***Percentage* (n=109)***
Awareness and Training	56	51%
Security Policy and Standards	30	28%
Security Budget	23	21%
Top Management	22	20%
Security Infrastructure	15	14%
Security Audit	11	10%
Security Responsibilities	9	8%
Organizational Structure	8	7%
Security Experts	3	3%
Change Management	3	3%
Communication and Collaboration	1	1%

**rounded to nearest percent*

4.4. Key Findings

As different countries face different cybersecurity challenges, the idea was to solicit and understand the prerequisites to cybersecurity implementation success. Therefore, respondents were asked to list at least three critical success factors for cybersecurity in their organisation. The survey results show, see Table 2, that the top five cybersecurity success factors for government organisations are:

- Awareness, training and education.
- Security policy, standards and procedures.
- Cybersecurity financing and resources.
- Top management support for cybersecurity.
- Cybersecurity audit and compliance.

Nearly, 51% (56/109) of respondents believe that government organizations should focus on awareness and training to make cybersecurity a success. Another 27% (30/109) of respondents believe that management should establish policy and standards while 21% (23/109) of respondents think that sufficient budgetary commitment to cybersecurity initiatives will help government organizations to achieve their organizational security objectives. Respondents also identified top management (20%) and security infrastructure (14%) as the fourth and the fifth critical success factors for cybersecurity implementation.

4.5. Recommendations

4.5.1. Awareness and Training

In [22], Fadi argues that educating and training users is must to combat IT security threats. He believes improving the security awareness among the normal users can prevent them becoming the *weakest link* in any organization or becoming an easy and soft target for the cyber criminals [22]. Awareness and training is also important for the legitimate users because people with authorized privilege and access rights bypassed rules to trade-off security against usability, people sometimes make biased decision, so that they gain maximum benefits for the cost of action or decision [23]. Close to 51% of survey respondents believe that awareness and training is the topmost critical success factor that can help government organizations to improve cybersecurity to achieve its business goals and objectives.

4.5.2. Cybersecurity Policy

According to [24], policy in general refers to “a plan or a course of action” that “influence and determine decisions, actions and other matters” of government, organization and business. In the context of cybersecurity, it is a formal statement of “set of rules that dictate acceptable and unacceptable behaviour within an organization”. In other words, the security policy is the foundation for planning, management and maintenance of cybersecurity. Policy drives the implementation of standards which further drives the implementation of practices, procedures and guidelines. Further, policy is a living document that has to be flexible, adaptable and constantly reviewed to reflect the change in environment. The survey results show that nearly 28% of respondents believe that cybersecurity policy is the second most important critical factor to ensure the success of cybersecurity implementation.

4.5.3. Security Budget

Budget underlies any policy initiatives to be undertaken by any government. Without budget and financial resources, it would be impossible to initiate any development activities and implement them successfully. The survey finding suggests that security budget (21%) is the third most important factor that the Bhutanese government should consider while implementing cybersecurity. Budget is central to other priority areas such as training and awareness, security policy and security infrastructure. Without budgetary commitment and resources, none of these critical factors can be implemented successfully.

4.5.4. Top Management Support

The success of cybersecurity efforts depends to a large extent on the commitment and support of the top management [25, 26]. Managerial issues are regarded as the most important security issues and requires management involvement to solve. In a worldwide survey conducted by Knapp et al, [27] found that ‘top management support’ to be the highest ranked issue among a list of 25 information security issues. Top management’s support and commitment is not only significant to planning, executing and governing of security decisions, but also important to demonstrate to security communities and stakeholders that their investment into security benefits them. Therefore, it is important for any organization to have competent and abled security managers to lead the security governance. Nearly, 20% of survey respondents identified management support as of one the critical success factors that government organization should consider for cybersecurity.

4.5.5. Security Infrastructure

Security infrastructure such as hardware and software (e.g., firewalls and intrusion detection systems) are equally important to meet organization's security requirements and implementation of access controls. Cybersecurity is often considered to be technical issue more than management issue. As a result, security mechanisms such as firewalls and antivirus solutions are widely implemented to protect information resources from security breaches. The survey results show that 14% of respondents view security infrastructure as the success factor for cybersecurity.

The study, therefore, recommends government organization to consider and adopt these critical success factors as priority areas to improve cybersecurity in Bhutan.

5. DISCUSSIONS

Cybersecurity may be global in nature but is highly localised to specific organisation in a particular country. No two countries have the same cybersecurity context and the level of maturity [28, 29]. Developing countries such as Bhutan, as described in the literature review, are at a different level of cyber maturity.

The survey results provide a broad perspective of cybersecurity and in particular the direction in which government in Bhutan needs to proceed in cybersecurity implementation. The critical success factors described in the survey findings are identified by the ICT professionals engaged in ICT activities in Bhutan. Therefore, it reflects the practical cyber challenges and the requirements to improve cybersecurity. The top two priorities identified in the survey were awareness and training, and security policy and standards. This suggests that most ICT professionals believe that the majority or most serious issues may be solved within the surveyed group. While there are some who believed that internal or external factors such as security budget, top management and security infrastructure were important, it is promising that the majority of staff were not externalising the problem.

Success factors in information security implementation in government organisations in Oman was explored based on information security experts view [30]. The five success factors identified in the study were: 1) Awareness and Training, ii) Management Support, iii) Budget, iv) Information Security Policy Enforcement and Adaptation, and v) Organisation's Mission. Another study carried out in Iran's Municipal Organisations based on the view of experts in the studied organisations suggests that top management support, information security policy and awareness and training programs are the most important success factors in implementing information security management systems. Furthermore, an empirical study [27] based on the survey of 874 certified information systems security professionals (CISSPs) suggest that top management, security budget and security awareness are among top ten information security issues. Another exploratory research of Yanus and Shin [31] suggests that security technologies, top management support and information awareness and training are factors critical for successful implementation of information awareness program.

The findings of this study in Bhutan shares many similarities and commonalities of success factors that are critical for successful implementation of cybersecurity and security related programs.

This survey was limited only to government organisations. Including survey participants from the corporate and private organisations may have led to different perspective and thinking. Furthermore, inclusion of survey participants of non ICT personnel may result in different findings. However, the survey results provide a list of conceptual areas which may be further investigated to validate their importance to cybersecurity effectiveness. Future work may include other organisations and groups to confirm the applicability of the reported success factors.

6. CONCLUSIONS

This paper presents the results of open-ended survey exploring critical success factors for cybersecurity implementation. This study has surveyed 159 Bhutanese ICT professionals about the key factors for Cyber security success. The results suggest that the top five priorities, in order of reported importance, are:

a) awareness, training and education – ICT professionals who are responsible for cybersecurity and ICT users affected by security issues must be made aware of their security responsibilities and trained in cybersecurity technologies,

b) policy, standards and procedures – policy is the cornerstone for planning and executing cybersecurity initiatives, while standards and procedures are necessary to achieve policy objectives and organisational vision,

c) Cybersecurity budget – budgetary commitment is essential not only for investment in cybersecurity technologies and infrastructure, but also for policy implementation and conduction of cybersecurity training and awareness,

d) top management support – competent leadership drives the success of the organisation. Top management support is essential to get the stakeholders support and secure budget for cybersecurity,

e) security infrastructure – effective cybersecurity needs security controls and tools (e.g., firewalls and antivirus) to mitigate cyber risk and prevent security breaches, and

f) cybersecurity audit process – compliance to cyber rules, policies and data standards are equally important. Cybersecurity audit process ensures that organisations meet the security requirements and remain up to date with changing environment.

The outcome of this research will have significant impact to both governmental organization and non-governmental organizations in terms of understanding the limited number of areas in which satisfactory results will ensure successful competitive performance for the individual, department or organisation. If implemented successfully, these factors would not only improve cybersecurity by reducing security breaches, but also meet organisational goals. However, the identified factors need to be further validated using different tools and techniques.

ACKNOWLEDGEMENTS

The authors would like to thank Miss Dechen Chhoeden, Department of IT and Telecom for her kind support for providing us with the contact list of ICT professionals.

REFERENCES

[1] N. Gyeltshen, "BoB transfers Nu 16M based on fake e-mail," in BBS Online, ed. Thimphu: Bhutan Broadcasting Service, 2016.

[2] B. Shmueli, "RCSC, BoB, RGoB Portal among tens of hacked websites," in ThimphuTech.com: Technology, food and happiness in Bhutan vol. 2014, ed: ThimphuTech.com, 2012.

[3] B. Shmueli, "DrukNet Servers Still Under Attack," in ThimphuTech.com: Technology, food and happiness in Bhutan vol. 2014, ed: ThimphuTech.com, 2012.

[4] B. Shmueli, "Hackers enjoy a free ride using RGoB, OAG, TCC, and other Bhutanese websites," in ThimphuTech.com: Technology, food and happiness in Bhutan vol. 2014, ed: ThimphuTech.com, 2012.

[5] B. Schmueli, "Are Viruses Clogging Bhutan's Information Highways?," in ThimphuTech.com: Technology, food and happiness in Bhutan, ed: ThimphuTech.com, 2010.

[6] MoIC, "Annual InfoComm and Transport Statistical Bulletin," Ministry of Information and Communications, Ed., 5 ed. Thimphu: Royal Governemtn of Bhutan, 2014.

[7] "Internet World Stats: Usage and Population Statistics," 2014, n.d.

[8] GNHC, "Eleventh Five Year Plan Volume I: Main Document," G. N. H. Commission, Ed., ed. Thimphu: GNHC, 2013.

[9] MoIC, "Bhutan e-Government Master Plan," Ministry of Information and Communications, Ed., ed: Royal Government of Bhutan, 2013.

[10] G2C, "G2C: Service Delivery Initiative," ed. Thimphu: Royal Government of Bhutan, n.d.

[11] P. Choejey, C. C. Fung, K. W. Wong, D. Murray, and D. Sonam, "Cybersecurity challenges for Bhutan," in Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2015 12th International Conference on, 2015, pp. 1-5.

[12] P. Choejey, C. C. Fung, K. W. Wong, D. Murray, and H. Xie, "Cybersecurity Practices for E-Government: An Assessment in Bhutan," presented at the The 10th International Conference on e-Business, Bangkok, Thailand, 2015.

[13] J. F. Rockart, "Chief executives define their own data needs," Harvard business review, vol. 57, pp. 81-93, 1978.

[14] C. V. Bullen and J. F. Rockart, "A primer on critical success factors," 1981.

[15] R. A. Caralli and W. R. Wilson, "Applying Critical Success Factors to Information Security Planning," DTIC Document2004.

[16] MoC, "Bhutan e-Readiness Assessment," T. D. o. I. Technology, Ed., ed: Ministry of Communications, 2003.

[17] J. Haller, S. A. Merrell, M. J. Butkovic, and B. J. Willke, "Best Practices for National Cyber Security: Building a National Computer Security Incident Management Capability," DTIC Document2010.

[18] ITU, "Cybersurity: Readiness Assessment for Establishing National CIRT," International Telecommunication Union2012.

[19] T. Roberts, "Building Cyber-Security Capacity in the Kingdom of Bhutan," G. C. S. C. Centre, Ed., ed: University of Oxford undated.

[20] B. Nono, "Proposing a Government PKI in Bhutan: A Solution to e-Government Security Requirements " 2011.

[21] J. Saldaña, The coding manual for qualitative researchers: Sage, 2009.

[22] F. A. Aloul, "The Need for Effective Information Security Awareness," Journal of Advances in Information Technology, vol. 3, 2012.

[23] D. Besnard and B. Arief, "Computer security impaired by legitimate users," Computers & Security, vol. 23, pp. 253-264, 5// 2004.

[24] M. E. Whitman and H. J. Mattord, Management of information security: Nelson Education, 2013.

[25] J. M. Torres, J. M. Sarriegi, and J. Santos, "Critical Success Factors and Indicators to Improve Information Systems Security Management Actions," Handbook of Research on Information Security and Assurance, vol. 160000, p. 140000, 2009.

[26] S. Posthumus and R. von Solms, "A framework for the governance of information security," Computers & Security, vol. 23, pp. 638-646, 2004.

[27] K. J. Knapp, T. E. Marshall, R. K. Rainer, Jr., and D. W. Morrow, "THE TOP INFORMATION SECURITY ISSUES FACING ORGANIZATIONS: WHAT CAN GOVERNMENT DO TO HELP?*," EDPACS, vol. 34, pp. 1-10, Oct 2006 2006.

[28] A. C. Tagert, "Cybersecurity challenges in developing nations," 3445893 Ph.D., Carnegie Mellon University, Ann Arbor, 2010.

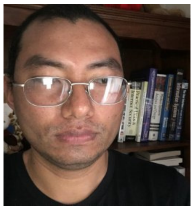
[29] K. P. Newmeyer, "Cybersecurity Strategy in Developing Nations: A Jamaica Case Study," 3616630 Ph.D., Walden University, Ann Arbor, 2014.

[30] M. Al-Awadi and K. Renaud, "Success factors in information security implementation in organizations," in IADIS International Conference e-Society, 2007.

[31] R. Yanus and N. Shin, "Critical Success Factors for Managing an Information Security Awareness Program," in Proceedings of the sixth Annual ISOneWorld Conference, 2007.

AUTHORS

PEMA CHOEJEY



Pema Choejey is currently studying Doctor of Philosophy (Ph.D) in Information Technology, School of Engineering and IT at Murdoch University, Australia. He has bachelor degree in Electronics and Communications Engineering from PSG College of Technology, Bharathiar University, India and master of science in Information Technology from King Mongkut’s University of Technology, Thailand. Prior to becoming a Ph.D student, he worked as the Chief ICT Officer and Head of Research Division for the Department of Information Technology and Telecom under the Ministry of Information and Communications, Bhutan.

CHUN CHE FUNG

Chun Che Fung received his B.Sc.(Hon.) and M.Eng. degrees from the University of Wales in 1981 and 1982 respectively. He was awarded a Ph.D degree from the University of Western Australia in 1994. Currently, he is Professor Emeritus at the School of Engineering and Information Technology, Murdoch University. Prior to his present position, he worked as Associate Professor and Associate Dean of Research at Murdoch University (2003-2015), Senior Lecturer at the School of Electrical and Computer Engineering, Curtin University (1988 to 2002), and the Department of Electronic and Communication Engineering, Singapore Polytechnic (1982 to 1988). His research interests are computational intelligence techniques and intelligent systems applications for practical problems.

DAVID MURRAY

David Murray received his Ph.D degree from Murdoch University. Currently, he is Senior Lecturer at the School of Engineering and Information Technology at Murdoch University. His research interests are in wireless networks, data communications and security. He has published in the areas of TCP Performance Enhancing Proxies, Wi-Fi performance, fast roaming, network measurement, routing protocols and security.

INTENTIONAL BLANK

DMIA: A MALWARE DETECTION SYSTEM ON IOS PLATFORM

Hongliang Liang, Yilun Xie and Yan Song

Beijing University of Posts and Telecommunications, Beijing, China
{hliang, xieyilun, yansong}@bupt.edu.cn

ABSTRACT

iOS is a popular operating system on Apple's smartphones, and recent security events have shown the possibility of stealing the users' privacy in iOS without being detected, such as XcodeGhost. So, we present the design and implementation of a malware vetting system, called DMIA. DMIA first collects runtime information of an app and then distinguish between malicious and normal apps by a novel machine learning model. We evaluated DMIA with 1000 apps from the official App Store. The results of experiments show that DMIA is effective in detecting malwares aimed to steal privacy.

KEYWORDS

iOS, Malware Detection, Dynamic Analysis, Machine Learning

1. INTRODUCTION

Apple iOS is one of the most popular and advanced operating systems for mobile devices on the market. By the end of January 2015, Apple had sold one billion iOS devices [1]. Apple exposes some APIs that can access to users' private data. This arises the privacy and security concerns. Because, for example, accessing to the users' location, can be used to track users across applications. If apps upload user's privacy without notifying users, we can regard these apps as malware. As the same, according to iOS developer license agreement [2], if an app use Private API, it is likely to be malware. Because Private APIs are functions in iOS frameworks reserved only for internal uses in built-in applications. They provide access to various device resources and sensitive information. After all, iOS apps face two threats: *abuse of security-critical Private APIs* and *stealing* (uploading without notifying the user) *privacy data in devices*.

To prevent third-party applications from performing malicious activities, Apple does review each app submission. And any violations of the App Store Review guidelines lead to rejection. It is generally believed that App Review is quite effective. However, recent work [3,4] shows that by constructing the names of Private APIs at runtime, it is possible to invoke Private APIs in third-party applications and still be able to pass the vetting process. Besides, there are several automated binary analysis systems [5, 6, 7, 8] proposed by security researchers to analyse iOS applications. However, the static analysis method in [5] could not resolve API names composed at runtime because of the runtime future and dynamic binding mechanism of Objective-C.

Natarajan Meghanathan et al. (Eds) : NETCOM, NCS, WiMoNe, GRAPH-HOC, SPM, CSEIT - 2016
pp. 63– 72, 2016. © CS & IT-CSCP 2016 DOI : 10.5121/csit.2016.61506

Dynamic approaches in [6, 7, 8] suffer from incomplete code coverage, thus would fail to detect uses of private APIs if malicious application authors place the invocations in complicated triggering conditions. And they could not find the private data uploading.

To improve the situation, we present DMIA in this paper. DMIA puts a monitor layer between system and application to catch the behaviours of an app, without the deficiencies (could not resolve API names composed at runtime) of static analysis caused by iOS runtime. We use 150 popular apps from App Store to train our classification model and it's equivalent to build a whitelist of the app behaviour. In summary, DMIA can solve some problems both in static and dynamic analysis tools. Monitor layer compensates the lack of static analysis which can't resolve API names composed at runtime. Machine learning model improve the problems of dynamic analysis which has high rate of false negatives due to the incomplete coverage of paths.

The main contributions of our paper are as follows:

(1) We insert a monitor layer between iOS system and applications to access applications' sensitive behaviours and network data. The layer can be regarded as a novel and effective dynamic binary instrumentation tool on iOS.

(2) We train a classification model of malicious behaviours based on machine learning method, which can distinguish malicious and normal applications.

The rest of the paper is organized as follows. We present the design of DMIA in section 2 and describe the implementation in section 3. Then we evaluate DMIA in section 4 and compare with related work in Section 5. Section 6 concludes the paper.

2. DESIGN

2.1. System Architecture

The general architecture of DMIA is depicted in Figure 1. DMIA consists of two parts: (1) The monitor layer between applications and the iOS system, (2) The classification model of malicious behaviours.

2.2. Monitor Layer between Applications and iOS System

The monitor layer is consisted of original network monitor, privacy function monitor, and special private APIs monitor.

Several tools like Wireshark can capture the network packages, but it's hard to handle with the issues of data encryption, packet loss, etc. Original network monitor of DMIA get original network data by hooking network functions. It outcomes the deficiencies of Wireshark and lessen the impact of encryption, through preset-value inspection which we will present in 2.3.

As we present before, one goal of DMIA is to detect malware by deciding whether it has uploaded private data or not. iOS will notice users to authorize privacy rights only at the first time to access it. Once a user has authorized it, he will not know when the app accesses his private data. So monitoring privacy function is important in DMIA. We hook those sensitive public APIs, such as CLLocationManager which is provided by CoreLocation framework to get the user location, AddressBook Framework for access to directories and so on to monitor the privacy related behaviour.

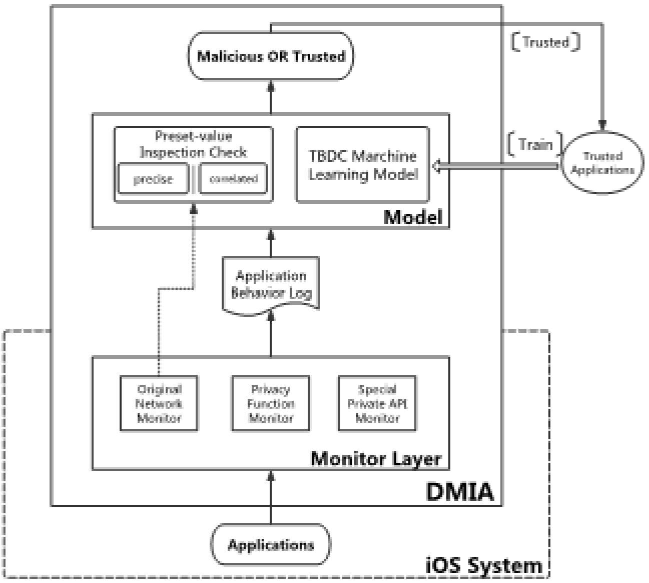


Figure 1. The architecture of DMIA

Private APIs are those functions in iOS frameworks reserved only for internal uses by built-in applications. They can be used to access to various device resources and sensitive information. To monitor private API abuse, we hook these related APIs which are selected according to the head files of private framework and related work about private API abuse, and our iOS development experience.

2.3. Malicious Behaviour Classification Technology Based on the Preset-Value Check

In order to detect whether an app will upload the user privacy, we designed the preset-value mechanism which consists of two kinds: ***precise*** and ***correlated***. We first describe the precise one. We forge privacy data in mobile phone, such as Reminder with specific text and address list with special phone number. We then collect mock privacy data and create a sensitive library based on it. Next, we match it with those network data obtained by the monitor layer. A full match indicates that the application is uploading privacy information illegally. Obviously, the ***precise*** method will fail if malware encrypt network data. To solve this problem, we introduce the ***correlated*** mechanism. Its main idea is the correlation detection. We also collect those mock privacy data to create the sensitive library, but we detect the relevance of them instead of a perfect match. By changing the content of the sensitive library regularly, DMIA monitors whether the communication data will change with it accordingly. If the correlation is greater than a threshold (0.6 is used in the paper), we think that the app uploads the privacy data.

2.4. Malicious Activity Classification Model Based on Machine Learning

Feature selection is a crucial step for machine learning. A reasonable feature will directly outperform the accuracy of most machine learning classifiers, despite some powerful models (e.g.

Long Short-Term Memory [9]) don't have to construct features manually. In section *Feature Vectors*, we discuss the feature selection strategy. In section *TBDC*, we propose our threshold-based dynamic classification model (TBDC).

Feature Vectors

Frequency of Sensitive APIs: malwares usually ask for more permissions than needed, and use them to obtain and upload sensitive information frequently. For example, a video app usually doesn't need to know who are in the user's contacts. Accordingly, the frequency of using sensitive APIs will be different between malicious and benign apps. Thus, we use the frequency of sensitive APIs as one feature.

In details, let f^{1i} denote the occurrence frequency of the *i*th API, where the superscript 1 means it is the first feature and $i \in [1,2,\dots,L]$ where *L* denoting the size of APIs that are being monitored, and it is computed as:

$$c_i = \begin{cases} c_i + 1 & if\ S_j = c_i \\ c_i & if\ S_j \neq c_i \end{cases} \quad (1 \leq j \leq |S|, 1 \leq i \leq L) \tag{1}$$

$$f^{1i} = c_i / |S| \tag{2}$$

Where S denotes the API sequence of an app, and is extracted from the system logs. |S| is the length of S. c_i is the occurrence number of the *i*th API in sequence S, which is initialized to zero.

Frequency of TBDC: The amount of sensitive API in iOS is very huge (In our experiment, we totally monitor 140 APIs). Intuitively, one app will just call a part of the APIs in their life cycle instead of all APIs (we prove that the conjecture is correct by our experiments). This phenomenon will lead to sparse feature vectors, which may increase the difficulty of model's training.

Consider this condition, we group the sensitive APIs into a much smaller set, which is based on their functions (e.g. Network, AddressBook). Assume the size of grouped API is L_g, we replace the API scope size L with L_g in equation (1) to compute the TBDC frequency. We define the feature vector obtained in this step as f^2.

Uncontrollable Behaviour Extraction: Frequency based features have enough power to represent the characteristics of different kinds of apps, but have limitations on one-class apps. For example, network-related APIs are used in video apps more often than in other apps, no matter they are malicious or benign.

In order to overcome this shortage, it is much important to know whether the behaviours of an app are under the user's control. We call the behaviours without requesting user's permissions as uncontrollable behaviours. For example, if there is a user interaction event (e.g. click button) just before a network request, the behaviour is regarded under the user's control (controllable behaviour).

In summary, a behaviour (e.g. URL request, address request) is defined as controllable behaviour only when it is just after a user interaction event. In the opposite, we define it as uncontrollable behaviour. We adopt the frequency of uncontrollable behaviours to generate the feature vector. In more detail, let $\text{Behavior}_{\text{all}}$ be all behaviours, which we are interested in, occurred in the API sequence. Alike, let $\text{Behavior}_{\text{uncon}}$ be those uncontrollable behaviours. Then, the feature vector f^3 is computed as $f^3 = \text{Behavior}_{\text{uncon}} / \text{Behavior}_{\text{all}}$.

Threshold-Base Dynamic Classification Model

Generally, before machine learning classifier get good performance, it needs a lot of data to training. Because, except the over fitting problem, more good data always lead to better performance, at least not worse than before. But when the training dataset is not enough for the classifier to learn necessary attributes, it just become a shot in the dark.

From this, we propose Threshold-Based Dynamic Classification Model (TBDC), which can own good result even when training dataset is small. The essential idea is, first, we train a regression model with the small dataset. Then, we compute a threshold of becoming malware according to the output of the regression model with the initial dataset as input. Finally, make the test samples' feature vectors as inputs, we can get the outputs of the learned regression model. If the output fall outside the threshold range, we classify it into malicious, otherwise, benign. And if there are new training samples, we will retrain the regression model to adjust the parameters, and generate new threshold dynamically, which makes TBDC have the ability to classify in a more fine-grain way.

In more detail, Let f be a vector consisting of all the feature vectors $[f^1,f^2,\ldots,f^m]$, where m is the amount of features (e.g. m = 3 in our feature space). Let $M \in \mathbb{R}^{n\times\Sigma_j^m|f^j|}$ be a matrix consisting of all samples' feature vector, where n quantifies the number of input samples, which is also treated as input matrix. For example, M_i is the *i*th row of M, which donates the *i*th sample's feature vector. Then, the output vector y and thresholds are computed as:

$$y = g(M) \tag{3}$$

$$[threshold_{min}, threshold_{max}] = h(y) \tag{4}$$

Where g is the regression function, h is a algorithm, which is used to compute the thresholds.

3. IMPLEMENTATION

In this paper, the monitor Layer runs between the jail-broken iOS system and applications. We use *Tai Chi tool* to jailbreak iOS 8.3 and *MobileSubstrate* to insert hooks at system level. We develop a dylib by iOS OpenDev and program with logo language. We debug and test our dylib on iPod Touch 5. Our preset-value inspection and TBDC model are developed by python.

3.1. Arrange Privacy Related Functions

In order to obtain more accurate and comprehensive information of privacy behaviour, we search Apple API documents based on all options in iOS system - Settings - Privacy. In the end, we collect 216 related functions. Then remove duplicate functions based on action and implement 89 hooks of key functions. Besides, we have also collect APIs about device information, such as

[UIDevice identifier For Vendor] (it can be used to get device UUID) and so on, total 15 functions. At last, to vet private API abuse, we export private API headers of iOS 8.3 SDK by class-dump. According to our development experience and function names, we sort out 31 privacy related functions. Then controlling of the existing research on private API abuse such as iRiS [10], we ultimately determine 36 private API function related privacy and hook them. At this point we have completed the work of arranging privacy related functions (140 totally).

3.2. Insert Monitor Layer

Hook the 140 functions sorted out by section 3.1. Program tweak by logo language and the program consist of 11 parts. NetworkHook, AddressBookHook, EKEventStoreHook, Calendar and Reminder, PhotoHook, MicHook, CameraHook, HealthHook, HomeKitHook, CLLocationHook, PrivateApiHook, OtherHook. Under the premise of keeping the original function of method, we append behaviour record to them. Thus, we can record the parameters, return values and call time into the system log according to the prescribed format. Finally, we compile the code into BehaviorMonitor.dylib and load it into iPod touch 5.

3.3. Implementation of TBDC

First, we give each API an index to map the text name into vector space. For example, *initWithRequest:delegate* is the first sensitive API that we monitored, thus, we index it with integer 1. Next, we extract app's API call sequence from the system log, and record it with API index. For example, a simple network request is achieved with *initWithRequest:delegate:startImmediately:* and *connection:WillSendRequest:redirectResponse* after it. So, we transpose this sequence into 2 11, where 2 and 11 are the index of the two APIs respectively. Then we construct the feature vectors as we discussed in section 2.4.1.

As for the regression function, we tried Support Vector Regression (SVR), which is based on Support Vector Machine (SVM), and Multilayer Perceptron (MLP). To get the threshold range, we simply set the minimum of benign samples' outputs as the minimum, and the maximum of benign samples' outputs as the maximum.

4. EVALUATION

In order to judge whether DMIA is effective and efficient in detecting malware, we have carried out massive experiments. Further more, for the two threats (abuse of security-critical Private APIs and stealing privacy data in devices) focused by DMIA, we expound them respectively in the end of evaluation as case studies.

We evaluate DMIA with 1000 applications from App Store. There are 24 categories in total: Children, education, shopping, photo & video, efficiency, food, live, fitness, journey, music, sport, business, news, tools, entertainment, social contact, newspapers and periodicals, finance, reference, navigation, medical treatment, books, weather and commodities guide. We download them from iTunes and install them in iPod Touch5, which is iOS 8.0 version. We run and capture every app's behaviour by Monitor Layer.

In the experiments, we collect 606132 pieces of text messages (over 64MB, size in total), which record behaviours of these apps. In these messages, about 430 thousand pieces (71%) are related to network API, about 48 thousand pieces (8%) are related to location API, and 30 thousand pieces (5%) are related to photo and camera API. 97 thousand pieces (16%) are related to all the rest APIs.

We make a statistic of that whether one app in particular category use one privacy related API or not. We assume that API calls of each normal app in particular category are similar. So, if there are a handful of apps (less than % 3) in particular category using a privacy related API, we think it suspicious. Following, we take Location API and AddressBook API for example. For the frequency of using Location API, navigation class applications are the highest (100%) and weather (98%), social (85%), food (81%), finance class applications and efficiency class applications are the lowest (5%). The APIs related to address book, are used by 76 of the 100 apps in social contact classification, in contrast, by 2 of the 100 apps in weather classification. So we think these two apps suspicious and analyse them carefully. Interestingly, they are not real weather-class apps. They just mark themselves as weather category when applying for app review. Among them, *pp assistant for phone* uploads users' privacy data without notification obviously and it is regarded as a malware. We review its *detail page* and *comments* on iTunes and find that its details screenshot is a game rather than a weather or assistant picture. 1177 of its 1283 comments are puzzling sentences and generated by robot obviously.

Case Studies

This paper focuses on the two kinds of threats in iOS system. Abuse of security-critical Private APIs and stealing (uploading without notifying the user) privacy data in devices. Here, we take i4Tools and Youmi SDK [11] for example to explain how DMIA resist the two threats and demonstrate the effectiveness of DMIA again.

i4Tools. In the lot-sizing tests, DMIA find the features of i4Tools are far away from normal value, which means it may be a malware. So we analyse its text carefully, which has 5600 lines. 4512 lines (80.5%) of them are related to network, 128 lines refer to geographical location information, 124 lines are private API. Especially, 84 of the 124 lines are about LSApplication class. So we know i4Tools break the iPhone developer agreement. What is more, according to preset-value detection, we find it still request network at a fixed time when screen interaction events don't happen. The correlation value between getting and uploading privacy is 0.85. It is greater than our threshold of 0.6. So it uploads data without permission or knowledge of the user. In a conclusion, it is a malware.

Youmi SDK. For apps using advertising app SDK, the proportion of malware in is much higher than others. Especially, almost all of the app texts containing youmi.com are judged to be abnormal by DMIA. So we suspect that the issue is in Youmi SDK. We download Youmi SDK from its official website and program a demo app according to its instructions. Then we test this demo with DMIA. And we get a total of 3221 lines information, of which 153 line involving private API. But it had unauthorized network transmission only when starting the app, and the requests at the rest of the time are all normal. So we can only say it violates the Apple's user agreement and abuse the user privacy data.

5. RELATED WORK

The work related to DMIA can be classified into two categories: (1) Privacy Leak Detection on iOS (2) Machine Learning Model.

5.1. Privacy Leak Detection on iOS

SecLab's PiOS [5] can detecting privacy leak of app. It creates hierarchical structure of class from binary file and build CFG. Analysing data stream to judge that weather privacy information transform from origin to leak point. This is a kind of static analysis. There are several shortcomings of PiOS such high false positive.

To overcome it, Peter Gilbert introduce some other ideas in 2011.6. They create AppInspector [12], a dynamic analysis tool. It can obtain application behaviour by analysing system call. Then summer out wither application access to privacy information or not.

Martin Szydlowski discussed the challenge on dynamic analysis of iOS app and developed a prototype. It can rack sensitive API calls by breakpoint debug and gets app UI model automatically [6]. 2012, Joorabchi and Mesbah implemented iCrawler. It can view the app UI and generate a model containing different UI state. This tool has accelerated the process of iOS app reverse [8]. Although the achieved coverage of their navigation technique looks promising when applied on a few open-source apps, it does not support simulation of any advanced gestures or external events. Moreover, the technique used by iCrawler is only applicable to standard UI elements, and, most notably, iCrawler has not been designed to perform privacy analysis.

Andreas Kurtz introduces DIOS, which is an iOS privacy leak analysis model based on dynamic API call sequence [7]. DIOS mainly includes three parts: Backend is used as the central data storage, worker is data interactive link between backend and iOS device and client is used for analysing the behaviour of the iOS App. DIOS can monitor privacy data access by hooking iOS API function. But the access to private data is not the same as privacy leak. And compared to the static analysis dynamic analysis has high false negative rate and low speed. In contrast, DMIA not only hook a greater variety of private functions, but also monitor the Private APIs and the network. And based on application behaviour, it can determine that it is normal access or privacy theft by preset-value inspection and TBDC model.

5.2. Machine Learning Model

Resent work by Gorla et al. [13] try to use app descriptions and sensitive APIs to check app behaviour in Android platform. They cluster apps that have analogical behaviours into one category and selected the most used APIs the feature of that category. An app will be classified depending on whether it's APIs is accord with the category's, which it belongs, APIs. But they do not construct any features with the sensitive APIs.

DroidADDMiner [14] is a machine learning model based on FlowDroid [15]. It adopts data flow analysis of sensitive APIs to capture the semantics information of malware. But it relies on big training dataset to get good performance.

Sundarkumar et al. [16] tried to use API information to characterize Android malware. They use text mining and topic modelling, combined with machine learning classifier, to detect malwares. But due to the shortage of static analysis, which their works mainly based on, there are false negative and false positive problems.

Some other systems [17, 18] also use static analysis to get API information, as part of their features. Their features also contain other information (e.g. permissions, package information) which they think crucial. But they all suffer at static analysis shortage and big training dataset.

6. CONCLUSION

In this paper, we present the design and implementation of a malware vetting system, called DMIA. It first collects application behaviour information and original network data via its monitor layer. The monitor layer can be considered as a novel dynamic binary instrumentation tool on iOS. Then, DMIA captures violations of stealing users' privacy by the novel *machine learning model*. Finally, our experiments with 1000 applications show that DMIA is powerful in detecting malwares. In our future work, we aim to provide DMIA as a usual app without requiring users to jailbreak their devices. Users can detect their apps by DMIA and upload the results to our server. We hope optimize our training set by this crowdsourcing technique and make DMIA more powerful.

REFERENCES

[1] Gize BusinessInsider. Apple has shipped 1 billion ios devices. http://www.businessinsider.com/apple-ships-one-billion-ios-devices-2015-1.

[2] Apple. ios developer program license agreement. http://www.thephoneappcompany.com/ios_program_standard_agreement_20130610.pdf..

[3] J. Han, S. M. Kywe, Q. Yan, F. Bao, R. Deng,D. Gao, Y. Li, and J. Zhou. Launching generic attacks on ios with approved third-party applications. In Applied Cryptography and Network Security, pages 272-289. Springer, 2013.

[4] T. Wang, K. Lu, L. Lu, S. Chung, and W. Lee. Jekyll on ios: When benign apps become evil. In Usenix Security, volume 13, 2013.

[5] M. Egele, C. Kruegel, E. Kirda, and G. Vigna. Pios: Detecting privacy leaks in ios applications. In NDSS, 2011.

[6] M. Szydlowski, M. Egele, C. Kruegel, and G. Vigna. Challenges for dynamic analysis of ios applications. In Open Problems in Network Security, pages 65-77. Springer, 2012.

[7] A. Kurtz, A. Weinlein, C. Settgast, and F. Freiling. Dios: Dynamic privacy analysis of ios applications. Technical Report CS-2014-03, Department of Computer Science, June 2014.

[8] M. E. Joorabchi and A. Mesbah. Reverse engineering ios mobile applications. In Reverse Engineering (WCRE), 2012 19th Working Conference on, pages 177-186. IEEE, 2012.

[9] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.

[10] Zhui Deng, Brendan Saltaformaggio, Xiangyu Zhang, Dongyan Xu: iRiS: Vetting Private API Abuse in iOS Applications. ACM Conference on Computer and Communications Security 2015: 44-56

[11] https://www.theiphonewiki.com/wiki/Malware_for_iOS#Youmi_Ad_SDK_.28October_2015.29

[12] Peter Gilbert, Byung-Gon Chun, Landon P. Cox, Jaeyeon Jung. Automating Privacy Testing of Smartphone Applications.

[13] A. Gorla, I. Tavecchia, F. Gross, and A. Zeller, Checking app behavior against app descriptions, in 36th International Conference on Software Engineering, ICSE' 14, Hyderabad, India - May 31 - June 07, 2014, 2014, pp. 1025-1035.

[14] Yongfeng Li, Tong Shen, Xin Sun, Xuerui Pan, Bing Mao: Detection, Classification and Characterization of Android Malware Using API Data Dependency. SecureComm 2015: 23-40

[15] Arzt, S., Rasthofer, S., Fritz, C., Bodden, E., Bartel, A., Klein, J., Le Traon, Y., Octeau, D., McDaniel, P.: Flowdroid: precise context, flow, field, object-sensitive and lifecycle-aware taint analysis for android apps. In: Proceedings of the 35th ACM SIGPLAN Conference on.

[16] G. G. Sundarkumar, V. Ravi, I. Nwogu, and V. Govindaraju, Malware detection via API calls, topic models and machine learning, in IEEE International Conference on Automation Science and Engineering, CASE 2015, Gothenburg, Sweden, August 24-28, 2015, 2015, pp. 1212-1217.

[17] P. P. K. Chan and W.-K. Song, Static detection of Android malware by using permissions and API calls, in 2014 International Conference on Machine Learning and Cybernetics, Lanzhou, China, July 13-16, 2014, 2014, pp. 82-87.

[18] A. Sharma and S. K. Dash, Mining API Calls and Permissions for Android Malware Detection, in Cryptology and Network Security-13th International Conference, CANS 2014, Heraklion, Crete, Greece, October 22-24, 2014. Proceedings, 2014, pp. 191-205.

PERFORMANCE COMPARISON DCM VERSUS QPSK FOR HIGH DATA RATES IN THE MBOFDM UWB SYSTEM

Naziha NOURI[1], Asma MESSAOUDI[2] and Ridha BOUALLEGUE[3]

RL Innov'Com : Innovation of communicating and cooperative mobile,
Higher School of Communications of Tunis, Sup'Com Tunisia
[1]nourinaziha@yahoo.fr
[2]Asma.Messaoudi@enit.rnu.tn
[3]ridha.bouallegue@supcom.rnu.tn

ABSTRACT

This paper presents the advantage of using a new modulation scheme called dual carrier Modulation (DCM) compared to classical Quadrature Phase Shift Keying (QPSK) modulation. This comparison is done at data transmission broadband in Multiband OFDM system (MBOFDM) based on Ultra Wide Band UWB. Simulation results show that the use of the modulation DCM for high data rates is more efficient compared with QPSK.

KEYWORDS

ECMA368, MBOFDM, WPAN, QPSK, DCM.

1. INTRODUCTION

UWB communication, accepted by the Federal Communications Commission (FCC) [1] in 2002, is characterized by its low cost for high data rates over wireless personal area networks WPAN. Recently, the UWB system based on MB-OFDM has been exploited in many fields such as wireless personal area networks offering broadband over 480 Mbps, low power multimedia capabilities for PCs, user electronics, mobile and automotive market [2].

In the MB-OFDM physical layer, DCM has been suggested as a new modulation scheme for high data rates [3].

QPSK and DCM are exploited as modulation schemes for MB-OFDM in ECMA-368. QPSK constellation is used for data rates 200 Mb/s and lower, while DCM is used as a multi-dimensional constellation for data rates 320 Mb/s and higher.

This paper compares the performance between DCM and QPSK modulation for high rates within the ECMA-368 standard using saleh and vanzuella channel.

Natarajan Meghanathan et al. (Eds) : NETCOM, NCS, WiMoNe, GRAPH-HOC, SPM, CSEIT - 2016
pp. 73– 81, 2016. © CS & IT-CSCP 2016 DOI : 10.5121/csit.2016.61507

The second section presents the studied system in general, the third one describes the two modulations QPSK and DCM, and then a performance comparison is discussed in section IV, and finally a conclusion in last Section.

2. MB-OFDM SYSTEM ARCHITECTURE

The architecture of the MB-OFDM transmitter is shown in Figure 1. An error correction code (FEC (Forward Error Correction)) is then applied to provide resilience against transmission errors. The punching step is essential to get high data rates.

The encoded sequence is interleaved in three consecutive steps as will be explained later, followed by the binary coding operation or symbol mapping of an OFDM symbol.

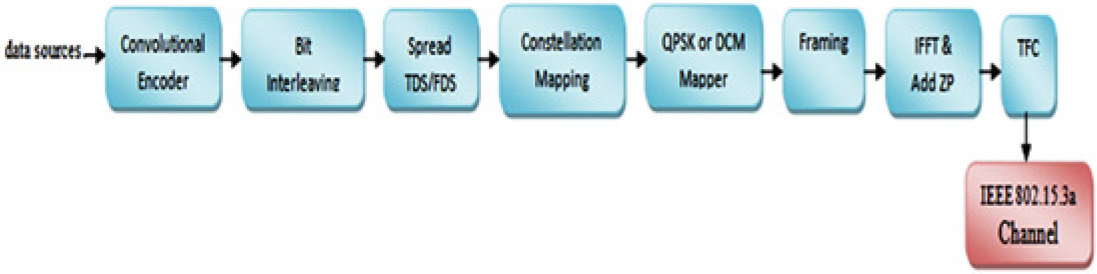


Figure 1: The transmitter architecture for the MB-OFDM system

The MB-OFDM system rates are summarized in Table 1 as they were defined in the ECMA-368 standard [4].

Tableau 1: Characteristics rates of MB-OFDM solution and number of bits per block

Data Rate (MB/s)	*Modulation*	*Code rate (R)*	*FDS*	*TDS*	*Coded Bits/6 OFDM Symbol (N_{CBP6S})*	*Info Bits/6 OFDM Symbol (N_{IB6S})*
53.3	QPSK	1/3	Yes	Yes	300	100
80	QPSK	1/2	Yes	Yes	300	150
106.7	QPSK	1/3	No	Yes	600	200
160	QPSK	1/2	No	Yes	600	300
200	QPSK	5/8	No	Yes	600	375
320	DCM	1/2	No	No	1200	600
400	DCM	5/8	No	No	1200	750
480	DCM	3/4	No	No	1200	900

3. QPSK AND DCM

3.1. QPSK

For rates below 200 Mb/s, the interleaved binary data is mapped to a constellation into QPSK Quadrature Phase.

$$d[k] = k_{MOD} \times [(2 \times b[2k] - 1) + j(2 \times b[2k+1] - 1)] \tag{1}$$

Where $k = 0,1,2,\ldots,$

And $k_{MOD} = {}^{1}/_{\sqrt{2}}$

3.2 DCM

For rates of 320 Mb/s and more, not using spreading (Tab. 1), the binary data is mapped to a multidimensional constellation using a dual-carrier modulation technique (DCM). Note that the first proposals MB-OFDM for IEEE 802.15.3a, including the proposal of September 2004, only consider a QPSK constellation for all data rates [6].

Indeed, the DCM modulation has been adopted by ECMA and recently amended in [7].

Passing from a QPSK constellation to a DCM scheme, an additional form of diversity may be obtained, which leads to an improvement in the overall range of the system.

This diversity is introduced by mapping four bits on two constellations of 16QAM (Quadrature Amplitude Modulation) shown in Figure 3.

The symbols obtained are then converted into bits which are separated by at least 200 MHz of bandwidth, which is approximately equivalent to 50 subcarriers.

The transmission of each obtained flow is carried out, respectively, by the first half and the second half of the subcarriers of the OFDM symbol. The different stages of the DCM modulation are listed as follows:

1. Binary input data coded and interleaved, b[i] where i = 0,1,2, ..., are divided into groups of 200 bits and converted into complex numbers 100 using a technique called double carrier modulation.
2. The 200 coded bits are grouped into 50 groups of 4 bits as shown in Table 2. Each group is represented as (b [g (k)], b [g (k) +1], b [g (k) + 50], b [g (k) 51]), where k ∈ [0, 49] and

$$g(k) = \begin{cases} 2k & k \in [0.24] \\ 2k + 50 & k \in [25.49] \end{cases} \tag{2}$$

3. Each group of 4 bits (b [g (k)], b [g (k) +1], b [g (k) +50], b [g (k) 51]) must be mapped to a constellation four-dimensional, and converted into two complex numbers (d [k], d [k + 50]).
4. Complex numbers must be normalized using a normalization factor :

$$k_{MOD} = {}^{1}/_{\sqrt{10}}$$

The use of non-adjacent subcarriers within this transmission makes it possible to increase the reliability of these high speeds modes towards the fading effect of the transmission channel.

The DCM technique is not applied to the low flow rates (200 Mb/s and below) as frequency diversity is better exploited through the use of FEC codes and TDS and FDS low flow techniques. Therefore, the expected diversity gain for these flows of DCM is minimal and the added complexity for DCM is not justified.

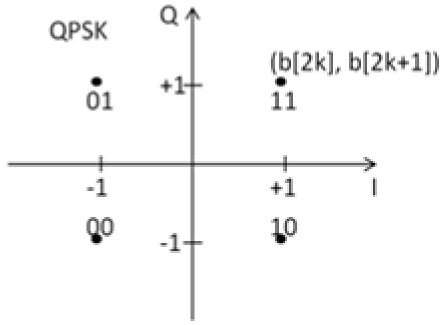


Figure 1: QPSK Constellation as standard ECM-368.

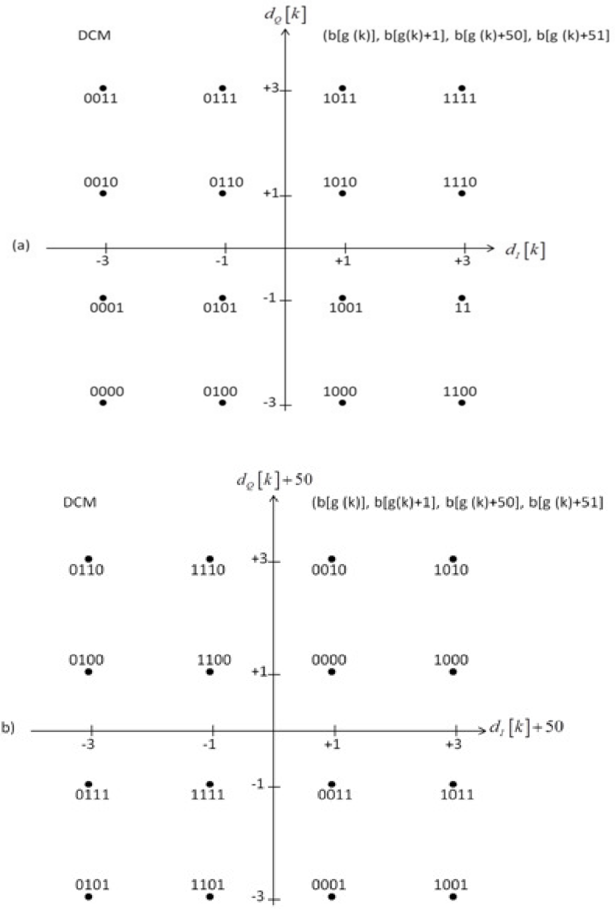
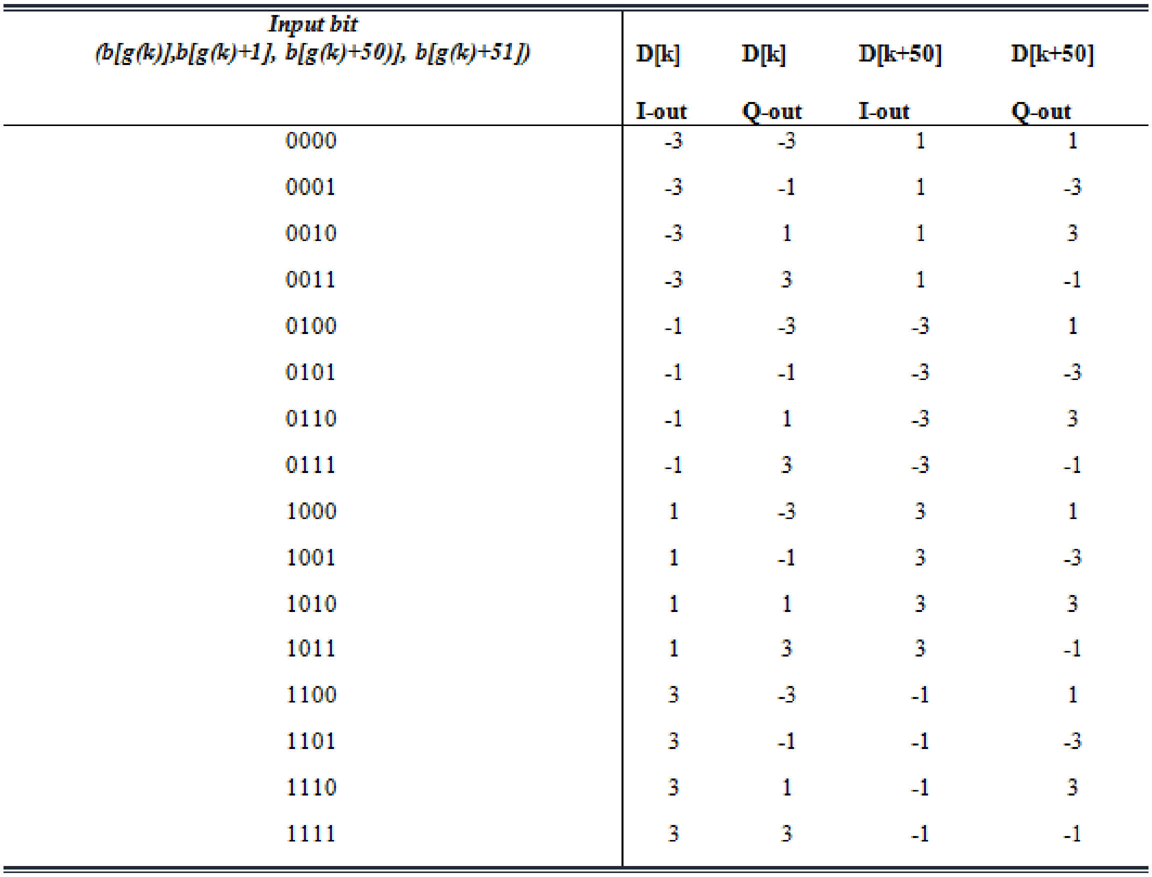


Figure 2: 16-QAM Constellations of the DCM modulation for MB-OFDM system.

Tableau 2: Encoding table of the dual carrier modulation (DCM).



Input bit *(b[g(k)],b[g(k)+1], b[g(k)+50)], b[g(k)+51])*	D[k]	D[k]	D[k+50]	D[k+50]
	I-out	Q-out	I-out	Q-out
0000	-3	-3	1	1
0001	-3	-1	1	-3
0010	-3	1	1	3
0011	-3	3	1	-1
0100	-1	-3	-3	1
0101	-1	-1	-3	-3
0110	-1	1	-3	3
0111	-1	3	-3	-1
1000	1	-3	3	1
1001	1	-1	3	-3
1010	1	1	3	3
1011	1	3	3	-1
1100	3	-3	-1	1
1101	3	-1	-1	-3
1110	3	1	-1	3
1111	3	3	-1	-1

4. SIMULATION RESULTS

4.1. Parameters

In this section, we present some simulations results per-formed on the MB-OFDM system described above in or-der to analyse its performance.

In these simulations, we consider an indoor environment.

In the following, the performance evaluation of the MB-OFDM system is achieved by simulating the transmission chain for the CM1 channel model (channel models), adopted by the working group IEEE 802.15.3a, whose characteristics are shown in Table 3.

The results, expressed in terms of bit error rate (BER), were averaged over 100 iterations of distinct channels of each model CM. Each new transmitted frame of OFDM symbols employs a new channel realization. The stop conditions being themselves tested every 100 received frames, i.e., until all the channels achievements are proven.

The made TFC are those of group 1 (Tab. 4). Thus, the system works using the US regulations set by the FCC.

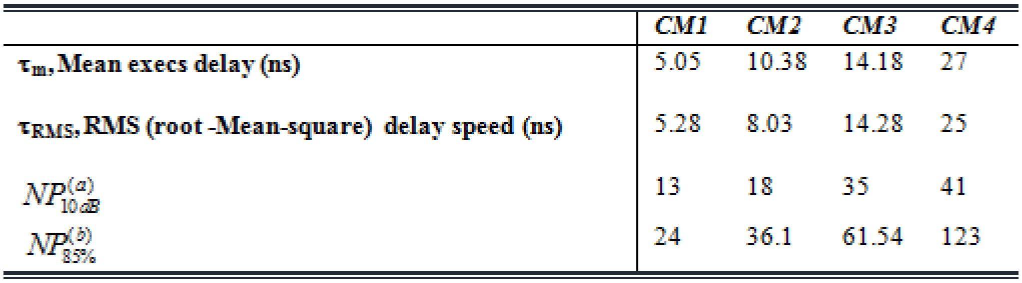
The simulation results are evaluated in terms of Eb / No, where Eb is the average energy per useful bit and No is the power density of AWGN (Additive white Gaussian noise), and perfect channel estimation is considered.

The performance of the MB-OFDM system is shown for various flow rates listed in Table 5. Note that the parameters considered here are those adopted in [5].

Replacing QPSK by DCM slightly improves the system performance to the detriment of the complexity of the higher-level system.

In reception, the used equalization technique is ZF. It is implemented from the coefficients obtained through a full assessment of the transmission channel.

Tableau 1: IEEE 802.15.3a Model Specifications for four distinct configurations.

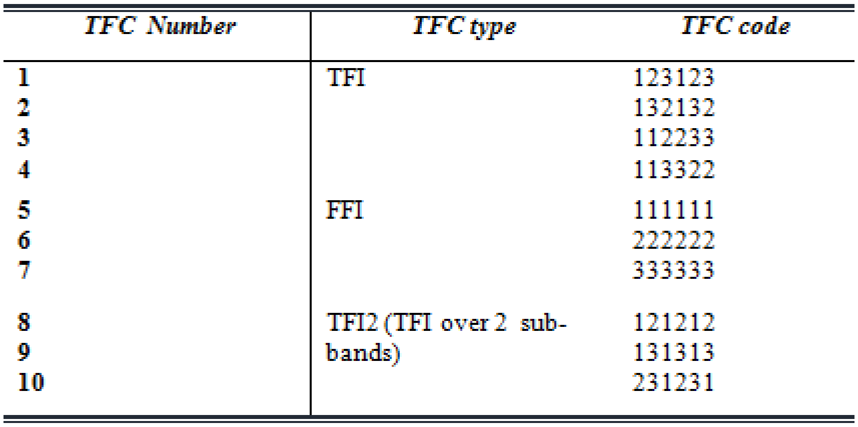


	CM1	*CM2*	*CM3*	*CM4*
τ_m, Mean execs delay (ns)	5.05	10.38	14.18	27
τ_{RMS}, RMS (root -Mean-square) delay speed (ns)	5.28	8.03	14.28	25
$NP_{10dB}^{(a)}$	13	18	35	41
$NP_{85\%}^{(b)}$	24	36.1	61.54	123

(a) Number of paths above the threshold of -10 dB relative to the dominant path

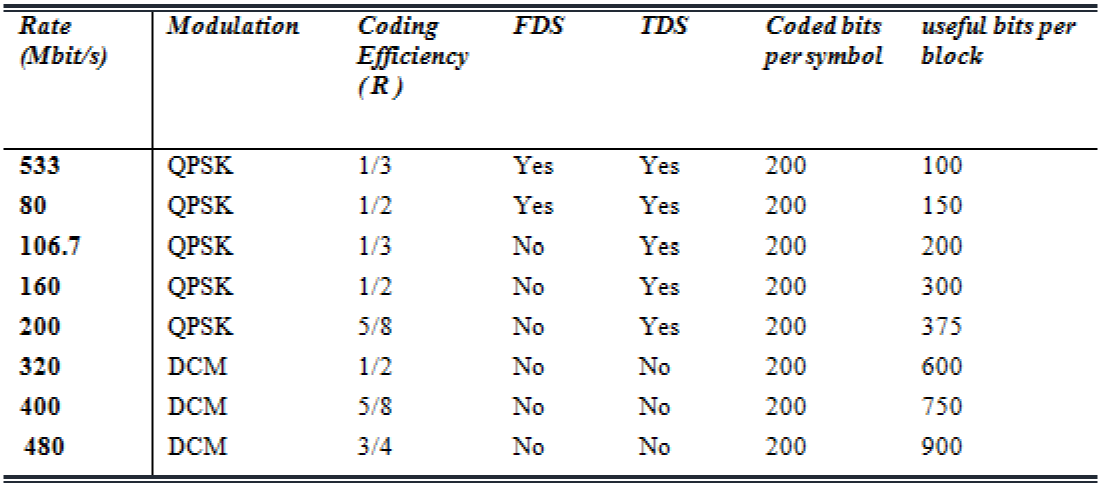
(b) Number of paths containing 85% of the energy of the impulse response

Tableau 2: Time-Frequency codes (TFC) for the band group 1 [4].



TFC Number	*TFC type*	*TFC code*
1	TFI	123123
2		132132
3		112233
4		113322
5	FFI	111111
6		222222
7		333333
8	TFI2 (TFI over 2 sub-bands)	121212
9		131313
10		231231

Tableau 3: Rates of MB-OFDM system based on the ECMA-368 standard [4].



Rate (Mbit/s)	*Modulation*	*Coding Efficiency (R)*	*FDS*	*TDS*	*Coded bits per symbol*	*useful bits per block*
533	QPSK	1/3	Yes	Yes	200	100
80	QPSK	1/2	Yes	Yes	200	150
106.7	QPSK	1/3	No	Yes	200	200
160	QPSK	1/2	No	Yes	200	300
200	QPSK	5/8	No	Yes	200	375
320	DCM	1/2	No	No	200	600
400	DCM	5/8	No	No	200	750
480	DCM	3/4	No	No	200	900

4.2. DCM vs QPSK

Dual carrier modulation (DCM) is being studied as a potential improvement of the classic modulation QPSK. DCM provides not only a diversity gain, but also a coding gain. DCM results are compared with those using conventional QPSK.

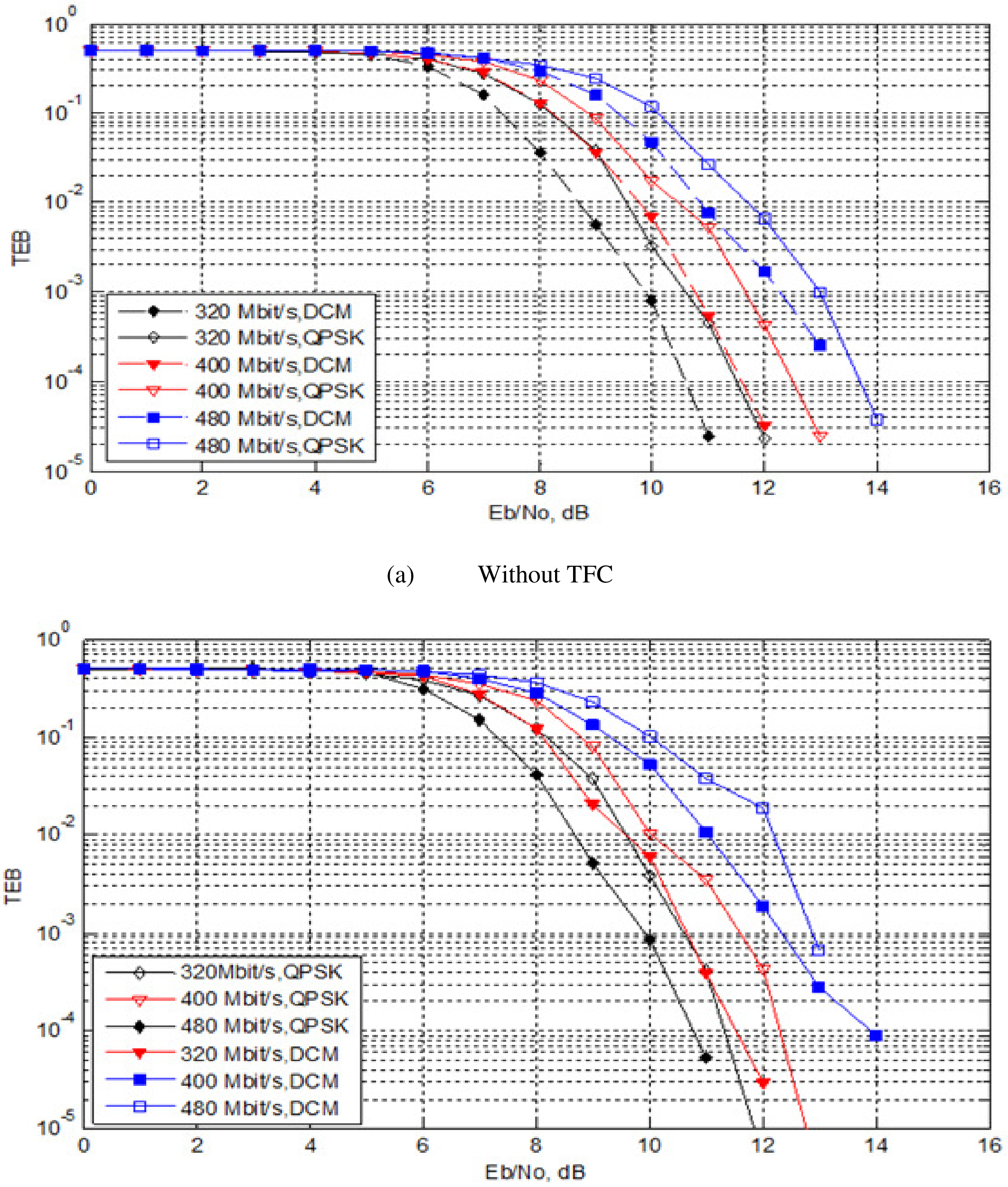
Therefore, the simulation results show that the DCM achieves lower BER compared to conventional QPSK.

Figure 4 shows that for a BER value of 10-3 a gain of about 0.7 dB is observed with a 320 Mbit/s throughput for both cases with or without TFC (Time Frequency Code).

According to these results, the DCM method is more efficient than QPSK. The advantages of DCM are retained even when the channel coding is applied.

The use of DCM provides the selection of an increased return code. DCM is not only robust, but also offers improvements in speed for a signal to noise ratio (SNR (signal to noise ratio)) given.

This improvement in performance is related to the characteristics of DCM, which provides additional diversity gain, first by spreading the symbol on two independent sub-carriers and secondly by the use of coding gain with constellation rearrangement.



(a) Without TFC

(b) With TFC

Figure 3: DCM versus QPSK for high data rates of 320 Mb/s, 400 Mb/s and 480 Mb/s Simulation in the case of CM1 (LOS) without TFC n°5 and with TFC n°1.

5. CONCLUSION

In this paper, a performance comparison between DCM and QPSK modulation was evaluated in a UWB system MBOFDM.

This evaluation shows the importance of using the new technology DCM only for high flow rates for both cases with and without TFC in MBOFDM system. Therefore, QPSK modulation is replaced by a multidimensional modulation DCM.

REFERENCES

[1] Federal Communication Commission, Revision of Part 15 of the Commission’s Rules Regarding Ultra-Wideband Transmission Systems, First Report and Order, ET Docket 98–153, FCC 02-48, Feb. 2002.

[2] D. Porcino and W. Hirt, "Ultra-wideband radio technology: Potential and challenges ahead, " IEEE Communication Magazine, vol. 41, pp.66-74, 2003.

[3] "MultiBand OFDM Physical Layer Specification Release 1.0," 2005

[4] ECMA, “High rateUltra Wideband PHY and MAC Standard “. Rapport ECMA-368 3nd edition, ECMA International, Décembre 2008.

[5] A. Batra et al., “Multi-band OFDM Physical Layer Proposal for IEEE 802.15 Task Group 3a “. IEEE P802.15-04/0493r1, Septembre 2004.

[6] G. F. Ross, “Transmission and reception system for generating and receiving base-band duration pulse signals for short base-band pulse communication system“, U.S. Patent 3,728,025 dated July 31, 1973.

[7] M. Hajjaj, Walid Chainbi, Ridha Bouallegue, “A Rhombic-DCM Constellation for MB-OFDM UWB Systems “,IEEE International Wireless Communica-tions & Mobile Computing (IWCMC), Dubrovnik ,p.256 - 261, Aug. 2015.

INTENTIONAL BLANK

WAVEFORM COMPARISON AND NONLINEARITY SENSITIVITIES OF FBMC, UFMC AND W-OFDM SYSTEMS

Changyoung An, Byeongjae Kim, and Heung-Gyoon Ryu

Department of Electronics Engineering, Chungbuk National University,
Cheongju, Korea 361-763
acy890217@naver.com, bj5236@nate.com, and ecomm@cbu.ac.kr

ABSTRACT

Recently, new waveforms for the 5th generation cellular system have been studied in many ways. UFMC, FBMC (filter bank multi-carrier) and W-OFDM (window orthogonal frequency division multiplexing) waveforms are very strong candidates as a new waveform for 5G system. In this paper, we have evaluated the spectrum characteristic and BER performance of the waveforms under the effect of nonlinear HPA. Also, we like to show the comparison of the time-frequency resources of each system because it would be very important to estimate the spectral efficiency and communication throughput. As simulation results, it is confirmed that OOB power of each system increases, and OOB power increase of FBMC system is the biggest. Additionally, we have confirmed that performance of every system is degraded by strength of HPA nonlinearity, and every system needs the PAPR reduction method for the nonlinear distortion compensation and power saving, even though it would be more complicated. Comparison table for the time-frequency resources requirement for the each modulation systems is included.

KEYWORDS

new waveform; OFDM; FBMC; UFMC; HPA nonlinearity

1. INTRODUCTION

The mobile traffic is being increased dramatically, because various mobile devices and multimedia services are being increased [1].Also, the growth of mobile traffic is being accelerated. It is difficult for the present mobile communication to support the mobile traffic required in the future [2].In order to solve the problem, studies for next generation 5G mobile communication has been carried actively [3-4].

Conventional orthogonal frequency division multiplexing (OFDM) based on multi-carrier has high-power out-of-band (OOB) [5]. This characteristic causes adjacent channel interference (ACI). OFDM uses a wide guard band in order to avoid ACI. It decreases spectral efficiency when a number of mobile devices simultaneously access a base station. Next generation mobile communication system requires high-level key performance indicators (KPIs). It is difficult for OFDM to satisfy the KPIs. Universal filtered multi-carrier (UFMC) and filter bank based multi-carrier (FBMC) are known as the candidate waveform for 5G mobile communication. When the

Natarajan Meghanathan et al. (Eds) : NETCOM, NCS, WiMoNe, GRAPH-HOC, SPM, CSEIT - 2016
pp. 83– 90, 2016. © CS & IT-CSCP 2016 DOI : 10.5121/csit.2016.61508

f-OFDM suggested by Huawei appeared in the first place, the filtered-OFDM system adopted one-filter system for the sharper OOB (out-of-band) spectrum characteristics, but they changed into the multiple filter system, which became very similar to the UFMC(universal filtered multi-carrier) system. These systems use filtering technique based on multi-carrier. These techniques have characteristic of low OOB power in comparison with conventional OFDM. Therefore, these systems have high spectrum efficiency. FBMC uses a filtering technique in each sub-carrier. UFMC uses a filtering technique in each sub-band [8-9].

However, these systems based on OFDM are vulnerable to non-linearity of high-power amplifier (HPA), like OFDM. OFDM has high peak-to-average power ratio (PAPR) because multi-carrier signals are overlapped. High PAPR causes nonlinear distortion in HPA because it saturates HPA. Similarly, UFMC and FBMC have high PAPR because these systems are based on multi-carrier [10-11]. In UFMC and FBMC system, if nonlinear distortion is caused by high PAPR, OOB power of these systems is increased. That is, advantage of these systems vanishes. Therefore, this drawback should be overcome in the candidate techniques for 5G mobile communication.

In this paper, in order to overcome the drawback, we focus on spectrum characteristic analysis and performance evaluation of FBMC and UFMC system under the effect of nonlinear HPA. Firstly, we describe and explain OFDM, UFMC, FBMC system. And then, we design the systems. Next, under linear environment, we analyse the spectrum characteristic of each system and evaluate bit error rate (BER) performance of each system. And then, under the effect of nonlinear HPA, we analyse spectrum characteristic of each system and evaluate bit error rate (BER) performance of each system.

Also, we like to show the comparison of the time-frequency resources of each system because it would be very important to estimate the spectral efficiency and communication throughput.

2. System Model

2.1. OFDM

In OFDM system, firstly, in transmitter of OFDM system, the data symbols are transformed into parallel stream from series stream by S/P block. The changed symbols are mapped onto each subcarrier by inverse fast Fourier transform (IFFT) operation. After IFFT operation, the time-domain signals are transformed into series stream from parallel stream by P/S block. And then, cyclic prefix (CP) is added in order to reduce the effect of inter-symbol interference (ISI). Finally, the RF signals are amplified by high-power amplifier (HPA). Receiver of OFDM system consists of reversed structure in comparison with OFDM transmitter. Additionally, in OFDM receiver, an equalizer is used in order to restore desired signal. The equalizer is very simple because of CP. OFDM receiver uses one-tap equalizer. OFDM system requires simple equalizer with one tap [12]. However, each subcarrier of OFDM system has high side-lobe power. As a result, channel capacity is decreased in OFDM system [12].

2.2. UFMC

UFMC filters each sub-band that consists of orthogonal multi-carrier in order to reduce OOB power [6]. In the UFMC system, each sub-band signal is transformed into series stream by P/S. Secondly, in UFMC receiver, the received signal is applied to RF chain. The received signal is transformed into baseband signal by RF chain. Baseband signal is converted into digital signal by ADC. And then, time-domain pre-processing is processed. After the process, the series data stream is transformed into a parallel data stream by S/P. The time-domain parallel data stream is converted to frequency-domain stream by 2N-FFT operation [6]. After 2N-FFT operation, oddnumbered data symbols are selected and equalized. Spectrum of UFMC system has lower OOB power in comparison with spectrum of OFDM system. This is good advantage. However, because UFMC system uses multi-carriers and multi-carriers are overlapped, UFMC system has high PAPR. High PAPR characteristic can distort signal of UFMC system [6].

2.3. FBMC

FBMC system filters each sub-carrier in order to reduce OOB power of spectrum [7]. In FBMC system, firstly, in transmitter of FBMC system, data symbols are transformed into parallel stream from series stream by S/P. The parallel symbols are modulated to offset quadrature amplitude modulation (OQAM) signal [7]. The modulated OQAM signal is transformed into a signal filtered by each sub-carrier by using the synthesis filter bank that consists of IFFT and poly phase network (PPN) [7]. Finally, the amplified FBMC signal is transmitted by antenna. Receiver of FBMC system consists of reversed structure in comparison with FBMC transmitter. FBMC system has lower OOB power in comparison with UFMC system and OFDM system. This is a good advantage. However, FBMC system has high system complexity. Additionally, because FBMC system uses multi-carrier, it has high PAPR.

2.4. W-OFDM

W-OFDM is a improved version of OFDM system. In the W-OFDM system, it does not use the filter but it uses the extension and windowing method on each OFDM symbol in order to reduce OOB power of spectrum.

2.5. HPA nonlinearity

In this paper, purposes are spectrum characteristic analysis and performance evaluation of OFDM, UFMC, FBMC and W-OFDM system under the effect of nonlinear HPA. Therefore, we have designed each system. In Saleh model, characteristics of AM-AM and AM-PM are as follows [13].

$$G[A(t)] = \frac{\alpha_A A(t)}{1+\beta_A A(t)^2} \tag{1}$$

$$\Phi[A(t)] = \frac{\alpha_\Phi A(t)^2}{1+\beta_\Phi A(t)^2} \tag{2}$$

Equation (1) shows AM-AM characteristic of Saleh model, nonlinear HPA model. A is amplitude of input signal. α_A and β_A are coefficients for adjusting amplitude of output signal. Equation (2) shows AM-PM characteristic of Saleh model. α_Φ and β_Φ are coefficients for adjusting phase of output signal.

The following formatting rules must be followed strictly. This (.doc) document may be used as a template for papers prepared using Microsoft Word. Papers not conforming to these requirements may not be published in the conference proceedings.

3. Simulation Results

Table 1 shows simulation parameters.

Table 1. Simulation parameters.

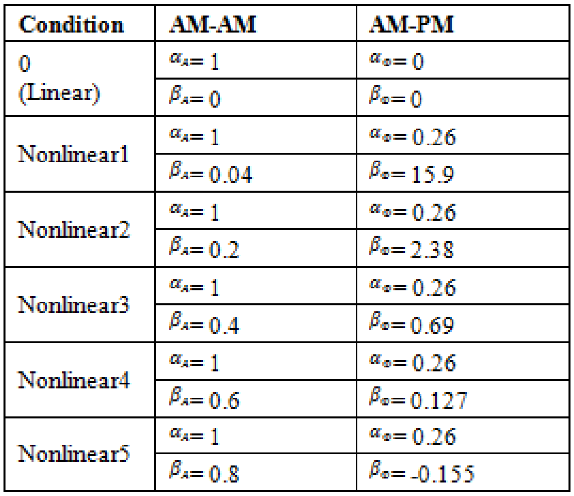
Parameter	Value
Modulation	QPSK
# of total subcarrier	64
# of used subcarrier	32
# of null subcarrier	32

Filter for FBMC	Phydyas prototype
	H0 = 1
	H1 = 0.97196
	H2 = 0.7071
	H3 = 0.235147

Filter for UFMC	Chebyshev Attenuation = 60dB,Length = 10
# of sub-band in UFMC	64/8
# of used sub-band in UFMC	4

Table 2 shows the considered HPA nonlinear conditions. Condition 0 is linear. Conditions 1 to 5 are nonlinear condition. Condition 1 is weak nonlinear condition. Condition 5 is strong nonlinear condition.

Table 2. Condition of HPA nonlinearity.



Condition	AM-AM	AM-PM
0 (Linear)	$\alpha_A = 1$	$\alpha_\Phi = 0$
	$\beta_A = 0$	$\beta_\Phi = 0$
Nonlinear1	$\alpha_A = 1$	$\alpha_\Phi = 0.26$
	$\beta_A = 0.04$	$\beta_\Phi = 15.9$
Nonlinear2	$\alpha_A = 1$	$\alpha_\Phi = 0.26$
	$\beta_A = 0.2$	$\beta_\Phi = 2.38$
Nonlinear3	$\alpha_A = 1$	$\alpha_\Phi = 0.26$
	$\beta_A = 0.4$	$\beta_\Phi = 0.69$
Nonlinear4	$\alpha_A = 1$	$\alpha_\Phi = 0.26$
	$\beta_A = 0.6$	$\beta_\Phi = 0.127$
Nonlinear5	$\alpha_A = 1$	$\alpha_\Phi = 0.26$
	$\beta_A = 0.8$	$\beta_\Phi = -0.155$

Table 3. Comparison of OOB power characteristic.

Condition	OFDM	UFMC	FBMC	W-OFDM
Linear	-26 dB	-83 dB	-120 dB	-90 dB
condition 1	-26 dB	-82 dB	-85 dB	-85 dB
condition 2	-26 dB	-74 dB	-75 dB	-75 dB
condition 3	-26 dB	-66 dB	-67 dB	-69 dB
condition 4	-26 dB	-61 dB	-62 dB	-63 dB

Table 3 shows OOB power comparison about each system. In this table, we have confirmed as follows. Under the HPA nonlinearity environment FBMC system shows the biggest change of OOB power, and OFDM system shows the smallest change of OOB power.

Figs. 1 to 4 show BER performances of each system. Each system has ideal performance under the linear condition or nonlinear condition. Under the nonlinear HPA environment, BER performance of every system is degraded. Additionally, FBMC system shows the smallest degradation of BER performance. However, even though FBMC system is the strongest against HPA nonlinearity, every system needs the PAPR reduction method for the nonlinear distortion compensation and power saving.

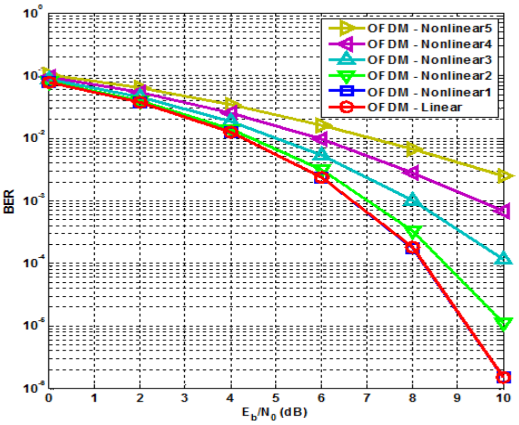


Figure 1. BER of OFDM system according in HPA conditions.

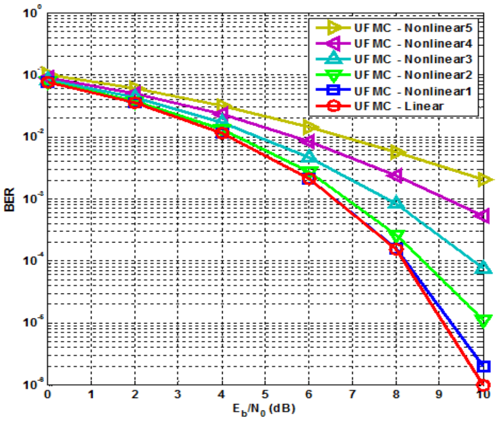


Figure 2. BER of UFMC system according in HPA conditions.

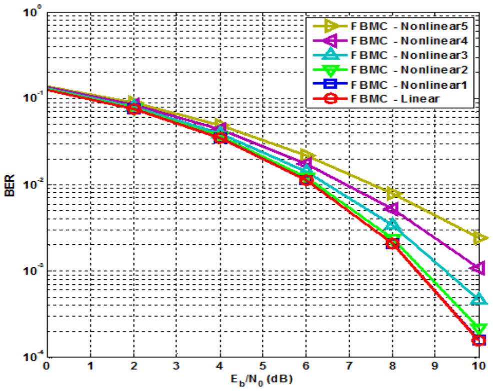


Figure 3. BER of FBMC system according in HPA conditions.

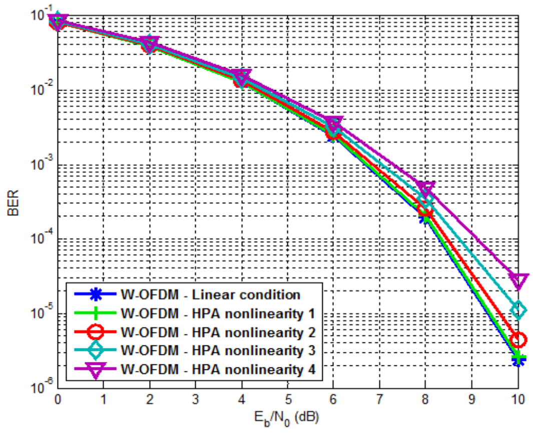


Figure 4. BER of W-OFDM system in HPA conditions.

Next, in order to compare the time-frequency resources of each candidates system, we have set some necessary conditions as in the below. Also, in the Table 4, we provide the comparison for the time-frequency resources requirement for the each modulation systems.

Allocated bandwidth = 20MHz
# of used sub-carriers = 16
# of transmission bits = 128
4QAM modulation(2bit) * 16 sub-carrier * 4 synthesis symbols = 128 bits
iFFT size = 64, CP length = 9
FBMC, Overlapping Factor (K) = 4
W-OFDM, Extension length = 6
OOB emission suppression (Frequency, 7.5Mhz Offset) / TTI length (Time)

Table 4. Comparison of the time-frequency resources.

	OFDM	UFMC	FBMC	W-OFDM
Linear	-26 dBc/ 292	-83 dBc / 292	-130 dBc / 480	-66 dBc/ 304
HPA condition 1	-26 dBc / 292	-82 dBc / 292	-85 dBc / 480	-66dBc/ 304
HPA condition 2	-26 dBc / 292	-74 dBc / 292	-75 dBc / 480	-66dBc/ 304
HPA condition 3	-26 dBc / 292	-66 dBc / 292	-67 dBc / 480	-65dBc/ 304
HPA condition 4	-26 dBc / 292	-63 dBc / 292	-65 dBc / 480	-63dBc/ 304

4. CONCLUSIONS

FBMC and UFMC systems are the strong modulation candidate for 5G mobile communication system. Since these systems are basically multicarrier system, it is important to study the nonlinearity sensitivity. In this paper, we have focused on spectrum characteristic analysis and BER performance evaluation of OFDM, FBMC, and UFMC system under the effect of nonlinear HPA. As simulation results, we have confirmed that if HPA nonlinearity rises in each system, OOB power of each system increases. The OOB power increase of FBMC system is the biggest. Additionally, we have confirmed that performance of every system is degraded by strength of HPA nonlinearity, and every system needs the PAPR reduction method for the nonlinear distortion compensation and power saving, even though it would be more complicated. Also, we like to show the comparison of the time-frequency resources of each system because it would be very important to estimate the spectral efficiency and communication throughput. We provide the comparison table for the time-frequency resources requirement for the each modulation systems.

ACKNOWLEDGEMENTS

“This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology(No.2013R1A2A2A01005849) and This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government(MSIP) (No. R0101-16-244, Development of 5G Mobile Communication Technologies for Hyper-connected Smart Services)”

REFERENCES

[1] Shanzhi Chen; Jian Zhao, "The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication," Communications Magazine, IEEE , vol. 52, no. 5, pp. 36-43, May 2014.

[2] Dahlman, E.; Mildh, G.; Parkvall, S.; Peisa, J.; Sachs, J.; Selén, Y.; Sköld, J., "5G wireless access: requirements and realization," Communications Magazine, IEEE, vol. 52, no. 12, pp. 42-47, December 2014.

[3] G. Wunder et al., “5GNOW: non-orthogonal, asynchronous waveforms for future mobile applications”, IEEE Commun. Mag., vol. 52, no. 2, pp. 97-105, Feb. 2014.

[4] P. Banelli et al., “Modulation Formats and Waveforms for the Physical Layer of 5G Wireless Networks: Who Will be the Heir of OFDM?”, in arXiv:1407.5947, July 2014.

[5] Schaich, F.; Wild, T., "Waveform contenders for 5G — OFDM vs. FBMC vs. UFMC," Communications, Control and Signal Processing (ISCCSP), 2014 6th International Symposium on, pp. 457-460, 21-23 May 2014.

[6] Vakilian, V.; Wild, T.; Schaich, F.; ten Brink, S.; Frigon, J.-F., "Universal-filtered multi-carrier technique for wireless systems beyond LTE," in Globecom Workshops (GC Wkshps), 2013 IEEE, pp. 223-228, 9-13 Dec. 2013.

[7] Farhang-Boroujeny, B., "OFDM Versus Filter Bank Multicarrier," in Signal Processing Magazine, IEEE, vol. 28, no. 3, pp. 92-112, May 2011.

[8] Wonsuk Chung; Beomju Kim; Moonchang Choi; Hyungju Nam; Hyunkyu Yu; Sooyoung Choi; Daesik Hong, "Synchronization Error in QAM-Based FBMC System," in Military Communications Conference (MILCOM), 2014 IEEE, pp. 699-705, 6-8 Oct. 2014.

[9] Mukherjee, Mithun; Shu, Lei; Kumar, Vikas; Kumar, Prashant; Matam, Rakesh, "Reduced out-of-band radiation-based filter optimization for UFMC systems in 5G," in Wireless Communications and Mobile Computing Conference (IWCMC), 2015 International, pp. 1150-1155, 24-28 Aug. 2015.

[10] Kollar, Zs.; Varga, L.; Czimer, K., "Clipping-Based Iterative PAPR-Reduction Techniques for FBMC," in OFDM 2012, 17th International OFDM Workshop 2012 (InOWo'12); Proceedings of , pp. 1-7, 29-30 Aug. 2012.

[11] Chafii, M.; Palicot, J.; Gribonval, R., "Closed-form approximations of the PAPR distribution for Multi-Carrier Modulation systems," in Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European, pp. 1920-1924, 1-5 Sept. 2014.

[12] Elshirkasi, A.M.; Siddiqi, M.U.; Habaebi, M.H., "Generalized Discrete Fourier Transform Based Minimization of PAPR in OFDM Systems," in Computer and Communication Engineering (ICCCE), 2014 International Conference on, pp. 205-208, 23-25 Sept. 2014.

[13] P. Drotar, J. Gazda, D. Kocur, and P. Galajda, “MC-CDMA performance analysis for different spreading codes at HPA Saleh model,” 18th Int. Conf. Radioelektronika, pp. 1-4, Prague, Apr. 2008.

AUTHORS

Changyoung An was born in Chungbuk, Republic of Korea in 1989. He received the B.S. degree and M.S. degree in the department of electronic engineering from Chungbuk National University in February 2013 and February 2013, respectively. He is currently working toward the Ph.D degree at the department of Electronic Engineering, Chungbuk National University, Republic of Korea. His research interests include wireless communication system, signal processing, antenna technology and 5G mobile communication.

Byeongjae Kim was born in Gyeongbuk, Republic of Korea in 1992. He received the B.S. degree in the department of electronic engineering from Chungbuk National University in 2016. Now he is currently working toward the master’s degree at the department of Electronic Engineering, Chungbuk National University, Republic of Korea. His research interests include wireless communication system, signal processing.

Heung-Gyoon Ryu (M’88) was born in Seoul, Republic of Korea in 1959. He received the B.S. and M.S. and Ph.D. degrees in electronic engineering from Seoul National University in 1982, 1984 and 1989. Since 1988, he has been with Chungbuk National University, Korea, where he is currently Professor of Department of Electrical, Electronic and Computer Engineering in Chungbuk National University. And he worked as chief director of RICIC (research institute of computer, information communication center) in Chungbuk National University from March 2002 to Feb 2004. His main research interests are digital communication systems, communication circuit design, spread spectrum system and communication signal processing. Since 1999, he has worked as reviewer of the IEEE transaction paper. He received the ‘2002 ACADEMY AWARD’ from the Korea Electromagnetic Engineering Society, Korea. He received the “BEST PAPER AWARD” at the 4th International Conference on Wireless Mobile Communications (ICWMC 2008) Athens, Greece, July 27-Aug.1, 2008. Also, He received the “BEST PAPER AWARD” at the International Conference on Advances in Satellite and Space Communications (SPACOMM 2009), Colmar France, July 20-25, 2009.

INTERFERENCE SUPPRESSING RECEIVER TECHNIQUE FOR WIRELESS AD HOC NETWORKS

Sunho Park and Byonghyo Shim

Institute of New Media and Communication, Department of Electrical and
Computer Engineering, Seoul National University, Seoul, Korea
sunhopark@islab.snu.ac.kr, bshim@snu.ac.kr

ABSTRACT

Recent works on ad hoc network study have shown that achievable throughput can be made to scale linearly with the number of receive antennas even if the transmitter has a single antenna. In this paper, we propose a method pursuing robustness in ad hoc network system when the channel state information (CSI) of interferers is unavailable. The non-parametric linear minimum mean square error (MMSE) filter is exploited to achieve large fraction of the MMSE filter transmission capacity employing the perfect covariance matrix information. The key feature ingredient to make our approach effective is to exploit the autocorrelation of received signal, which obtains the covariance matrix information without transmission rate loss. From the numerical results, we show that the proposed scheme brings substantial transmission capacity gain over conventional MMSE filter using sample covariance matrix.

KEYWORDS

Ad hoc network, channel estimation, receiver technique, minimum mean square error

1. INTRODUCTION

In the decentralized wireless network (ad hoc network), multiple transmitter-receiver pairs transmit simultaneously without the benefit of fixed infrastructure. Due to the uncoordinated nature of communication, multiple transmitters communicate simultaneously, and hence incur substantial interference which severely decreases the transmission rate. In an ad hoc wireless network, exploiting multiple receive antennas has been one of the promising solutions to increase data rate and deal with inter-user interference. Prior works on investigating the transmission capacity with multiple receive antennas [1]–[3] considered some specific multiple antenna configurations. In [1] the maximal ratio combining (MRC) only provides array gain while [2] considers full zero forcing to remove the strongest interferers but no array gain is provided. It is shown in [3] that both conventional MMSE filter and partial zero forcing (PZF) provide a benefit that network-wide throughput scales linearly with the number of receive antennas even if each transmitter has only a single antenna. All these promising gains are achieved assuming perfect channel state information at the receiver (CSIR). Although perfect CSIR is reasonable for initial state of research, further investigation of imperfect CSIR would be interesting since practical network in general has imperfect channel state information due to estimation errors. Recently, as a way to deal with imperfect CSIR issue, Jindal et al. proposed the filter employing sample covariance matrix which can be estimated by listening to the interference and noise observations [3]. However, the shortcoming of this filter using sampled covariance matrix is that data rate loss

Natarajan Meghanathan et al. (Eds) : NETCOM, NCS, WiMoNe, GRAPH-HOC, SPM, CSEIT - 2016
pp. 91– 102, 2016. © CS & IT-CSCP 2016 DOI : 10.5121/csit.2016.61509

is substantial since the covariance matrix is sampled in an inactive mode of the desired transmitter.

In this paper, we propose a technique that employs nonparametric linear minimum mean square error (MMSE) receive filter for improving network-wide throughput. To be specific, when the CSIR of all the interferers is unavailable, which is true for real ad hoc network scenarios, the proposed scheme exploits the autocorrelation of the received signal for MMSE operation. Even with the covariance matrix variation, the maximized SINR of the proposed method is identical to that of the conventional filter. Due to the fact that the autocorrelation of the received signal is obtained the data transmission period, the proposed method achieves large fraction of optimal transmission capacity.

The rest of this paper is organized as follows. In Section 2, we describe the system model and the summary of the conventional MMSE filter. In Section 3, we present the proposed non-parameter linear MMSE method. Simulation results and conclusion are provided in Section 4 and Section 5.

2. AD HOC NETWORK

2.1. System Model

In an ad hoc network, the active transmitters are placed according to a 2-D homogeneous Poisson point process (PPP) of density λ ($transmitters/m^2$). Each single transmit antenna communicates with a receiver equipped with N antennas, where each receiver is randomly located at d meters away from the corresponding transmitter. Due to the Poisson process stationarity, we focus on a typical transmit-receive pair denoted by Tx_d and Rx_d, respectively. From the viewpoint of Rx_d, the set of interferers except Tx_d also sets a homogeneous PPP due to Slivnyak's Theorem [4]. The set of all the active transmitters is denoted by denoted by $\mathrm{A} = \{(X_i, \mathbf{h}_i), (d, \mathbf{h}_d), \lambda, i \in \mathbf{N}\}$ where X_i and $\mathbf{h}_i$ are the location and channel vector of the i th transmitting node with respect to the typical receiver.

Under the frequency-flat channel, the received signal $\mathbf{y}$ is

$$\mathbf{y} = d^{-\alpha/2}\mathbf{h}_d s_d + \sum_{i \in \mathrm{A}(\lambda)\setminus\{Tx_d\}} \left|X_i\right|^{-\alpha/2}\mathbf{h}_i s_i + \mathbf{w} \tag{1}$$

where $\alpha(>2)$ is a path-loss exponent, $\left|X_i\right|$ is the distance from the desired transmitter to the i th interferer, $\mathbf{h}_i$ is the channel vector $(N \times 1)$ from the i th interferer to the desired receiver, $\mathbf{w}$ is the complex Gaussian noise vector $(\mathbf{w} \sim CN(0, \sigma^2\mathbf{I}))$, and s_i is the symbol transmitted by the i th interferer $(E[|s_i|^2] = \rho)$. Without loss of generality, we assume that the distances $\left|X_i\right|$ are ordered so that the squared-distances $\left|X_1\right|^2$, $\left|X_2\right|^2$, ... follow the 1-D PPP with intensity $\pi\lambda$ [5].

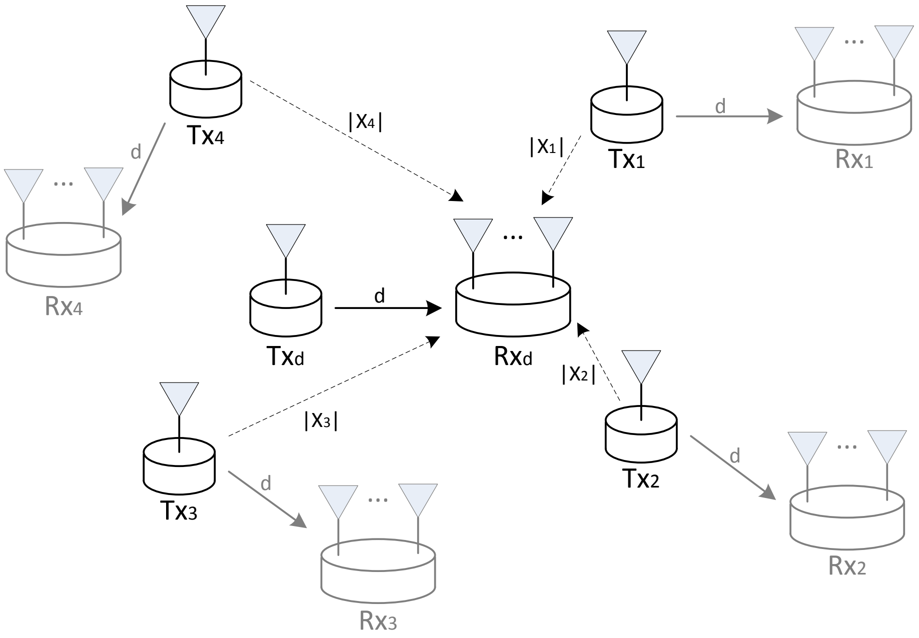


Figure 1. Desired Transmit-receive pair with interferers

2.2. Conventional MMSE Receiver

In this subsection, we describe the transmission capacity for the ad hoc network and review the conventional MMSE filter. With the inclusion of unit norm receive filter $\mathbf{v}_d$, the estimated symbol becomes $\hat{s}_d = \mathbf{v}_d^H \mathbf{y}$ and hence resulting signal-to-interference-and-noise ratio (SINR) is

$$SINR = \frac{\rho d^{-\alpha} \mathbf{v}_d^H \mathbf{h}_d \mathbf{h}_d^H \mathbf{v}_d}{\mathbf{v}_d^H (\sigma^2 \mathbf{I} + \rho \sum_{i \in \mathrm{A}(\lambda)} | X_i |^{-\alpha} \mathbf{h}_i \mathbf{h}_i^H) \mathbf{v}_d} . \tag{2}$$

When all transmitters send at the rate equal to $R = \log_2(1+\beta)$, a communication is regarded as successful if and only if the received SINR is larger than β. Hence the outage probability at SINR threshold β is $P_{out}(\lambda) = P[SINR \le \beta]$, which is an increasing function of λ. By the stationarity of the process, this outage probability approximates the network-wide packet error probability. Further, the maximum interferer density such that the outage does not exceed $\varepsilon (> 0)$ is $\lambda_\varepsilon = \max_\lambda \{\lambda : P_{out}(\lambda) \le \varepsilon\}$ where ε is a constant outage level, ensuring a typical transmission will succeed with probability $1-\varepsilon$. Then, the transmission capacity of the ad hoc network is

$$C(\varepsilon) = \lambda_\varepsilon (1-\varepsilon) \log_2(1+\beta) \qquad \text{bps/Hz}/m^2 , \tag{3}$$

by accumulating all the λ_ε simultaneous transmissions in the network [1], [6], [7].

Due to the fact that the SINR and λ_ε primarily depend on the receive filter, multiple receive antenna technique has received much attention as a means to mitigate interference [3], [7], [8]. It is well known that the MMSE filter optimally pursues balance between signal boost and interference suppression for maximizing the SINR [3]. The normalized MMSE receive filter is given by

$$V_d = \frac{\Sigma^{-1}\mathbf{h}_d}{\| \Sigma^{-1}\mathbf{h}_d \|} \tag{4}$$

where $\Sigma = \frac{1}{SNR}\mathbf{I} + d^{\alpha} \sum_{i \in \mathrm{A}(\lambda)\backslash Tx_d} | X_i |^{-\alpha} \mathbf{h}_i\mathbf{h}_i^H$ is the spatial covariance of the interference plus noise and $SNR = \frac{\rho d^{-\alpha}}{\sigma^2}$. Using (4) and (2), the maximized received SINR of the MMSE filter becomes

$$\begin{aligned} SINR_{MMSE} &= \frac{\rho d^{-\alpha}(\mathbf{h}_d^H\Sigma^{-1}\mathbf{h}_d)^2}{\mathbf{h}_d^H\Sigma^{-1}(\sigma^2\mathbf{I} + \rho \sum_{i \in \mathrm{A}(\lambda)} | X_i |^{-\alpha} \mathbf{h}_i\mathbf{h}_i^H)\Sigma^{-1}\mathbf{h}_d} \\ &= \mathbf{h}_d^H\Sigma^{-1}\mathbf{h}_d. \end{aligned} \tag{5}$$

When the CSIR of all the interferers is unavailable, the receiver should estimate the interfering channels information to design the optimal MMSE filter. Note that since the desired channel can be estimated accurately via pilot symbols, the primary concern of the ad hoc network is the interfering channels estimation. The MMSE with imperfect CSIR [3] estimates the sampled covariance matrix by listening to interferer transmissions in the absence of desired signal. If the desired transmitter remains inactive for K symbols duration, the receiver can employ the K observations to organize the sample covariance as

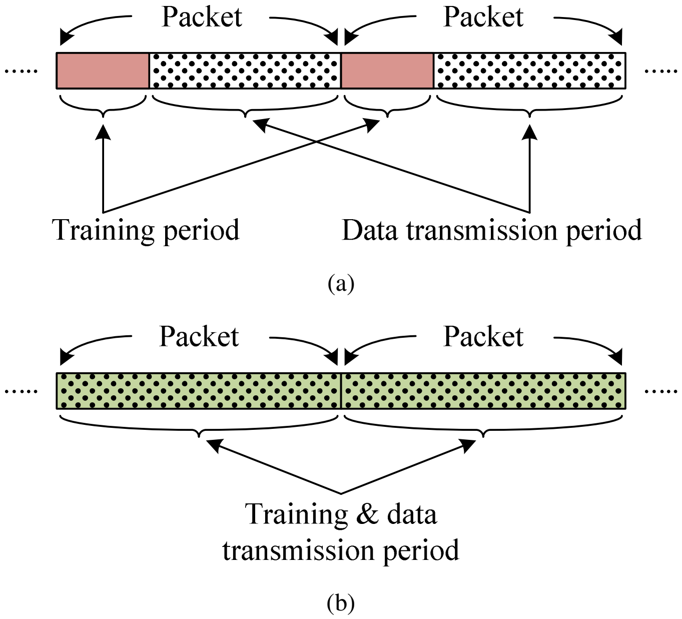
$$\hat{\Sigma} = \frac{1}{K}\sum_{i=1}^{K}\mathbf{r}_i\mathbf{r}_i^H \tag{6}$$

where $\mathbf{r}_i$ represents the i th observation including interference and noise. By replacing $\hat{\Sigma}$ with Σ in (4), the resulting SINR becomes $SINR = \frac{(\mathbf{h}_d^H\hat{\Sigma}^{-1}\mathbf{h}_d)^2}{\mathbf{h}_d^H\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\mathbf{h}_d}$. Under the assumption that all the interferers send independent Gaussian symbols, the expected SINR with respect to the $\hat{\Sigma}$ distribution is [9]

$$E\left[\frac{(\mathbf{h}_d^H\hat{\Sigma}^{-1}\mathbf{h}_d)^2}{\mathbf{h}_d^H\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\mathbf{h}_d}\right] = \left(1 - \frac{N-1}{K+1}\right)\mathbf{h}_d^H\Sigma^{-1}\mathbf{h}_d. \tag{7}$$

3. Non-Parametric Linear MMSE Receiver

In this section, we present the proposed filter based on the non-parametric linear MMSE estimation. There are three following drawbacks in the conventional MMSE receiver. First, when the receiver estimates K observations, the desired transmitter should be turned off for K symbols duration. The training duration will be substantial and cause the substantial transmission data rate loss even though the receiver provide good covariance matrix for sufficient K. Second, the covariance matrix information should be updated when the channel state is changed. If the channel is changing per T symbols period, the effective rate is decreased by the factor of $\frac{T-K}{T}$. Finally, K should increase to attain the identical SINR in (5) when the N increases (see Fig. 2).



(a)

(b)

Figure 2. Packet structure: (a) conventional method and (b) proposed method.

3.1. Non-Parametric Linear MMSE

The key distinction of the proposed method over the conventional MMSE filter is that the desired channel information is additionally incorporated on top of the observations of the interference and noise. From [10], the estimated desired symbol of the linear MMSE is given by

$$\begin{aligned}\hat{s}_d &= \mathbf{v}_d^H \mathbf{y} = R_{s_d \mathbf{y}} R_{\mathbf{yy}}^{-1} \mathbf{y} \\ &= \mathbf{h}_d^H \left(\frac{1}{SNR}\mathbf{I} + \sum_{i\in \mathrm{A}(\lambda)\backslash\{Tx_d\}} \mathbf{h}_i \mathbf{h}_i^H + \mathbf{h}_d \mathbf{h}_d^H \right)^{-1} \mathbf{y} \\ &= \mathbf{h}_d^H (\Sigma + \mathbf{h}_d \mathbf{h}_d^H)^{-1} \mathbf{y}.\end{aligned} \tag{8}$$

Following theorem explains the fact that the linear MMSE filter in (8) can achieve the maximum SINR of (5) regardless of the inclusion of the desired channel information.

Theorem 3.1: The linear MMSE filter using non-parametric autocorrelation of the received signal $R_{\mathbf{yy}}$ is

$$\mathbf{v}_d = (\Sigma + \mathbf{h}_d \mathbf{h}_d^H)^{-1} \mathbf{h}_d \tag{9}$$

and the corresponding SINR becomes

$$SINR = \mathbf{h}_d^H \Sigma^{-1} \mathbf{h}_d \tag{10}$$

Employing the Sherman-Morrison formula [10], the linear MMSE receiver with the autocorrelation of the received signal achieves the maximum transmission capacity. By modifying the Theorem 3.1, the SINR of the proposed algorithm under imperfect CSIR condition is obtained. The sample covariance of the proposed receiver is $\hat{\Sigma}_d = \hat{\Sigma} + \mathbf{h}_d \mathbf{h}_d^H$ where $\hat{\Sigma}$ represents the observations of the noise plus interference. With the knowledge of $\mathbf{h}_d$, the receiver can compute the filter $\hat{\mathbf{v}}_d = \hat{\Sigma}_d^{-1} \mathbf{h}_d$ and the resulting SINR becomes

Following theorem explains the fact that the linear MMSE filter in (8) can achieve the maximum SINR becomes

$$SINR_{prop} = \frac{(\mathbf{h}_d^H \hat{\Sigma}_d^{-1} \mathbf{h}_d)^2}{\mathbf{h}_d^H \hat{\Sigma}_d^{-1} \Sigma \hat{\Sigma}_d^{-1} \mathbf{h}_d}. \tag{11}$$

One can show that the expected SINR is

$$E\left[SINR_{prop}\right] = \left(1 - \frac{N-1}{M+1}\right) \mathbf{h}_d^H \Sigma^{-1} \mathbf{h}_d. \tag{12}$$

Note that the (12) based upon the sample covariance $\hat{\Sigma}_{prop}$ is precisely a factor of $1 - \frac{N-1}{M+1}$ smaller than the expected SINR with perfect knowledge of $R_{\mathbf{yy}}$. This factor is increasing in M and converges to one as $M \to \infty$ because $\hat{\Sigma}_{prop} \to \Sigma_{prop}$ as $M \to \infty$.

3.2 Alternative Form of the Non-Parametric Linear MMSE

Due to the fact that the exist of $R_{\mathbf{yy}}^{-1}$ is not always guaranteed, we provide the alternative form of the non-parametric linear MMSE filter. One can find that

$$\begin{aligned} \mathbf{v}_d^H &= \mathbf{h}_d^H (\Sigma^{-1} - \Sigma^{-1} \mathbf{h}_d (1 + \Sigma_h)^{-1} \mathbf{h}_d^H \Sigma^{-1}) \\ &= (\mathbf{I} - \Sigma_h (1 + \Sigma_h)^{-1}) \mathbf{h}_d^H \Sigma^{-1} \\ &= (1 + \Sigma_h)^{-1} \mathbf{h}_d^H \Sigma^{-1} \end{aligned} \tag{13}$$

Using the Eigen-decomposition [11]

$$\Sigma = \begin{bmatrix} \mathbf{U}_S & \mathbf{U}_N \end{bmatrix} \begin{bmatrix} \Lambda_S + \frac{1}{SNR}\mathbf{I} & 0 \\ 0 & \frac{1}{SNR}\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{U}_S^H \\ \mathbf{U}_N^H \end{bmatrix}, \tag{14}$$

then

$$\Sigma^{-1} = \mathbf{U}_S \left(\Lambda_S + \frac{1}{SNR}\mathbf{I} \right)^{-1} \mathbf{U}_S^H + SNR(\mathbf{U}_N \mathbf{U}_N^H). \tag{15}$$

Plugging (15) into (13), we have

$$\begin{aligned}\mathbf{v}_d^H &= \frac{1}{SNR}\left(\frac{1}{SNR}+\frac{1}{SNR}\Sigma_h\right)^{-1}\mathbf{h}_d^H\Sigma^{-1}\\ &=\left(\frac{1}{SNR}+\frac{1}{SNR}\mathbf{h}_d^H\mathbf{F}\mathbf{h}_d+\mathbf{h}_d^H\mathbf{P}_I^{\perp}\mathbf{h}_d\right)^{-1}\mathbf{h}_d^H\left(\frac{1}{SNR}F+\mathbf{P}_I^{\perp}\right)\end{aligned} \qquad (16)$$

where $\mathbf{F}=\mathbf{U}_S\left(\Lambda_S+\frac{1}{SNR}\mathbf{I}\right)^{-1}\mathbf{U}_S^H$ and $\mathbf{P}_I^{\perp}=\mathbf{U}_N\mathbf{U}_N^H$. By employing the singular value decomposition $\mathbf{h}_d^H\mathbf{P}_I^{\perp}=\mathbf{V}\Lambda\mathbf{U}^H$ and choosing $\mathbf{W}$ such that the matrix $[\mathbf{V}\quad\mathbf{W}]$ is unitary. Then, (16) can be resolved onto the basis $[\mathbf{V}\quad\mathbf{W}]$ as follows

$$\begin{aligned}\mathbf{v}_d^H &= [\mathbf{V}\quad\mathbf{W}]\begin{bmatrix}\frac{1}{SNR}(\mathbf{I}+\mathbf{V}^H\mathbf{h}_d^H\mathbf{F}\mathbf{h}_d\mathbf{V})+\Lambda^2 & 0\\ 0 & \frac{1}{SNR}\mathbf{I}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{V}^H\\ \mathbf{W}^H\end{bmatrix}\mathbf{h}_d^H\left(\frac{1}{SNR}\mathbf{F}+\mathbf{P}_I^{\perp}\right)\\ &=\mathbf{V}\left(\frac{1}{SNR}\left(\mathbf{I}+\mathbf{V}^H\mathbf{h}_d^H\mathbf{F}\mathbf{h}_d\mathbf{V}\right)+\Lambda^2\right)^{-1}\mathbf{V}^H\mathbf{h}_d^H\left(\frac{1}{SNR}\mathbf{F}+\mathbf{P}_I^{\perp}\right)\end{aligned} \qquad (17)$$

since $\mathbf{W}^H\mathbf{h}_d^H=0$ [11]. If we consider the interference-limited regime $\left(\frac{1}{SNR}\to 0\right)$, we have

$$\mathbf{v}_d^H=\mathbf{V}\Lambda^{-2}\mathbf{V}^H\mathbf{h}_d^H\mathbf{P}_I^{\perp}=\mathbf{V}\Lambda^{-1}\mathbf{U}^H=(\mathbf{P}_I^{\perp}\mathbf{h}_d)^{\dagger}=(\mathbf{h}_d^H\mathbf{P}_I^{\perp}\mathbf{h}_d)^{-1}\mathbf{h}_d^H\mathbf{P}_I^{\perp}. \qquad (18)$$

By letting $R_{\mathbf{yy}}=R_A+\frac{1}{SNR}\mathbf{I}=AA^H+\frac{1}{SNR}\mathbf{I}$ where $R_A=\mathbf{h}_d\mathbf{h}_d^H+\sum\mathbf{h}_i\mathbf{h}_i^H$ and $A=[\mathbf{h}_d\quad\mathbf{h}_1\quad\Lambda\quad\mathbf{h}_n]$ [11]. Therefore, (18) can be expressed as

$$\mathbf{v}_d^H=(\mathbf{h}_d^H R_A^{\dagger}\mathbf{h}_d)^{-1}\mathbf{h}_d^H R_A^{\dagger}. \qquad (20)$$

Since $R_{\mathbf{yy}}^{\dagger}=R_A^{\dagger}$ in the interference-limited regime, (9) can be reduced to

$$\mathbf{v}_d=\hat{R}_{\mathbf{yy}}^{\dagger}\mathbf{h}_d(\mathbf{h}_d^H\hat{R}_{\mathbf{yy}}^{\dagger}\mathbf{h}_d)^{-1} \qquad (21)$$

where $\hat{R}_{\mathbf{yy}}=\frac{1}{M}\mathbf{Y}\mathbf{Y}^H$ is the sample correlation matrix obtained from the received signal set $\mathbf{Y}=[\mathbf{y}_1\quad\mathbf{y}_2\quad\Lambda\quad\mathbf{y}_M]$.

Note that the desired transmitter should be turned off for K symbols period in order to attain fairly good sample covariance matrix. Also note that no such requirement is necessary for the proposed approach in (21). While K is the sampling overhead in the packet transmission, M in (21) can be freely selected within the range of the packet length.

4. SIMULATION AND DISCUSSION

In this section, we compare the transmission capacity of the proposed technique with the conventional MMSE, MMSE with imperfect CSI (MMSE with K samples), as well as maximum ratio combining (MRC) and zero forcing (ZF) schemes (full ZF and partial ZF). While the MRC maximizes the desired signal power and the full ZF selects the filter orthogonal to $N-1$ interferer channels, the partial ZF employs some of the receive degrees of freedom for interference annihilation while exploiting the rest of degrees of freedom to boost desired signal power (readers are referred to [3] for details).

The simulation setup is based on the 2-D PPP transmitters which are realized on the square distances. The SINR and outage probability are computed and compared to determine maximum density over several thousand iterations. We assume that the elements of each transmitter's channel vectors are i.i.d. zero mean complex Gaussian random variables with unit variance, which almost surely ensures that the desired channel $\mathbf{h}_d$ and interferer channels $\{\mathbf{h}_i\}_{i=1}^{\Lambda(\lambda)}$ have full column rank. For comparison, the number of cancelled interferers θN is considered in the partial ZF [3].

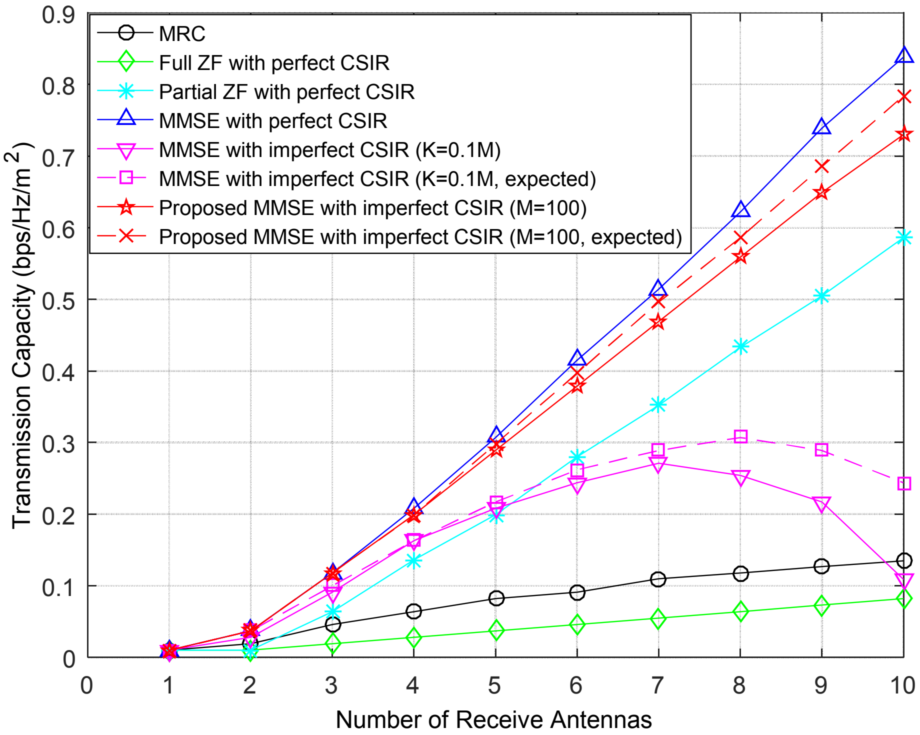


Figure 3. Transmission Capacity versus N for $\varepsilon = 0.1$, $\beta = 1$, $\alpha = 3$, $d = 1$, and $K = 10$.

In Fig. 3, we plot the transmission capacity as a function of N. Note that $K = 10$ (10% of packet length) and M is the packet length. We observe that the proposed scheme, MMSE, and PZF show linearly increasing transmission capacity, whereas MMSE with K samples, MRC, and full ZF exhibit much poorer scaling. In particular, although the proposed filter leaves a performance gap from MMSE, the transmission capacity of the proposed method is larger than the MMSE with K samples and PZF, and the gain gets larger as N increases. Due to the scaling factor of

$\left(1-\frac{N-1}{K+1}\right)$, the expected SINR of the MMSE with K samples is smaller than that of the MMSE with full CSI, the transmission capacity of the MMSE with K samples is decreased when K is a fixed number and N goes to large number.

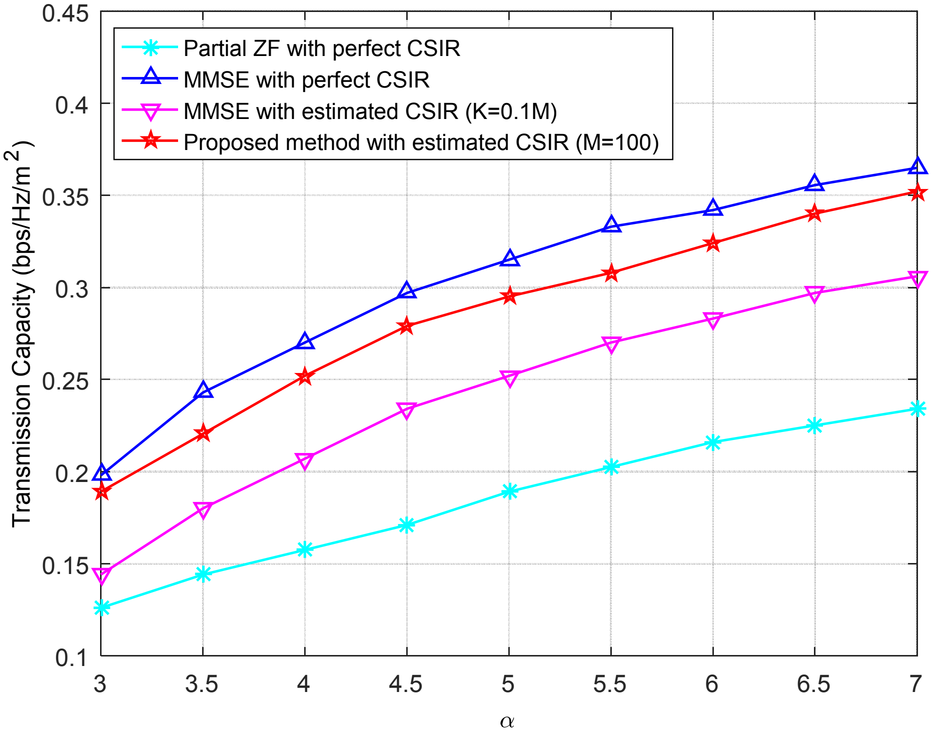


Figure 4. Transmission Capacity versus α for $\varepsilon = 0.1$, $\beta = 1$, $d = 1$, $N = 6$, and $SNR = 10dB$.

In Fig. 4, we plot the transmission capacity as a function of path loss exponent. The transmission capacity increases with the path loss exponent due to the fact that while the quality of the desired Tx-Rx pair is reduced by a higher α, the effect of interference is also decreased. Note that the proposed method outperforms the MMSE with K samples and partial ZF at all path loss exponent regime. This result shows that this interference degradation has a more significant effect on the ad hoc network system.

Finally, in order to solidify our conclusions, we plot the transmission capacity as a function of the number of blocks. From Fig. 5, we observe that the transmission curves of the MMSE and partial ZF are consistent due to full CSI structure. On the contrary, the transmission capacities of the proposed method and MMSE with K samples are increasing function of the number of blocks since the accuracy of the covariance matrix is increased when the number of blocks grows. Although the transmission capacity of the MMSE with K samples is close to the that of the optimal MMSE filter, it is impractical since the transmission data rate loss also increases. The results indicate that the proposed method is competitive option in real ad hoc network scenarios.

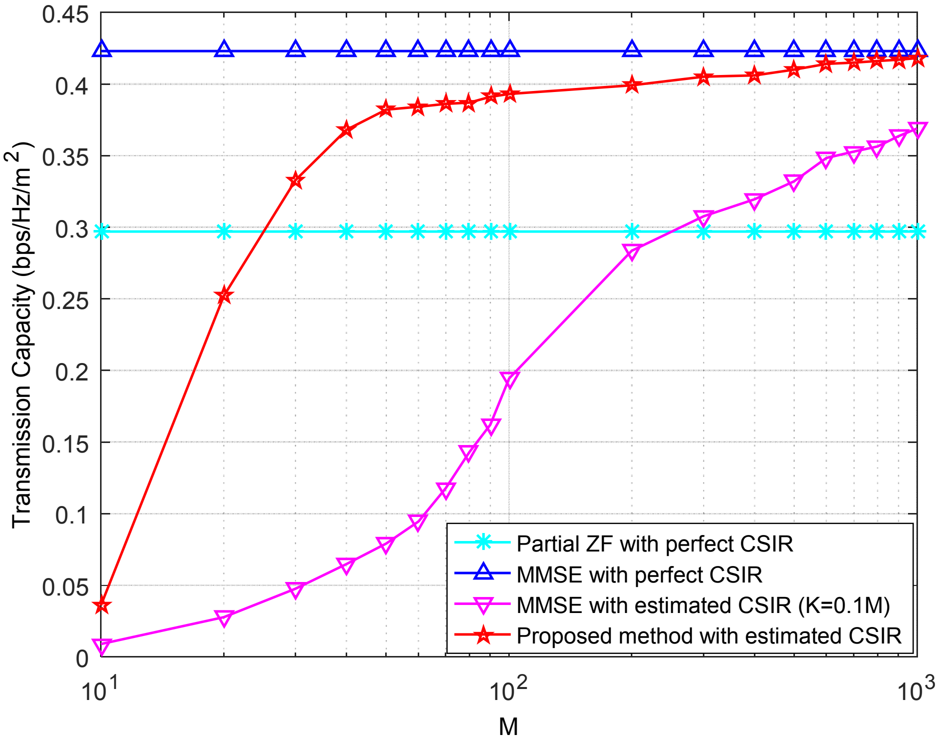


Figure 5. Transmission Capacity versus number of samples for $\varepsilon = 0.1$, $\beta = 1$, $\alpha = 3$, $d = 1$, $N = 6$, and $SNR = 10dB$.

5. Conclusion

In this paper, we investigated an approach based on MMSE filter achieving robustness of real ad hoc network in which the CSIR of all interferers is unavailable. Motivated by the fact that the MMSE with imperfect CSIR brings significant transmission rate loss due to the inactive mode of the desired transmitter, we employed the non-parametric linear MMSE filter to achieve large fraction of the MMSE filter transmission capacity. We observe from the transmission capacity performance that the proposed method outperforms the MMSE receiver under imperfect CSIR condition and conventional receive antenna algorithms. Future study needs to be directed towards the investigation of the performance when the interfering transmitters are heterogeneous.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government(MSIP)(2014R1A5A1011478) and the ICT R&D program of MSIP/IITP[B0717-16-0023, Fundamental technology development of transmission, modulation, and coding techniques for low-power and low-complexity 5G-based IoT environments with massive connectivity].

REFERENCES

[1] A. Hunter, J. G. Andrews, and S. Weber, “The transmission capacity of ad hoc networks with spatial diversity,” IEEE Trans. Wireless Commun., vol. 7, no. 12, pp. 5058-5071, Dec. 2008.

[2] K. Huang, J. G. Andrews, D. Guo, R. W. Heath, Jr., and R. Berry, “Spatial interference cancellation for multi-antenna mobile ad hoc networks,” IEEE Trans. Inf. Theory, submitted. [Online]. Available:arxic.org/abs/0807.1773v2

[3] N. Jindal, J. G. Andrews, and S. weber, “Multi-antenna communication in ad hoc networks: achieving MIMO gains with SIMO transmission,” IEEE Trans. Comm., vol. 59, no. 2, pp. 529-540, Feb. 2011.

[4] S. Weber, J. G. Andrews, and N. Jindal, “An overview of the transmission capacity of wireless networks,” IEEE Trans. Comm., vol. 58, no. 12, pp. 3593-3604, Dec. 2010.

[5] M. Haenggi, “On distances in uniformly random networks,” IEEE Trans. Inf. Theory, vol. 51, no. 10, pp. 3584-3586, Oct. 2005.

[6] S. Weber, X. Yang, J. G. Andrews, and G. de Veciana, “Transmission capacity of wireless ad hoc networks with outage constraints,” IEEE Trans. Inf. Theory, vol. 51, no. 12, pp. 4091-4102, Dec. 2005.

[7] J. Blomer and N. Jindal, “Transmission capacity of wireless ad hoc networks: successive interference cancellation vs. joint detection,” Proc. IEEE Intl. Conf. Commun. (ICC), Dresden, Germany, June 2009.

[8] S. Govindasamy, D. W. Bliss, and D. H. Staelin, “Spectral efficiency in single-hop ad-hoc network wireless netowrks with interference using adaptive antenna arrays,” IEEE J. Sel. Areas Commun., vol. 25, no. 7, pp. 1358-1369, Sep. 2007.

[9] I. Reed, J. Mallet, and L. Brennan, “Rapid convergence rate in adaptive arrays,” IEEE Trans. Aerospace Electron. Syst., vol. 10, no. 6, pp. 853-863, Nov. 1974.

[10] S.\ M.\ Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, Prentice Hall, 1998.

[11] L. Scharf and M. McCloud, “Blind adaptation of zero forcing projections and oblique pseudo-inverses for subspace detection and estimation when interference dominates noise,” IEEE Trans. Sig. Proc., vol. 50, no. 12, pp. 2938-2946, Dec. 2002.

AUTHORS

Sunho Park received the B.S., M.S., and Ph.D. degrees from the School of Information and Communication from Korea University, Seoul, in 2008, 2010, and 2015, respectively. From September 2015 to September 2016, he was with Institute of New Media and Communications, Seoul National University, as a Senior Researcher. He is currently a research assistant professor in Seoul National University, Seoul, Korea. His research interests include wireless communications and signal processing.

Byonghyo Shim received the B.S. and M.S. degrees in control and instrumentation engineering from Seoul National University, Korea, in 1995 and 1997, respectively. He received the M.S. degree in mathematics and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign (UIUC), USA, in 2004 and 2005, respectively.

From 1997 and 2000, he was with the epartment of Electronics Engineering, Korean Air Force Academy as an Officer (First Lieutenant) and an Academic Full-time Instructor. From 2005 to

2007, he was with Qualcomm Inc., San Diego, CA, USA, as a Staff Engineer. From 2007 to 2014, he was with the School of Information and Communication, Korea University, Seoul, as an Associate Professor. Since September 2014, he has been with the Department of Electrical and Computer Engineering, Seoul National University, where he is presently an Associate Professor. His research interests include wireless communications, statistical signal processing, estimation and detection, compressive sensing, and information theory.

Dr. Shim was the recipient of the 2005 M. E. Van Valkenburg Research Award from the Electrical and Computer Engineering Department of the University of Illinois and 2010 Hadong Young Engineer Award from IEIE. He is currently an Associate Editor of the IEEE WIRELESS COMMUNICATIONS LETTERS, Journal of Communications and Networks, and a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (JSAC).

A DYNAMIC ROUTE DISCOVERY SCHEME FOR HETEROGENEOUS WIRELESS SENSOR NETWORKS BASED ON POLYCHROMATIC SETS THEORY

Dong Wang[1], Xinheng Wang[1] and Hong-Hsu Yen[2]

1School of Engineering and Computing,
University of the West of Scotland, Paisley, Scotland, UK
dong.wang@uws.ac.uk, xinheng.wang@uws.ac.uk
2Department of Information Management,
Shih Hsin University, Taipei 116, Taiwan
hhyen@cc.shu.edu.tw

ABSTRACT

With the development of new networking paradigms and wireless protocols, nodes with different capabilities are used to form a heterogeneous network. The performance of this kind of networks is seriously deteriorated because of the bottlenecks inside the network. In addition, because of the application requirements, different routing schemes are required toward one particular application. This needs a tool to design protocols to avoid the bottlenecked nodes and adaptable to application requirement. Polychromatic sets theory has the ability to do so. This paper demonstrates the applications of polychromatic sets theory in route discovery and protocols design for heterogeneous networks. From extensive simulations, it shows the nodes with high priority are selected for routing, which greatly increases the performance of the network. This demonstrates that a new type of graph theory could be applied to solve problems of complex networks.

KEYWORDS

Dynamic routing, heterogeneous networks, wireless sensor networks, polychromatic sets

1. INTRODUCTION

Wireless communications play an important role in modern communications because of excellent performance and capability in mobility. With the development of wireless protocols and the new paradigm of networking, sensor nodes equipped with different protocols may join together to form a heterogeneous network. Because of the difference in wireless protocols and the ability of communications of these heterogeneous sensor nodes, the design of routing protocols inside the heterogeneous networks is becoming difficult.

Natarajan Meghanathan et al. (Eds) : NETCOM, NCS, WiMoNe, GRAPH-HOC, SPM, CSEIT - 2016
pp. 103– 115, 2016. © CS & IT-CSCP 2016 DOI : 10.5121/csit.2016.61510

Graph theory is a common approach used to design protocols for networking. Traditionally only one metric is considered when a routing protocol is designed. However, in a heterogeneous network, multiple metrics need to be considered. Therefore, it is essential to investigate new types of graph theories to design algorithms and protocols for modern complex heterogeneous networks.

Polychromatic Sets (PS) theory has the ability to consider multiple properties of a set member and describe the relations among the properties, which enables it an appropriate tool to design efficient protocols for networking. Previous research has demonstrated that it could be applied to design protocols for large-scale sensor networks [1]. In this paper, the research work has been extended to heterogeneous networks to investigate the possibility of applying polychromatic sets theory to design protocols for more complex networks.

The performance requirement of a network is always different under various application situations. For example, in the situation of using a wireless sensor network to monitor the fire in a forest, the network should keep its nodes to work as long as possible. Therefore, energy saving is a key task to extend the lifetime of the network. In this case, during normal monitoring process, most devices could be turned into the sleep mode and only a few devices work regularly in order to decrease the energy consumption. In case of detecting the fire, the network is required to operate with high throughput and reliably. Therefore, the routing protocols of such a wireless sensor network need to be adapted to the application, which means the routing protocol should be dynamic.

A heterogeneous sensor network consists of different sensor nodes, such as temperature sensor, humidity sensor, smog sensor and camera sensor. These devices may operate differently in terms of frequency, bandwidth and protocol. Frequency, bandwidth, wireless protocol, energy level of the nodes, and location are inherent characteristics of a sensor node. Throughput, delay, package delivery ratio are performance parameters of the network. All these parameters are defined as the properties of the sensor nodes. Some of them are used as metrics to design the routing protocols in this paper.

In this paper, eight network property parameters, namely location, frequency, residual energy, energy consumption, power mode, bandwidth, throughput, and delay, are considered to design the routing protocols in heterogeneous wireless sensor networks based on PS theory. A dynamic route discovery scheme is proposed for designing the routing protocols. Simulation results demonstrate that performance requirements could be reached under different applications.

The remainder of this paper is organized as follows. Section 2 introduces the related research work. Definitions of metrics used in this paper are described in Section 3. The methodology to define the route discovery based on PS theory is presented in Section 4. PS theory and design of new routing protocols based on PS theory are presented in Section 5. Simulation results are presented in Section 6 with results analysis. Finally this paper is concluded in Section 7.

2. RELATED WORK

Traditional graph theory is not suitable for designing protocols for complex modern networks. New types of graph theories, such as random geometric graphs theory [2], directed graphs network model [3] and distributed graph theory [4], were developed to design the protocols.

However, even these new developed graph theories are not fully functional in terms of designing protocols by considering both the properties of the network nodes and links. Limitations of these graph theories are still obvious.

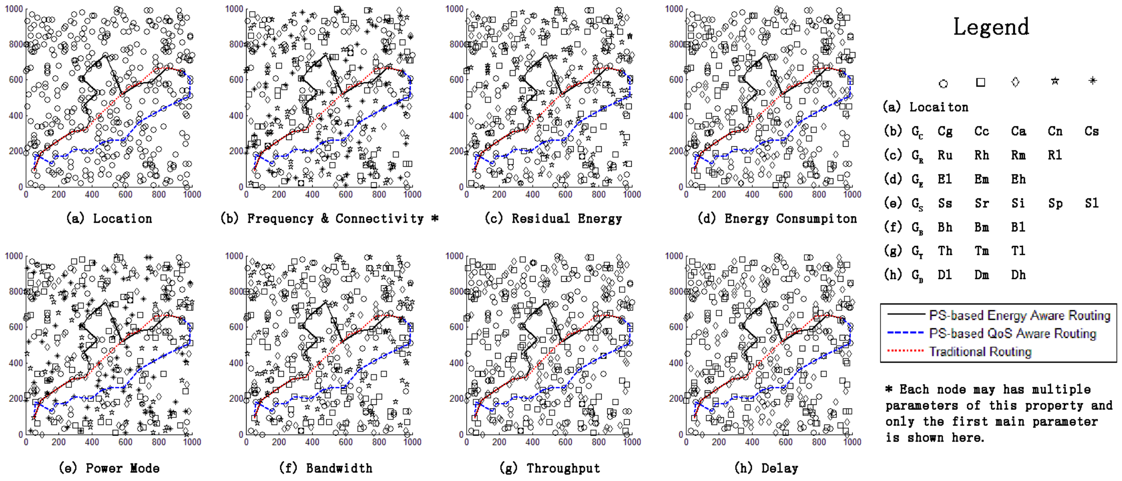
Polychromatic graph based on PS theory is different from previously developed graphs. PS theory has the ability to describe multiple properties of a set member, which makes it an ideal tool to consider the properties of both the network nodes and links. Previous work in [1] demonstrated the feasibility.

Currently, there is little research conducted on multiple metrics aware routing for wireless networks. Normally, for wireless sensor networks, energy consumption [5] or QoS based routing [6] are two major concerns. Multi-constrained QoS routings [6] is a hot research topic, where quite a lot of research has been done. For example, Ben-Othman et al. proposed an energy and QoS aware multi-path routing for wireless sensor networks [7]; Yao et al. provided an energy-efficient, delay-aware and lifetime-balanced data collection [8]; Maurya et al. presented an energy efficient routing for heterogeneous wireless sensor networks by considering distance, energy and load parameters of a network [9]. However, even for multi-constraints QoS routing, the parameters used for designing routing protocols are predefined and, therefore, it is difficult to adapt to the dynamic nature of modern heterogeneous wireless networks.

The aim of the research work presented in this paper is to address this challenge and design a dynamic routing protocols towards application requirements.

3. DEFINITION OF NETWORK PARAMETERS

As shown in figure 1, eight main parameters of a heterogeneous wireless sensor network are considered for designing routing protocols, including location, frequency, node residual energy, energy consumption level, power mode, bandwidth, throughput, and delay.



(a) Location (b) Frequency & Connectivity * (c) Residual Energy (d) Energy Consumpiton

(e) Power Mode (f) Bandwidth (g) Throughput (h) Delay

* Each node may has multiple parameters of this property and only the first main parameter is shown here.

Figure 1. Route Discovery based on Different Metrics

Location

The location of a node is a popular metric in wireless sensor networks for designing geographical routing protocols. Based on GPS or other localization techniques [10], it is easy for the nodes to acquire the information of their location. A node's location is defined as $L(x_L, y_L, z_L)$ in the three-dimensional plane or $L(x_L, y_L)$ in the two-dimensional plane.

Frequency

In a heterogeneous network, network nodes may be equipped with facilities operating at different frequencies. The nodes can only connect with other nodes operating on the same frequency band and protocol. Some of these nodes are able to operate at multiple frequencies. In this scenario, five popular protocols are considered and defined as: C_g, *5GHz (based on IEEE 802.11ac protocol)*; C_c, *2.4GHz (based on IEEE 802.11n protocol)*; C_a, *2.4GHz (based on IEEE 802.11g protocol)*; C_n,*2.4GHz (based on IEEE 802.11b protocol)*; and C_s, *2.4GHz (based on IEEE 802.15.4 protocol)*.

Residual Energy

Since most of the sensor nodes are powered by battery, energy consumption is one of the biggest concerns for wireless sensor networks. In order to extend the lifetime of the networks, sensor nodes with high residual energy are more often used during data transmission. The low residual energy nodes are protected and less used to save the energy. In order to meet the requirement of applying set theory in designing protocols, four levels of nodes' residual energy are defined as: R_u, *Unlimited Residual Energy*; R_h, *High Residual Energy*; R_m, *Medium Residual Energy*; and R_l, *Low Residual Energy*.

Energy Consumption

The energy consumption of different types of nodes in a heterogeneous network is always different. The energy consumption of nodes is defined with three levels: E_l, *Low Energy Consumption*; E_m, *Medium Energy Consumption*; and E_h, *High Energy Consumption*.

Power Mode

In wireless sensor networks, network nodes have four power modes: sending, receiving, idle, and sleep. In some special applications, another mode, silence, is also considered. It could save the transfer time by considering and choosing the relay nodes working on high priority modes (like idle) than those nodes working on low priority modes (like sending). The five parameters of power mode are defined as: S_s, *Sending*; S_r, *Receiving*; S_i, *Idle*; S_p, *Sleep*; and S_l, *Silence*.

Bandwidth

Bandwidth is an important metric to describe the capability of communications in wireless networks. The bandwidth of links connecting various devices is normally different. In heterogeneous networks, the largest residual bandwidth between various devices is defined with four parameters: B_e, *Extra-high Bandwidth*; B_h, *High Bandwidth*; B_m, *Medium Bandwidth*; and B_l, *Low Bandwidth*.

Throughput and Delay

The throughput and delay of networks are two important metrics to describe the QoS of wireless networks. In this scenario, the available throughput is defined with three parameters: T_h, High Throughput; T_m, Medium Throughput; and T_l, Low Throughput. The delay of networks is also defined with three parameters: D_l, *Low Delay*; D_m, *Medium Delay*; and D_h, *High Delay*.

In summary, eight main metrics of heterogeneous wireless sensor networks have been collected and introduced. In next section, the multiple metrics aware heterogeneous network will be defined with this eight metrics by PS theory.

4. NETWORK DEFINITIONS BASED ON PS

In this section, the definition of the networks based on PS theory is described.

4.1. Nodes Set

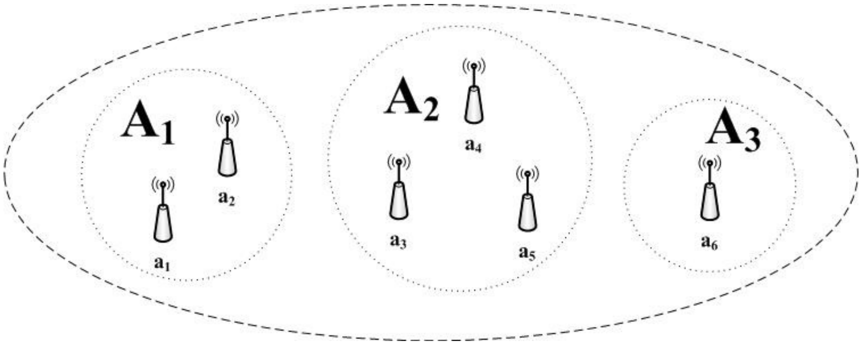


Figure 2. Illustration of a Network and Grouping

As shown in figure 2, a network consists of six nodes. The nodes set A is defined as:

$$A = \{a_1, a_2, \Lambda, a_i, \Lambda, a_n\} \tag{1}$$

where n is the number of nodes. These nodes can be divided into several groups based on their locations or other features. For example, in the network shown in figure 2, node a_1 and a_2 belong to set A_1, node a_3, a_4 and a_5 belong to set A_2 ,and node a_6 belongs to set A_3.

4.2. Colour Set

In PS theory, each property is defined as one individual colour. A colour set F is defined to include all the individual colours.

$$F = \{f_1, f_2, \Lambda, f_j, \Lambda, f_{n_F}\} \tag{2}$$

where f_j is an individual colour and n_F is the number of colours/properties defined in *F*.

In the case of this paper, all the colours, excluding locations, are presented as:

$$F = \{C_g, C_c, C_a, C_n, C_s, R_u, R_h, R_m, R_l, E_l, E_m, E_h, S_s, S_r, S_i, S_p, S_l, B_h, B_m, B_l, T_h, T_m, T_l, D_l, D_m, D_h\} \tag{3}$$

These colours could be also classified into several groups, as:

$$F\begin{cases} G_C = \{C_g, C_c, C_a, C_n, C_s\} \\ G_R = \{R_u, R_h, R_m, R_l\} \\ G_E = \{E_l, E_m, E_h\} \\ G_S = \{S_s, S_r, S_i, S_p, S_l\} \\ G_B = \{B_h, B_m, B_l\} \\ G_T = \{T_h, T_m, T_l\} \\ G_D = \{D_l, D_m, D_h\} \end{cases} \tag{4}$$

In addition, location is utilized for the geographical routing, which will be discussed in next section.

Based on the definition of individual colours, the relations of nodes and their properties/colours can be expressed as a matrix:

$$[A \times F] = \begin{array}{c} \begin{array}{ccccc} C_g & K & f_j & K & D_h \end{array} \\ \left[\begin{array}{ccccc} c_{11} & K & c_{1j} & K & c_{1,26} \\ K & K & K & K & K \\ c_{i1} & K & c_{ij} & K & c_{i,26} \\ K & K & K & K & K \\ c_{n1} & K & c_{nj} & K & c_{n,26} \end{array}\right] \end{array} \begin{array}{c} \\ a_1 \\ K \\ a_i \\ K \\ a_n \end{array} \tag{5}$$

$$c_{ij} = \begin{cases} 1, & f_j \in F_j \\ 0 \end{cases}$$

$n_F = 26$ based on the definitions of all the properties.

4.3. Unity Colour Set

In a network, some nodes share some common properties. These common properties are defined as a unity colour set F_f, which is a subset of colour set F as:

$$\begin{cases} F_f = \{f_{f1}, f_{f2}, K, f_{fj}, K, f_{fo}\} \\ F_f \subset F \end{cases} \tag{6}$$

The selected nodes are also defined as a set A_f, which is a subset of nodes set A. The relations between sets F_f and A_f is presented as a matrix, similar to sets *A* and *F*:

$$[A_f \times F_f] = \begin{array}{c} \begin{array}{ccccc} f_{f1} & \mathrm{K} & f_{fj} & \mathrm{K} & f_{fo} \end{array} \\ \left[\begin{array}{ccccc} c_{11} & \mathrm{K} & c_{1j} & \mathrm{K} & c_{1o} \\ \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} \\ c_{i1} & \mathrm{K} & c_{ij} & \mathrm{K} & c_{io} \\ \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} \\ c_{m1} & \mathrm{K} & c_{mj} & \mathrm{K} & c_{mo} \end{array}\right] \end{array} \begin{array}{c} \\ a_1 \\ \mathrm{K} \\ a_i \\ \mathrm{K} \\ a_m \end{array} \tag{7}$$

$$c_{ij} = \begin{cases} 1, & f_{fj} \in F_j \\ 0 \end{cases}$$

4.4. Available Paths Set

If two nodes share the same properties, a link/path could be established between these two nodes. The total paths available inside a network could be defined as :

$$[A \times A(F)] = \begin{array}{c} \begin{array}{ccccc} F(A_{k1}) & \Lambda & F(A_{kj}) & \Lambda & F(A_{kl}) \end{array} \\ \left[\begin{array}{ccccc} c_{11} & \Lambda & c_{1j} & \Lambda & c_{1l} \\ \Lambda & \Lambda & \Lambda & \Lambda & \Lambda \\ c_{i1} & \Lambda & c_{ij} & \Lambda & c_{il} \\ \Lambda & \Lambda & \Lambda & \Lambda & \Lambda \\ c_{m1} & \Lambda & c_{mj} & \Lambda & c_{ml} \end{array}\right] \end{array} \begin{array}{c} \\ a_1 \\ \Lambda \\ a_i \\ \Lambda \\ a_m \end{array} \tag{8}$$

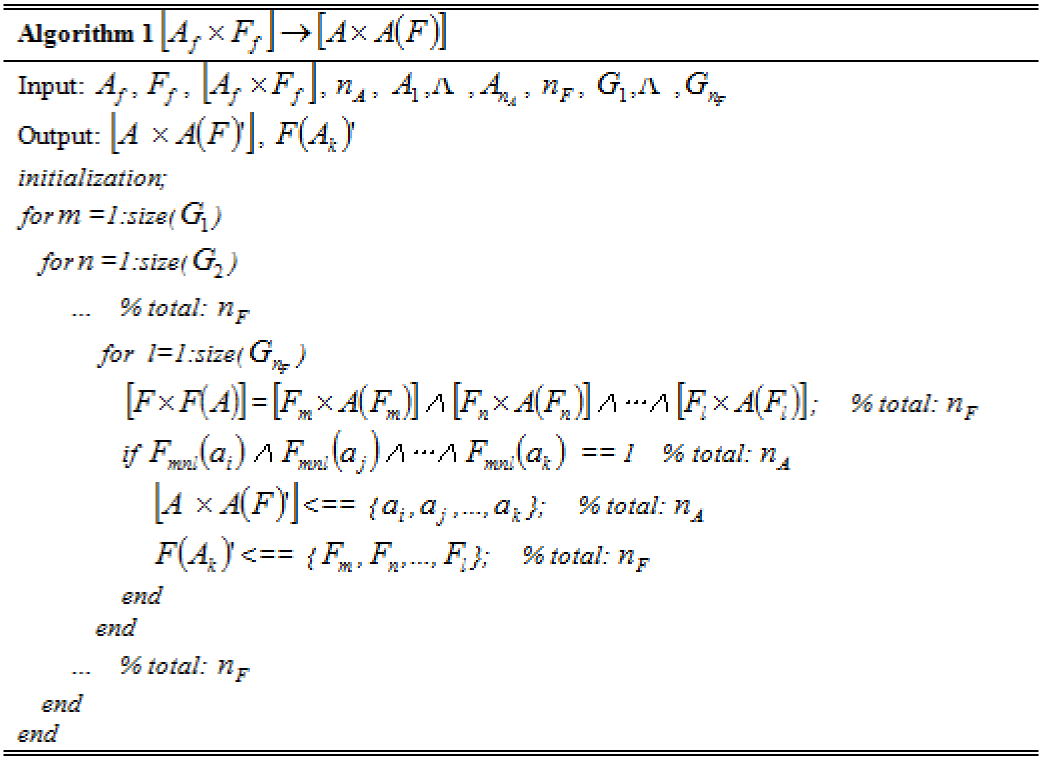
$$c_{ij} = \begin{cases} 1, & a_i \in A_{kj} \\ 0 \end{cases}$$

where $F(A_{kj})$ is a set of colours that a group of nodes A_{kj} in the corresponding column have all of those colours, defined as:

$$\begin{cases} F(A_{kj}) = \{f_{k1}, f_{k2}, \Lambda, f_{ki}, \Lambda, f_{ko}\} \\ F(A_{kj}) \subset F_f \\ \forall f_{ki} \subset F(A_{kj}): \\ [A_{kj} \times f_{ki}] = [1 \quad \Lambda \quad 1 \quad \Lambda \quad 1]^T \end{cases} \tag{9}$$

where $\lfloor A_{kj} \times f_{ki} \rfloor$ is an all-ones matrix.

A route could be established by using the available paths from (9). Algorithm 1 can be used to find the available paths $[A \times A(F)]$. In algorithm 1, $[F \times F(A)]$ is the conjunction of different nodes' properties, $[F_m \times A(F_m)]$ is a rank of node's properties, F_m and $F_{mnl}(a_i)$ are the tags whether a_i has all the properties F_{mnl} or not. Based on this algorithm, 7 parameters are required for input: a set of nodes A_f, a set of colours F_f, a set of nodes' properties $\lfloor A_f \times F_f \rfloor$, number of node groups n_A, node groups $A_1, \Lambda\ , A_{n_A}$, number of property groups n_F and metric groups G_1, $\Lambda\ , G_{n_F}$. The set of selected nodes $\lfloor A\ \times A(F)' \rfloor$ and the parallel set of their matching metrics $F(A_k)'$ are output.



```
Algorithm 1 $\lfloor A_f \times F_f \rfloor \rightarrow \lfloor A \times A(F) \rfloor$
Input: $A_f$, $F_f$, $\lfloor A_f \times F_f \rfloor$, $n_A$, $A_1, \Lambda\ , A_{n_A}$, $n_F$, $G_1, \Lambda\ , G_{n_F}$
Output: $\lfloor A\ \times A(F)' \rfloor$, $F(A_k)'$
initialization;
for m =1:size( $G_1$ )
  for n =1:size( $G_2$ )
    ...  % total: $n_F$
      for  l=1:size( $G_{n_F}$ )
        $[F \times F(A)] = [F_m \times A(F_m)] \wedge [F_n \times A(F_n)] \wedge \cdots \wedge [F_l \times A(F_l)]$;   % total: $n_F$
        if $F_{mnl}(a_i) \wedge F_{mnl}(a_j) \wedge \cdots \wedge F_{mnl}(a_k)$  == 1   % total: $n_A$
          $\lfloor A\ \times A(F)' \rfloor$<== { $a_i, a_j, ..., a_k$ };   % total: $n_A$
          $F(A_k)'$ <== { $F_m, F_n, ..., F_l$ };   % total: $n_F$
        end
      end
    ...  % total: $n_F$
  end
end
```

5. GEOGRAPHICAL ROUTING BASED ON PS

Location based routing [11] is a popular routing scheme for multi-hop networks including heterogeneous wireless sensor networks. The data packets are transmitted to the destination node(s) based on their location rather than their ID or logical address. In this paper, location based routing is also applied to demonstrate the performance of routing protocols based on PS theory.

5.1. Greedy Forwarding Routing

As shown in figure 3, a general location based routing, greedy forwarding routing (GFR) [12], is selected for comparison. In traditional GFR, every node broadcasts its location information to its neighbour nodes so that the current node knows the location information of other nodes in its signal coverage area. The location of the destination node is also known.

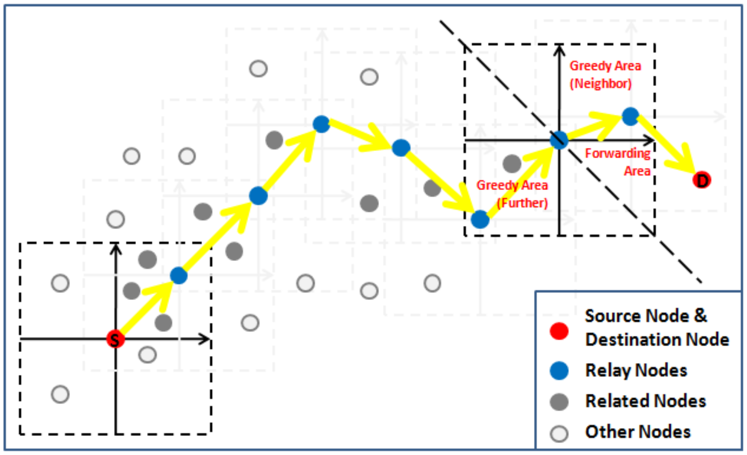


Figure 3. Greedy Forwarding Routing

The area around a node is divided into four zones based on the distance to the destination. As shown in figure 3, the nearest to the destination node is the forwarding area and the other two near areas are neighbour greedy areas and further greedy area, respectively. Whenever a route is to be established, the node in forwarding area which is the nearest to the destination should be selected as the next hop node. If there is no such a suitable node in the forwarding area, the suitable node in neighbour greedy areas should be selected. The node in further greedy area will be selected as a last resolution if necessary.

5.2. Next Hop Selection Based on PS

Route discovery based on PS theory is different from traditional. As shown in figure 4, when the destination node is not in the signal coverage area of current node, the current node is classified as a node group A_1 by itself and the nodes in the considered area are classified as node group A_2. After collecting the properties of these nodes in A_2, the nodes having the expected properties are included in an available paths set. The node which is the nearest to the destination node in this set is selected as the next hop node. In figure 4, node a_3 with high priority property is selected as the next hop node even though node a_2 with low priority in the same area is much nearer to the destination node than node a_3.

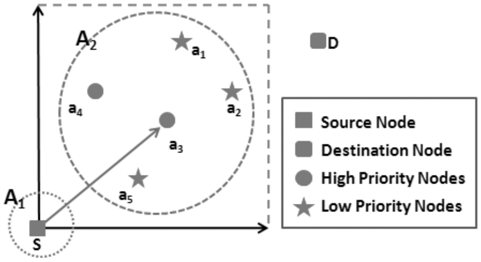


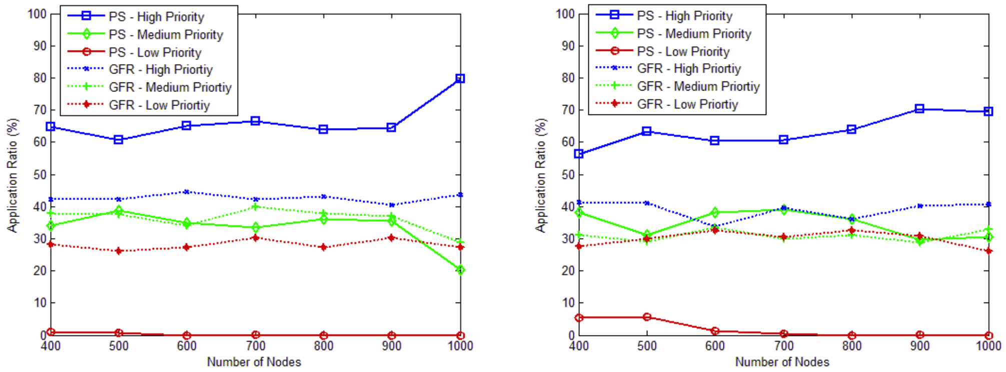
Figure 4. PS-based Metrics Aware Next Hop Node Selection

6. SIMULATIONS AND ANALYSIS

A simulation platform based on PS theory is built and developed for simulating the heterogeneous wireless sensor networks in Matlab. 400~1000 nodes are deployed in a 1000m×1000m area randomly. These nodes have eight different properties as defined in section 3. In a simplified simulation, we assume that all the nodes can connect with the neighbour nodes around it based on traditional GFR. Energy efficient routing is designed by considering three metrics (Residual Energy, Energy Consumption and Power Mode) and QoS aware routing is designed by considering Bandwidth, Throughput and Delay. The discovery radius of each node is set as $100\sqrt{2}$ m. The optimized routes are selected based on energy efficient routing, QoS aware routing and traditional GFR routing, which are shown in figure 1. The performance of the routing protocols is evaluated by the number of participation nodes with different priorities, throughput and delay.

6.1. Application Ratio of Nodes with Different Metrics

In a wireless sensor network, nodes are classified into different priority groups based on requirements. For example, nodes with high residual energy and low energy consumption level are classified as high priority nodes and nodes with high throughput are also classified as high priority nodes. Three levels, namely high priority, medium priority and low priority, are used to classify the nodes. The number of nodes at each priority group is similar. Figure 5 shows the results of number of nodes selected for routing with different priorities based on PS routing and traditional GFR routing. In GFR, the ratios of applying nodes with different priorities are similar, between 35% and 45%. In PS-based routing, the application of nodes with high priority is more than 60%. At the same time, few of low priority nodes are utilized and none of low priority node is applied when the number of nodes is more than 700. The PS-based routing improves the application of high priority nodes and decreases the application of low priority nodes.



(a) Energy Aware Routing (b) QoS Aware Routing

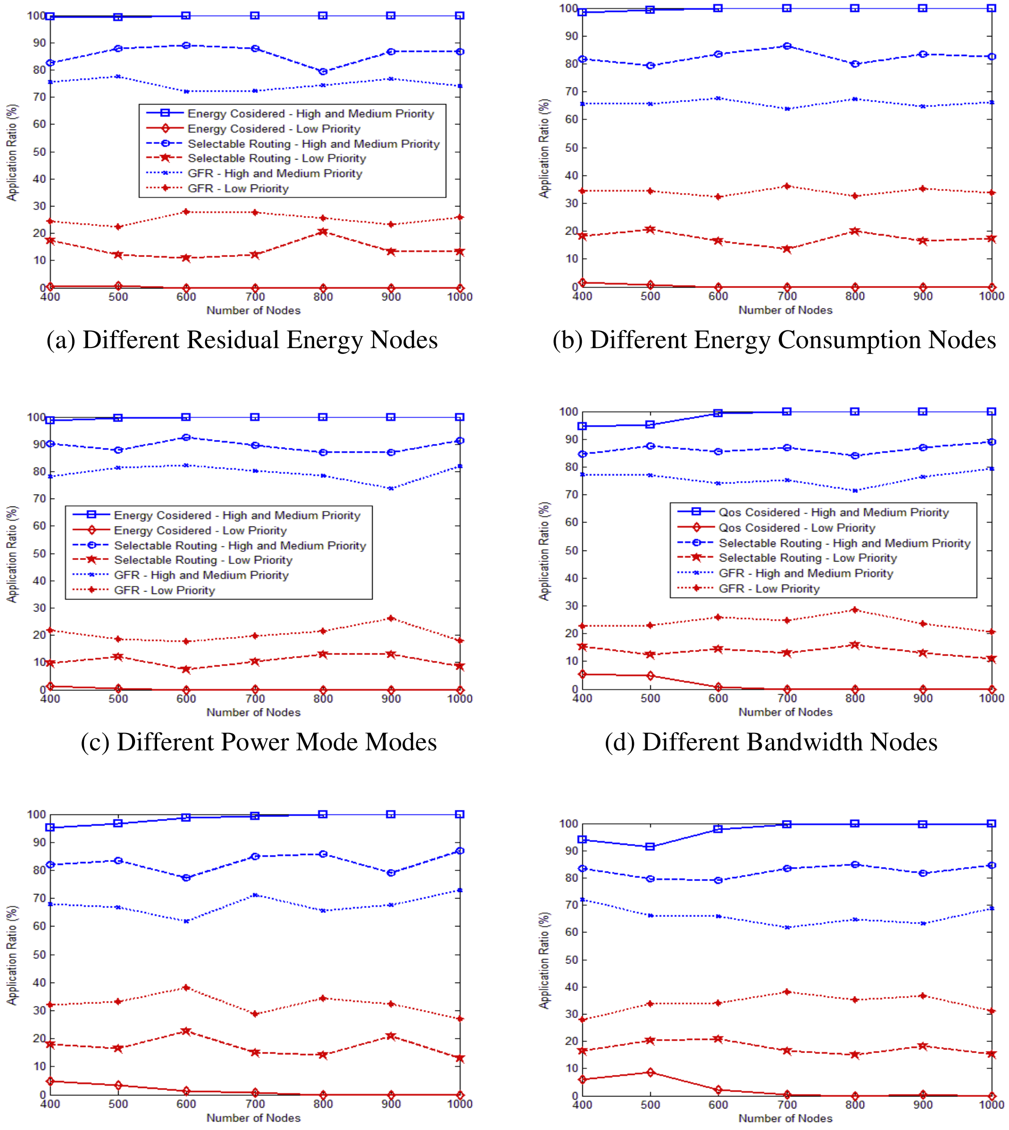
Figure 5. Application Ratio of Nodes with Different Priority

Figure 6 shows the ratios of numbers of nodes with different priority under other six evaluation metrics. For all the scenarios, PS based routing significantly increased the ratios of applying high priority nodes and decreased the ratios of applying low priority nodes. This is one advantage of PS based routing protocols.

6.2. Throughput and Delay

Figure 7 shows the results of throughputs and figure 8 shows the results of delay for protocols based on GFR and PS. In figure 7, the bandwidth of some nodes in the simulated network is 11Mbps, such as IEEE 802.11b protocol, the average throughput of GFR network is limited by these nodes. However, PS-based routing is possible to avoid using these nodes with low bandwidth and then improves the overall network throughput.

The same reason applies to delay. Since PS has the ability to select nodes with short delay, the overall network delay is shorter than GFR routing. The results are shown in figure 8.



(a) Different Residual Energy Nodes

(b) Different Energy Consumption Nodes

(c) Different Power Mode Modes

(d) Different Bandwidth Nodes

(e) Different Throughput Nodes

(f) Different Delay Nodes

Figure 6. Application of Nodes with Different Metrics on Different Modes

7. CONCLUSION

In this paper, a PS-based route discovery and routing protocol design are proposed to improve the performance of the network. Multiple metrics of heterogeneous networks are considered in this network model. The performance of heterogeneous networks was improved prominently. This demonstrates the feasibility of applying PS theory in designing protocols for heterogeneous networks.

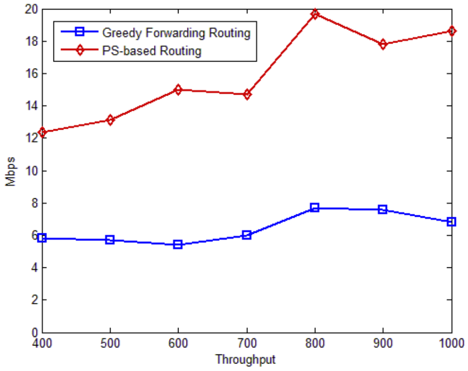


Figure 7. Throughput

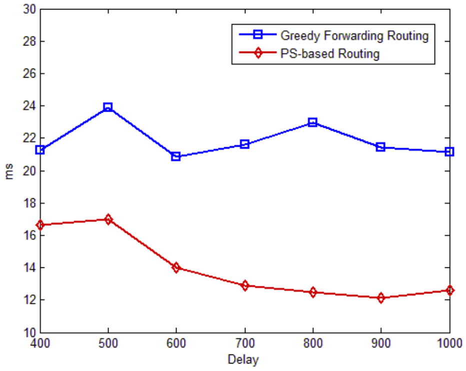


Figure 8. Delay

REFERENCES

[1] Xinheng Wang & Shancang Li, (2013) “Scalable routing modeling for wireless ad hoc networks by using polychromatic sets”, IEEE Systems Journal, Vol. 7, No. 1, pp50-58.

[2] M. Haenggi, J. G. Andrews, F. Baccelli, O. Dousse& M. Franceschetti, (2009) “Stochastic geometry and random graphs for the analysis and design of wireless networks”, IEEE Journal on Selected Areas in Communications, Vol. 27, No. 7, pp1029-1046.

[3] Yanhua Li & Zhi-Li Zhang, (2010) “Random walks on digraphs: A theoretical framework for estimating transmission costs in wireless routing”, IEEE INFOCOM, pp1-9.

[4] Qing Chen, Qian Zhang & Zhisheng Niu, (2009) “A graph theory based opportunistic link scheduling for wireless ad hoc networks”, IEEE Transactions on Wireless Communications, Vol. 8, No. 10, pp5575-5585.

[5] Xiao Chen, Zanxun Dai & Hongchi Shi, (2013) “EgyHet: An energy-saving routing protocol for wireless heterogeneous sensor networks”, IEEE ICNC, pp778-782.

[6] Xiaoxia Huang & Yuguang Fang, (2008) “Multiconstrained QoS multipath routing in wireless sensor networks”, Wireless Networks, Vol. 14, No. 4, pp465-478.

[7] Jalel Ben-Othman & Bashir Yahya, (2010) “Energy efficient and QoS based routing protocol for wireless sensor networks”, Journal of Parallel and Distributed Computing, Vol. 70, No. 8, pp849-857.

[8] Yanjun Yao, Qing Cao& A. V. Vasilakos, (2015) “EDAL: An energy-efficient, delay-aware, and lifetime-balancing data collection protocol for heterogeneous wireless sensor networks”, IEEE/ACM Transactions on Networking, Vol. 23, No. 3, pp810-823.

[9] Sonam Maurya & A. K. Daniel, (2014) "An Energy Efficient Routing Protocol under Distance, Energy and Load Parameter for Heterogeneous Wireless Sensor Networks", International Conference on Information Technology (ICIT), pp161-166.

[10] S. Li, X. Wang, S. Zhao, J. Wang & L. Li, (2014)"Local semidefinite programming-based node localization system for wireless sensor network applications", IEEE Systems Journal, Vol. 8, No. 3, pp879-888.

[11] F. Cadger, K. Curran, J. Santos & S. Moffett, (2013) "A survey of geographical routing in wireless ad-hoc networks", IEEE Communications Surveys & Tutorials, Vol. 15, No. 2, pp621-653.

[12] J. Na, D. Soroker & C. Kim, (2007) "Greedy geographic routing using dynamic potential field for wireless ad hoc networks", IEEE Communications Letters, Vol. 11, No. 3, pp243-245.

INTENTIONAL BLANK

AN EFFICIENT DEPLOYMENT APPROACH FOR IMPROVED COVERAGE IN WIRELESS SENSOR NETWORKS BASED ON FLOWER POLLINATION ALGORITHM

Faten Hajjej, Ridha Ejbali and Mourad Zaied

Research Group on Intelligent Machines (REGIM-Lab) Sfax, Tunisia
hajjej.faten.tn@ieee.org
ridha_ejbali@ieee.org
mourad.zaied@ieee.org

ABSTRACT

Wireless Sensor Networks (WSNs) are experiencing a revival of interest and a continuous advancement in various scientific and industrial fields. WSNs offer favorable low cost and readily deployable solutions to perform the monitoring, target tracking, and recognition of physical events. The foremost step required for these types of ad-hoc networks is to deploy all the sensor nodes in their positions carefully to form an efficient network. Such network should satisfy the quality of service (QoS) requirements in order to achieve high performance levels. In this paper we address the coverage requirement and its relation with WSN nodes placement problems. In fact, we present a new optimization approach based on the Flower Pollination Algorithm (FPA) to find the best placement topologies in terms of coverage maximization. We have compared the performance of the resulting algorithm, called FPACO, with the original practical swarm optimization (PSO) and the genetic algorithm (GA). In all the test instances, FPACO performs better than all other algorithms.

KEYWORDS

WSN, Sensors Deployment problem, Coverage, FPA.

1. INTRODUCTION

Technological innovations in miniaturization, power management and wireless communication in the recent years have enabled the progress of wireless networks, which have attracted a growing interest for many applications and fields, such as military sensing, physical air traffic control, video surveillance, traffic surveillance, industrial and manufacturing automation, security, environment monitoring, and building and structural monitoring.

A wireless Sensor Network (WSN), which is a targeted wireless network, consists of a significant number of miniaturized electronic devices, called sensors, distributed over a specified area in order to sense the environment and communicate the accumulated information from the

Natarajan Meghanathan et al. (Eds) : NETCOM, NCS, WiMoNe, GRAPH-HOC, SPM, CSEIT - 2016
pp. 117– 129, 2016. © CS & IT-CSCP 2016 DOI : 10.5121/csit.2016.61511

monitored field to other networks (e.g., the internet). In WSN, sensors have limited resources, typically the energy resources, and the calculation capabilities, as well as the storage capacity. Therefore, most studies and researches on WSNs have focused on the optimization of resources in order to enhance the performances and meet the quality of service (QoS) requirements. Determining the sensor field topologies is a key challenge in sensor resource management. Consequently, WSN performance is powerfully influenced by the deployment topology of sensor nodes, which affect QoS metrics, such as energy consumption, sensor lifetime, and sensing coverage equally [1].

In the literature, the deployment topology can be classified according to; the placement methodology that can be either random placement or grid-based placement (deterministic placement), the optimization of performance metrics such as connectivity, sensing coverage, energy consumption and lifetime, and the roles the deployed node, which can be regular, relay, cluster-head, or base-station, plays [2]. However, the placement techniques can be further categorized into static and dynamic whether the optimization is performed at the time of deployment or whiles the network is working, respectively. The choice of the deployment scheme depends on many properties [2]. Therefore, many studies considered that for some cases random placement becomes the only option due to the environment characteristics [3] [4] and deployment cost, and time. Figure 1 shows the different categories of node placement strategies.

Our major focus in this paper is on how to choose the optimal nodes deployment that can achieve maximal coverage of the monitored area [5]. Thus, optimal nodes placement issue is a problem that has been proven NP-hard for most formulations of sensor deployment [6].

The coverage metric is a decisive metric that can be considered as a measure of permanence and QoS for WSN. Coverage in a WSN is to ensure that the Region of Interest (RoI) is monitored with high reliability in order to have the necessary information on the supervised phenomenon [7]. Coverage issues can be commonly classified into two types: target coverage problem and area coverage problem. The former ensures the monitoring of only certain specific points which have fixed positions in the area of interest, while the latter is concerned with the supervision of the whole deployment area. Target coverage can be categorized as Q-coverage or simple coverage. For simple coverage, each target should be monitored by at least one sensor node. For Q-coverage, each target has to be monitored by at least Q different working sensor nodes.

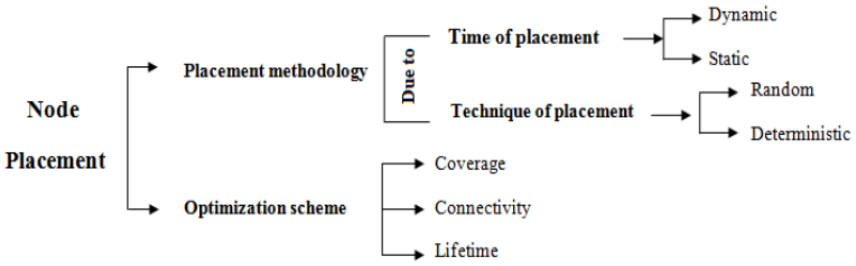


Figure 1. Sensor node placement methodologies

The connectivity metric is as important as coverage in wireless sensor networks. A WSN is defined as connected if, and only if, there exists at least one route between each pair of nodes. Thus, connectivity depends on the existence of paths and can therefore be directly affected by changes of topology. For this reason an optimal deployment strategy have to maximize coverage with respect to the connectivity constraint.

Nature constantly inspires research in the field of optimization. While genetics, ants and particle swarm algorithms are famous examples, other nature inspired optimization algorithms emerge regularly. Flower Pollination Algorithm (FPA) is novel global optimization algorithm inspired from pollination process of flowers. FPA is simple and very powerful; in fact, it can outperform both genetic algorithm (GA) and particle swarm optimization (PSO) according to [8].

To find the best nodes deployment that would achieve maximal coverage of the targeted area without affecting network connectivity, a new approach based on FPA is introduced to enhance coverage in a wireless sensor network. We considered a centralized topology and an area coverage problem with random sensor deployment. Here, different scenario was tested. The proposed approach was able to maximize the total coverage area for the considered scenarios. The remainder of this paper is organized as follows. Section 2 gives a literature survey about different deployment algorithms. The problem formulation is presented in section 3. Section 4 specifies the proposed FPA based deployment approach. In section 5, the simulation results and discussion are given. Finally, section 6 concludes the paper.

2. LITERATURE SURVEY

Over the last years, researchers attempted to tackle the nodes deployment problem in WSN through various optimization processes both by mathematical programming and through nature inspired techniques. This problem was sometimes modelled as a one single objective problem, in special cases deal with several objectives through well selected weights.

Yu et al. proposed a node placement algorithm for mobile sensor networks based on the strength of van der Waals in order to improve the total coverage area. In fact, the proximity relationship of nodes is defined by the Delaunay triangulation method, the frictional force is inserted into the equation of force, the force calculated generate an acceleration in the movement of nodes. To evaluate whether the nodes are uniformly distributed over the deployment field an evaluation metric named pair correlation function was introduced in [9]. The Genetic Algorithm (GA) was introduced as a solution for coverage holes problem in WSN [10]. This approach found the optimal positions and the number of mobile nodes that have to be added to the initial deployment schema. Simulation results prove that this algorithm has optimized network coverage in terms of overall coverage ratio and additional number of mobile nodes. Sengupta et al. addressed the problem of achieving an optimal trade-off between coverage, energy consumption, and lifetime in WSN by using the multi-objective evolutionary algorithm (MOEA). They developed an enhanced version of Multi-objective evolutionary algorithm based on differential evolution (MOEA/D-DE) known as MOEA/DFD which includes the fuzzy dominance [11]. Sakamoto et al. proposed a simulation approach founded on Particle Swarm Optimization (PSO). They focused on the size of giant component and number of covered mesh clients (NCMC), which are important objective functions to optimize Wireless Mesh Networks (WMNs) [12]. In their work, the authors of [13] proposed a modified version of the original artificial bee colony (ABC); in fact, they change the updating equation of onlooker bee and scout bee [14]. Indeed, some new parameters, such as forgetting and neighbors factor for accelerating the convergence speed and probability of mutant for maximizing the coverage rate were introduced [15]. Comparing their approach with the deployment topology based on the traditional ABC and PSO algorithm, they found that the former achieved better performance in terms of coverage and speed of convergence with less moving distance sensor.

3. Problem Formulation

The deployment of sensor nodes in WSN is to find the placement nodes topology or find the coordinates of the sensor nodes in the two-dimensional plane. The most important concerns for WSN are how improving the performances and optimizing the resources. Thus, an optimal placement strategy ought to be considered to achieve the required goal. Here our objective is to find an optimal placement schema that maximizes the coverage area without losing network connectivity. For this, the following different mathematical models are described.

3.1. Preliminary

Sensor nodes in WSN are characterized by their positions in the 2D plane (x, y), sensing radius R_s, and communication radius R_c. Given a multi-hop WSN, where all nodes collaborate in order to ensure cooperative communication. Such network, can be defined as a linked graph, G = {V, E}, where V is the set of vertices representing sensors and E is the set of edges representing links between the sensors. Let u ϵ V and v ϵ V, (u, v) belongs to E if, and only if, u can send a message directly to v (we say that v is neighbor of u). We assume that R_c is identical for all nodes. Let d(u, v) be the distance between the nodes u and v, the set E can be defined as follows:

$$E = \left\{ (u,v) \in V^2;\ \mathrm{d(u,v)} \le \mathrm{R}_C \right\}$$

The network coverage is defined by the sensing radius of the sensor nodes, whereas the network connectivity is specified by the communication radius of the nodes.

3.2. Connectivity

Definition 1 (Node Degree). Given an undirected graph G. The degree Deg(u), of a vertex u ϵ V is specified as the number of a vertex u ϵ V is specified as the number of neighbors of u [16].

Definition 2 (k-Node Connectivity). A graph is considered to be connected if for every pair of nodes, there exists a single hop or a multi-hop path connecting them; otherwise the graph is called disconnected. A graph is considered to be Q-connected if for any pair of nodes there are at least Q reciprocally separate paths connecting them [16].

3.3. Binary Sensing Model

The coverage in WSN defined as the total area covered by a set of sensor nodes deployed in the region of interest (ROI). This region is considered as m × n grids, each grid point size was equal to 1 and denoted as G(x, y) (Figure. 2).

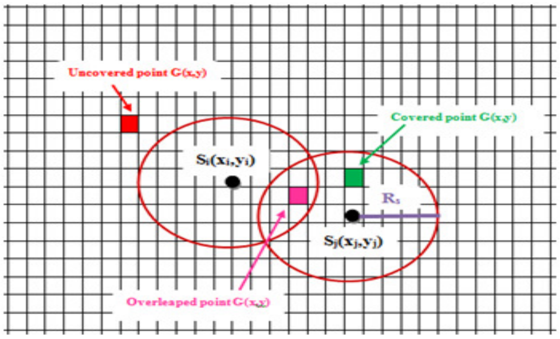


Figure 2. Sensor coverage in sensing field

Generally, the zone covered by a sensor node is a disk with radius equals to sensing radius of the sensor. The binary sensing model considered that each grid point within the sensing radius of a node can be considered as covered with probability equal to "1" and the point out of the sensing range was set as "0" since it cannot be covered (Eq1). Thus, the coverage of the whole area is proportional to the grid points that can be covered by at least one sensor $S_i(x_i, y_i)$ [17].

$$P = \begin{cases} 1, \; if & \sqrt{(x - x_i)^2 - (y - y_i)^2} \leq R_s \\ 0, & otherwise \end{cases}$$

4. The Proposed Approach

This work interested by the node sensors deployment problem in WSN. In fact, we deal with area coverage problem for random placement topology with predefined number of sensors. Here, the main purpose was to improve the quality of coverage without affecting network connectivity constraint. Evidently, to supply connected coverage to a zone, the set of disks used much cover all points in that region and the connectivity graph of all the R_c-disks must form a single connected component in a graph theoretic sense. The proposed approach, named Flower Pollination Coverage Optimization approach (FPCOA), was a centralized approach based on FPA, aimed to deploy all the sensor nodes in their positions carefully to form a WSN with maximal coverage area.

4.1. Fitness function

The binary model was considered as sensing model (Section 2.3). The proposed approach is a mono-objective deployment approach designed to optimize one objective function, namely the ratio of total coverage target area. It is given by:

$$P(x, y, S) = 1 - \prod_{i=1}^{N} \left(1 - P(x, y, S_i\right)$$

With N is the number of sensor nodes and $P(x, y, S_i)$ is the probability that a grid point G(x, y) is covered by a sensor S_i. So, the total coverage area is defined as:

$$TotCovArea = \sum_{x=1}^{m} \sum_{y=1}^{n} P(x, y, S)$$

And the ratio of total coverage area is given by:

$$Total\ Coverage\ ratio = \frac{TotCovArea}{TotalGridArea}$$

4.2. Constraints

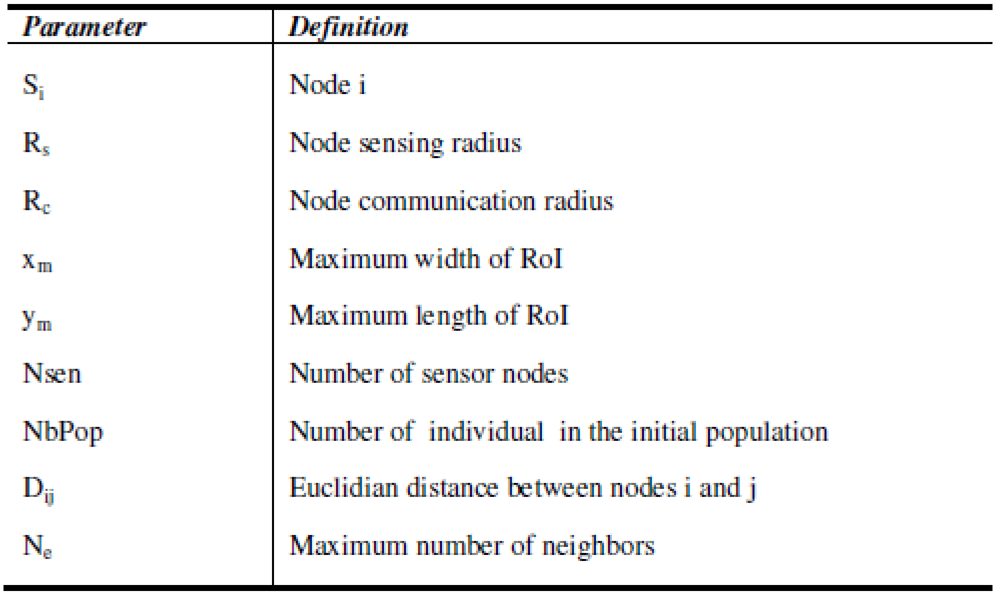
The network connectivity is taken as a constraint in this optimization problem. Therefore one path, at least, must exist from the sensor node to the sink node, to guarantee connectivity

4.3. Flower Pollination Coverage Optimization Algorithm (FPCOA)

The proposed approach composed of two main steps. The first step was the creation of the initial population (Algorithm 1). The second step was the performing of the optimization process based on FPA (Algorithm 2).

Initial Population. To create the initial population we considered that each individual was represented by a vector of all sensor nodes position (x, y) in RoI. The WSN parameters are described in Table 1.

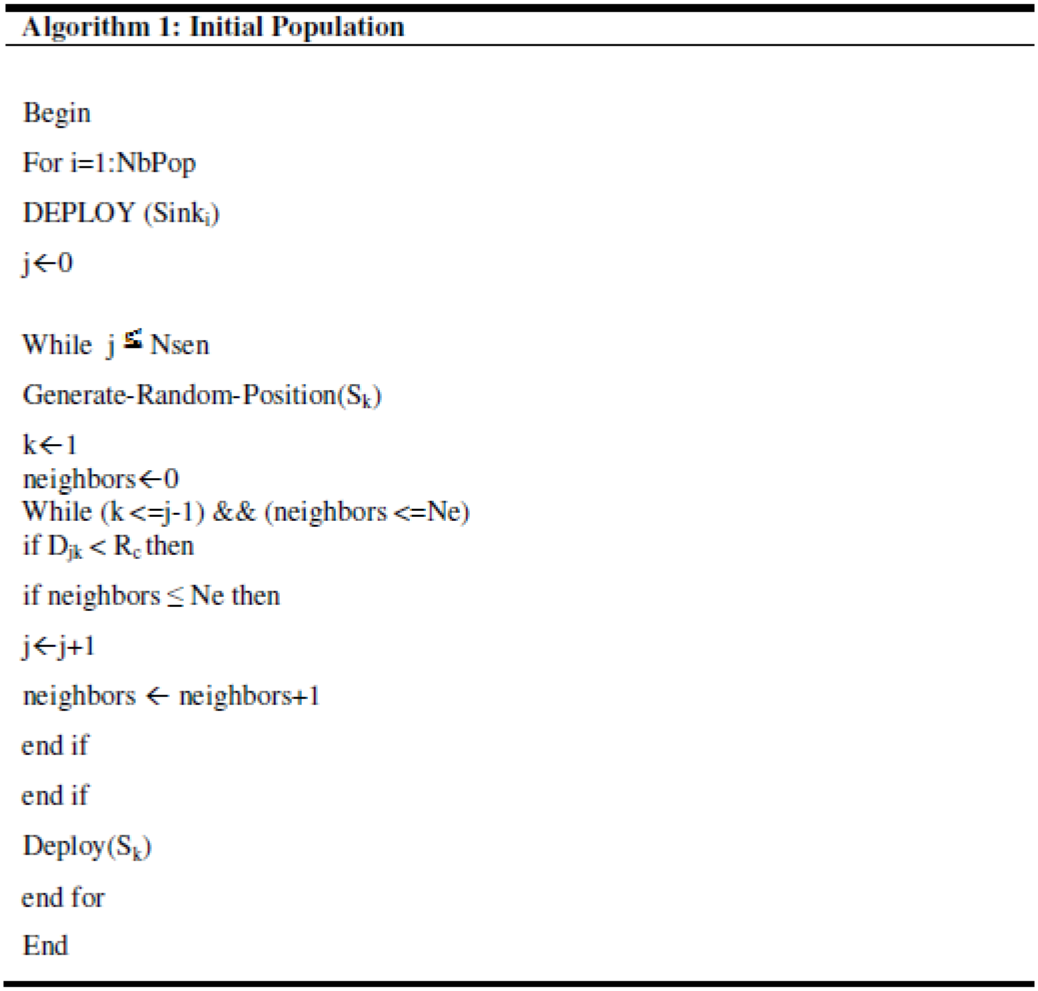
Table 1. Parameters of WSN.



Parameter	*Definition*
S_i	Node i
R_s	Node sensing radius
R_c	Node communication radius
x_m	Maximum width of RoI
y_m	Maximum length of RoI
Nsen	Number of sensor nodes
NbPop	Number of individual in the initial population
D_{ij}	Euclidian distance between nodes i and j
N_e	Maximum number of neighbors

To create initial population, we began by generating the position of the sink node at the centre of RoI (i.e., at $x_m/2$ and $y_m/2$) for each individual. Then, we deployed the remaining sensors by taking into consideration the connectivity constraint. Actually, network connectivity is assumed to be full if the distance between two sensors is less than the communication radius (R_c) of the sensor. The distance is defined as the Euclidean distance between two sensors. In addition, to insure a sufficient distribution in RoI, we controlled the number of neighbors of each deployed node that should be less than a predefined number N_e (see Algorithm1).

Table 2. Pseudo code of Initial Population Creation.



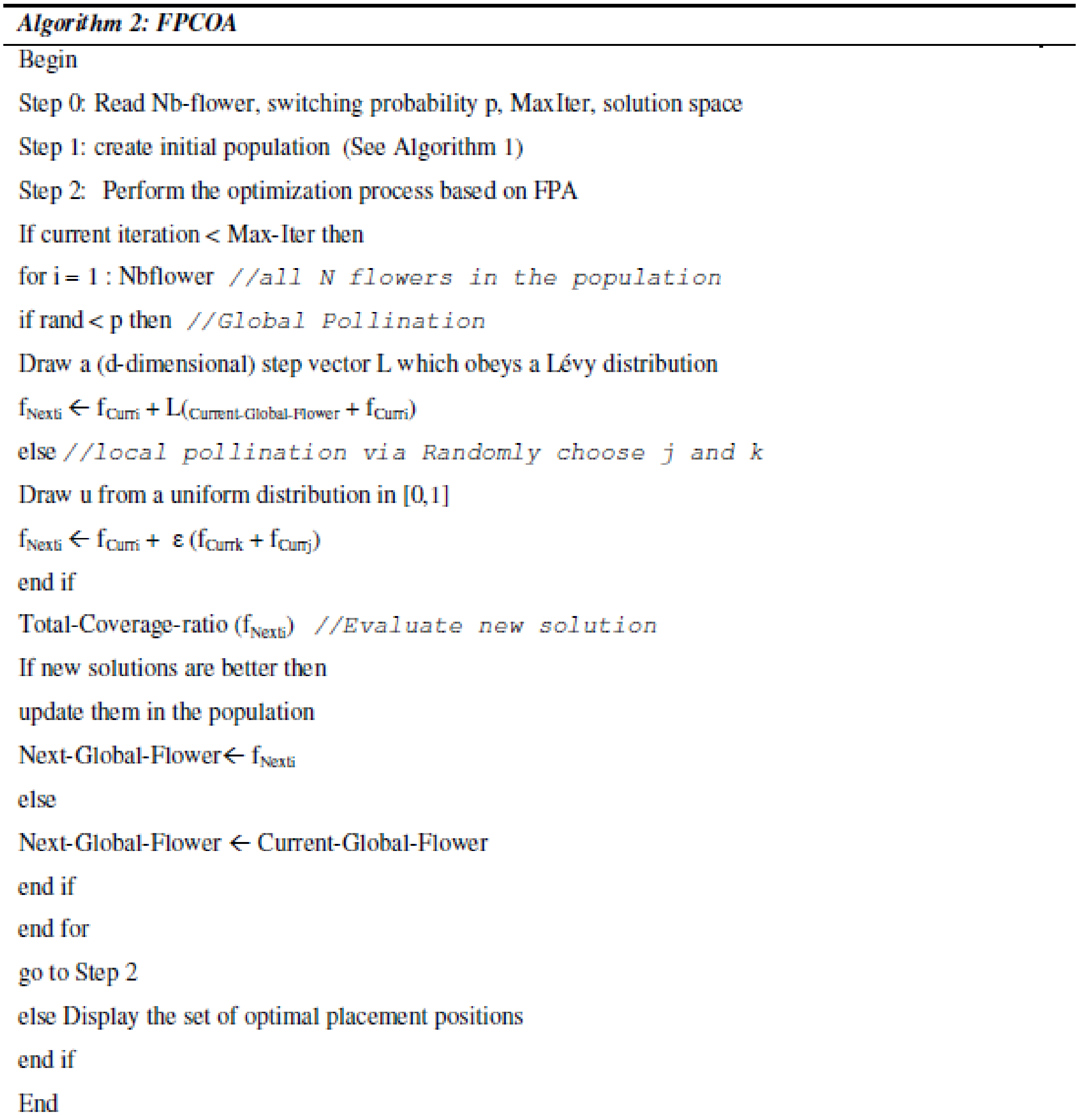
Algorithm 1: Initial Population

```
Begin
For i=1:NbPop
DEPLOY (Sink$_i$)
j←0

While j $\leq$ Nsen
Generate-Random-Position(S$_k$)
k←1
neighbors←0
While (k <=j-1) && (neighbors <=Ne)
if D$_{jk}$ < R$_c$ then
if neighbors $\leq$ Ne then
j←j+1
neighbors ← neighbors+1
end if
end if
Deploy(S$_k$)
end for
End
```

Flower Pollination Algorithm (FPA). metaheuristics are generic algorithms, often inspired from nature, designed to solve challenging optimization problems [17] [18]. Here, we considered one of the most recent metaheuristic algorithms named Flower Pollination Algorithm (FPA), developed by Xin-She Yang in the year 2012 [8] for the global optimization problems. FPA inspired from the flower pollination process of flowering plants. In nature, flowers pollination process resulting from the transfer of pollen, typically, by pollinators such as insects, birds, bats and other animals. In fact, pollination process can be commonly classified into two types: self-pollination and cross-pollination. The former can occur by the pollen of the same flower. The latter can take place by pollen of a flower of a different plant [20] [21]. FPA has the following four rules:

1. *Cross-pollination is considered as global pollination process with pollen carrying; pollinators performing Lévy flights.*
2. *Self-pollination is considered as local pollination.*
3. *Flower constancy can be defined as the reproduction probability is proportional to the similarity of the two flowers involved.*
4. *Global and local pollination is controlled by a switch probability p ϵ [0, 1].*

Table 3. Pseudo Code of Flower Pollination Coverage Optimization.



Algorithm 2: FPCOA

```
Begin
Step 0: Read Nb-flower, switching probability p, MaxIter, solution space
Step 1: create initial population  (See Algorithm 1)
Step 2:  Perform the optimization process based on FPA
If current iteration < Max-Iter then
    for i = 1 : Nbflower //all N flowers in the population
        if rand < p then //Global Pollination
            Draw a (d-dimensional) step vector L which obeys a Lévy distribution
            $f_{Nexti} \leftarrow f_{Curri} + L(_{Current-Global-Flower} + f_{Curri})$
        else //local pollination via Randomly choose j and k
            Draw u from a uniform distribution in [0,1]
            $f_{Nexti} \leftarrow f_{Curri} + \varepsilon (f_{Currk} + f_{Currj})$
        end if
        Total-Coverage-ratio ($f_{Nexti}$) //Evaluate new solution
        If new solutions are better then
            update them in the population
            Next-Global-Flower $\leftarrow f_{Nexti}$
        else
            Next-Global-Flower $\leftarrow$ Current-Global-Flower
        end if
    end for
    go to Step 2
else Display the set of optimal placement positions
end if
End
```

Here each flower was represented by a vector of all sensor nodes position (x, y) in RoI, f_{Curr1}, f_{Curr2}, …,f_{CurrN} was the flower population at iteration t, f_{Next1}, f_{Next2}, …,f_{NextN} was the flower population at iteration t +1, Nbflower was the total number of flower and the Current-Global - Flower is the best solution found among all solutions at the current generation or iteration t. To imitate the movement of pollinator [22], FPA uses Lévy flight. That is, we draw L > 0 from a Lévy distribution:

$$L \sim \frac{\lambda\Gamma(\lambda)\sin(\frac{\pi\lambda}{2})}{\pi}\frac{1}{s^{1+\lambda}} \quad (s\ ?\ \ s_0\ ?\ \ 0)$$

The pseudo-code of FPA is presented in Table3.

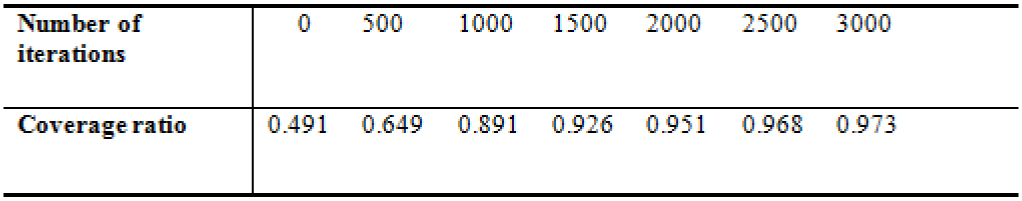
5. SIMULATION AND RESULTS

To validate the proposed approach, some simulations were undertaken. We used a binary sensing model the nodes are initially randomly distributed. The network is homogeneous, i.e., all sensors have the same deployment parameters such as the sensing and communication radius. Simulations were carried out using MATLAB R2016a. The algorithm was run a maximum number of iterations of 3000. The average of 10 runs was recorded. For the simulations, we considered a square area divided into a number of squares of 1 m2 each. The center of each of these squares is taken as the demand point to detect by at least one sensor node. In this section, the performance of the proposed FPCOA is evaluated with regard to the total coverage ratio. Moreover, the obtained results were compared with those obtained with two metaheuristics algorithms, namely, PSO and GA and, finally, the effect of the number of randomly deployed sensor nodes was discussed.

5.1. Efficiency of the proposed approach

In order to test the performances of FPCOA, we considered a square area with each side 100m in length. We considered also that the number of sensors and the communication radius R_c as well as the sensing radius R_s as constant values. Here the number of sensors was set as 15, R_c as 15m and R_s as 15m.

Table 4. Deployment results of FPCOA.



Number of iterations	0	500	1000	1500	2000	2500	3000
Coverage ratio	0.491	0.649	0.891	0.926	0.951	0.968	0.973

As seen in Table 4, the effective coverage area was improved significantly over the 3000 iterations. The decrease in the standard deviation values can be explained by the stability of the algorithm with larger numbers of iterations. In fact, FCPOA improved the coverage ratio by 48.2% compared with the random initial distribution. To highlight this improvement, the best deployments obtained by the FCPOA for initial and final configurations are shown in Figure 2 and Figure 3, respectively, where the colored areas represent detected coverage areas.

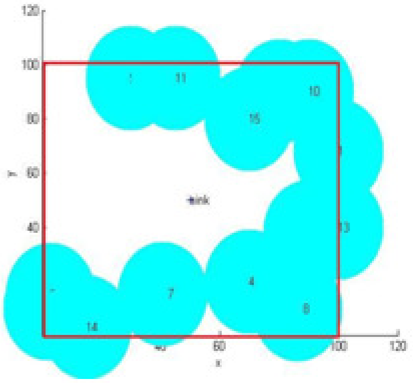


Figure 3. Initial configuration of Sensors

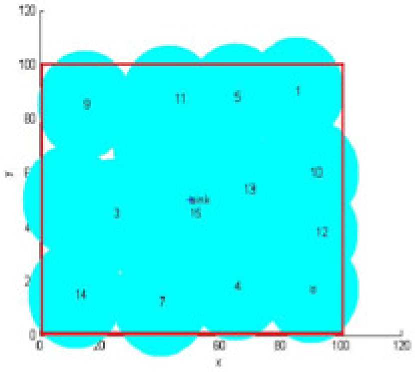


Figure 4. Final configuration of Sensors

5.2. Comparison with other approaches

To evaluate the efficiency of our proposed approach we choose to compare our results with those obtained with GA and PSO, respectively. Figure 5 gives the comparison of the coverage rate tested on the same initial population for the three approaches.

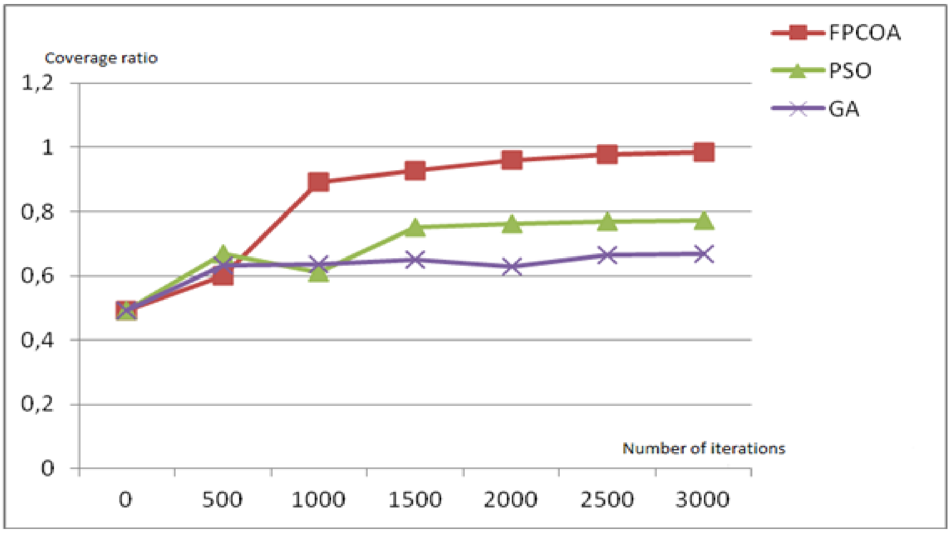


Figure 3. Comparison of total coverage ratio with GA and PSO

From this figure, we can find that after the nodes reached stable distribution and obtained the optimal placement topology, the proposed algorithm has better coverage rate than the other two approaches. The results of the proposed approach clearly outperform both than GA and PSO respectively. This figure shows that FPCOA gives a much more stable performance in total coverage than both the two algorithms.

5.3. Effect of Number of Sensor Nodes

In order to show the effect of number of sensor on the total coverage ratio for the propose approach, we considered that the sensor nodes were randomly deployed in a 50m×50m sensor field, the communication radius R_c was set as 5m and the sensing radius R_s was set as 5m.

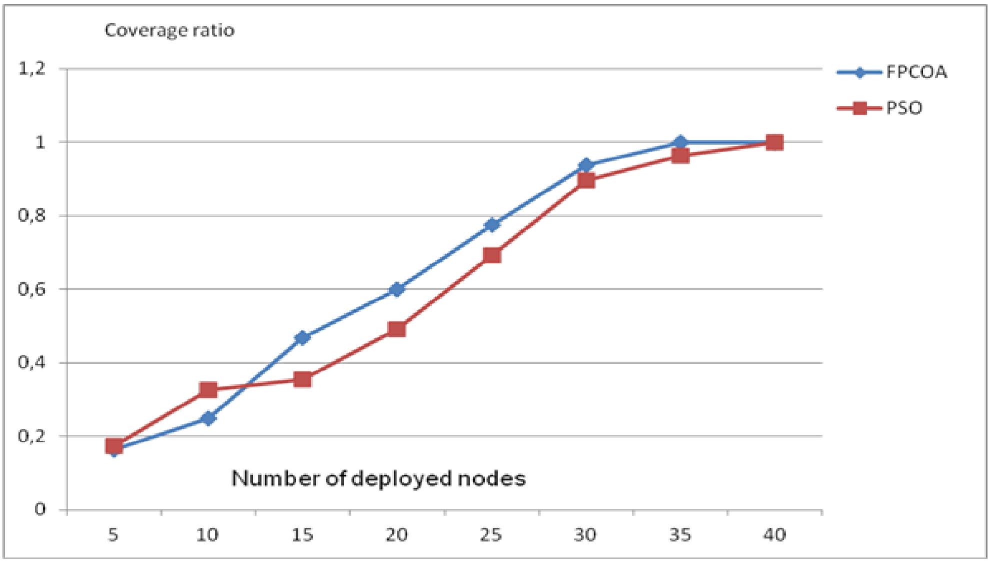


Figure 4. The Coverage Ratio vs. Number of Sensor Nodes

Figure 6 shows the coverage ratio when adding sensor nodes to the network for both of FPCOA and PSO. As shown, the coverage ratio increases as the number of deployed nodes increases. This figure indicates that the proposed approach offers higher coverage with less sensor nodes. FPCOA requires around 32 sensor nodes to get 100% coverage compared to PSO which requires 34 sensor nodes. Thus, it can be said that FPCOA is able to offer higher coverage with the lowest cost.

6. CONCLUSION

In this paper, the sensor placement problem for WSN is addressed. A deployment approach based on FPA was proposed. This approach can find the optimal placement topology in terms one QoS metric. The simulations results of the different scenarios prove that our proposed approach achieved the optimal placement regarding coverage maximization and connectivity constraint. In a future work, we will incorporate other QoS metrics like energy consumption and deal with multi-objective node placement problem for the WSNs.

REFERENCES

[1] F. Oldewurtel, P. Mhnen, (2010) “Analysis of enhanced deployment models for sensor networks,” in Vehicular Technology Conference, pp. 1-5.

[2] M. Younisa, K. Akkayab. (2008) “Strategies and techniques for node placement in wireless sensor networks: A survey”, Ad Hoc Networks. Vol. 6, No. 4, pp. 621-655.

[3] F. Y. S. Lin, P. L. Chiu, (2005) “A near-optimal sensor placement algorithm to achieve complete coverage-discrimination in sensor networks”, IEEE Communications Letters, Vol. 9, No. 1, pp. 43-45.

[4] M. Zaied, C. Ben Amar, M. A Alimi “Award a new wavelet based beta function” International conference on signal, system and design, SSD03 1, 2003, pp. 185-191 .

[5] R Ejbali, Y Benayed, M Zaied, A. M Alimi “Wavelet networks for phonemes recognition” International conference on systems and information processing , 2009.

[6] L. W. X. Cheng, D-Z Du, B. Xu, (2008), “Relay sensor placement in wireless sensor networks”, Journal of Wireless Networks, Vol. 14, No. 3. pp. 347-355.

[7] R. G. J. Wang, S. Das, (2010) “A survey on sensor localization”. Journal of Control Theory and Applications, Vol. 8, No.1, pp.2-11.

[8] X. S. Yang, “Flower pollination algorithm for global optimization” in Unconventional Computation and Natural Computation, vol. 7445, 2012, pp.240-249.

[9] X. Yu, N. Liu, W. Huang, X. Qian, and T. Zhang, (2013) “A node deployment algorithm based on van der waals force in wireless sensor networks”, International Journal of Distributed Sensor Networks, pp. 1-8.

[10] O. Banimelhem, M. Mowafi, W. Aljoby, (2013) “Genetic algorithm based deployment in hybrid wireless sensor networks”, Communications and Network, Vol. 5 No. 4, pp. 273-279.

[11] S. Senguptaa, S. Dasb, M.D. Nasira, B.K. Panigrahic, (2013) “Multi-objective node deployment in wsns : In search of an optimal trade of among coverage, lifetime, energy consumption, and connectivity”, Engineering Applications of Articial Intelligence, Vol. 26, No. 1, pp. 405-416.

[12] S. Sakamoto, T. Oda, M. Ikeda, L. Barolli , (2015) “Design and implementation of a simulation system based on particle swarm optimization for node placement problem in wireless mesh networks,” Intelligent Networking and Collaborative Systems (INCOS), pp. 164-166.

[13] X. Yu, J. Zhang, J. Fan, and T. Zhang, (2013) A faster convergence artificial bee colony algorithm in sensor deployment for wireless sensor networks, International Journal of Distributed Sensor Networks, Vol. 9, No. 10 pp. 2-9.

[14] B. Guedri, M. Zaied, C. Ben Amar “Indexing and images retrieval by content” High Performance Computing and Simulation (HPCS), pp. 2011.

[15] A. El Adel, M. Zaied, C. Ben Amar “Learning wavelet networks based on Multiresolution analysis: Application to images copy detection Communications”, Computing and Control Applications (CCCA), 2011.

[16] D. B. West, (2006) “Introduction to graph theory”, 2nd ed., Prentice-Hall.

[17] A. Hossain, P. K. Biswas, S. Chakrabarti, (2008) “Sensing Models and Its Impact on Network Coverage in Wireless Sensor Network,” pp1-2.

[18] M. Zaied, R. Mohamed, C. Ben Amar, (2012) "A power tool for Content-based image retrieval using multiresolution wavelet network modeling and Dynamic histograms", International REview on Computers and Software (IRECOS), vol. 7 No. 4.

[19] R. Ejbali, M. Zaied, C. Ben Amar, (2012) "Multi-input Multi-output Beta Wavelet Network: Modeling of Acoustic Units for Speech Recognition", International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 3, No. 4.

[20] A. Elfes, "Occupancy grids: a stochastic spatial representation for active robot perception," in Autonomous Mobile Robots: Perception, Mapping and Navigation, vol. 1, S. S. Iyenger and A. Elfes, Editors, IEEE Computer Society Press, pp. 60-70, 1991.

[21] O. Jemai, R. Ejbali, M. Zaied, C. Ben Amar "A speech recognition system based on hybrid wavelet network including a fuzzy decision support system" Seventh International Conference on Machine Vision (ICMV), pp. 503-944, 2015.

[22] R. Ejbali, M. Zaied, C. Ben Amar, "Intelligent approach to train wavelet networks for Recognition System of Arabic Words", International Conference on Knowledge Discovery and Information Retrieval , 2010.

AUTHORS

Faten Hajjej received the graduate degree in Computer Engineering from the National School of Engineers of Sfax, University of Sfax, in 2009. She has been pursuing the Ph.D degree with the Research Group on Intelligent Machines (REGIM-Lab), University of Sfax, under the supervision of Dr. Ridha EJBALI and Prof. Mourad Zaied. His research interests include the internet of things, wireless sensor network, multi objective optimization, nature inspired optimization algorithms.

Ridha Ejbali received the Ph.D degree in Computer Engineering, Master degree and computer engineer degree from the National Engineering School of Sfax Tunisia (ENIS) respectively in 2012, 2006 and 2004. He was assistant technologist at the Higher Institute of Technological Studies, Kebili Tunisia since 2005. He joined the faculty of sciences of Gabes Tunisia (FSG) where he is an assistant in the Department computer sciences since 2012. His research area is now in pattern recognition and machine learning using Wavelets and Wavelet networks theories.

Mourad Zaied received the HDR, the Ph.D degrees in Computer Engineering and the Master of science from the National Engineering School of Sfax respectively in 2013, 2008 and in 2003. He obtained the degree of Computer Engineer from the National Engineering School of Monastir in 1995. Since 1997 he served in several institutes and faculties in university of Gabes as teaching assistant. He joined in 2007 the National Engineering School of Gabes (ENIG) as where he is currently an associate professor in the Department of Electrical Engineering. He is a member of the REsearch Group on Intelligent Machines laboratory (REGIM) http://www.regim.org in the National Engineering School of Sfax (ENIS) since 2001. His research interests include Computer Vision and Image and video analysis. These research activities are centered on Wavelets and Wavelet networks and their applications to data classification and approximation, pattern recognition and image, audio and video coding and indexing.

INTENTIONAL BLANK

SHARP OR BLUR: A FAST NO-REFERENCE QUALITY METRIC FOR REALISTIC PHOTOS

Fan Zhang

Lenovo Research, Shenzhen, China
zhangfanhk@lenovo.com

ABSTRACT

There is an increasing demand on identifying the sharp and the blur photos from a burst of series or a mass of collection. Subjective assessment on image blurriness takes account of not only pixel variation but also the region of interest and the scene type. It makes measuring image sharpness in line with visual perception very challenging. In this paper, we devise a no-reference image sharpness metric, which combines a set of gradient-based features adept in estimating Gaussian blur, out-of-focus blur and motion blur respectively. We propose a dataset-adaptive logistic regression to build the metric upon multiple datasets, where over half of the samples are realistic blurry photos. Cross validation confirms that our metric outperforms the-state-of-the-art methods on the datasets with a total of 1577 images. Moreover, our metric is very fast, suitable for parallelization, and has the potential of running on mobile or embedded devices.

KEYWORDS

Image sharpness, No reference metric, out-of-focus, motion blur, logistic regression

1. INTRODUCTION

With fast-growing consumer electronics camera technology, such as phone camera, wearable camera, vehicle camera and aerial camera, there are challenging times in exploring easily attained photos. Those photos might be captured in a causal way, without a stable support to camera or intent focusing on scene. One demand is to discard the useless blurry photos in a mass of collection for efficient album management. Another challenge is to pick up the clearest photo from a sequence of burst shooting, for either photo snap under unavoidable camera shaking or photo recognition crowd-sourcing service like *CamFind*. To address the automatic and instant photo selection, there is a growing interest in a photo sharpness (or blurriness) metric in line with human visual perception.

Blur analysis has been widely studied but not well solved. Blur identification in computer vision society aims at estimating the type and the amount of blur [19, 16, 28, 26]. It tells in-focus regions from out-of-focus ones or moving regions from still backgrounds, forming a blurriness map to guide segmentation [5], super-resolution [9], shape-from-focus [20], depth-from-defocus [32], motion-from-blur [11], defocus magnification [1], and deblur [11, 25]. However blur identification pays attention to local blurriness in radiometric intensity rather than the overall quality in perception. Focus measure in electronic microscope and camera design [13, 20] can

Natarajan Meghanathan et al. (Eds) : NETCOM, NCS, WiMoNe, GRAPH-HOC, SPM, CSEIT - 2016
pp. 131– 145, 2016. © CS & IT-CSCP 2016 DOI : 10.5121/csit.2016.61512

tackle a focus image series, but probably fail when the scene is moving or changing. Quality metrics SSIM, widely used in signal processing society, match perception fairly well, but they need a reference image, known as the full-reference metrics. Recent studies develop the no-reference sharpness metric [14, 2] with the guidance of subjective image quality database and even unlabeled images. The subjective databases provide the MOS (subjective mean opinion scores) as the ground truth of blur extent, and the unlabeled data are synthesized with the same distortion type and level as the subjective databases [29, 27]. However, most databases use the synthetic images, which are generated from a limited set of source images with spatially-invariant Gaussian convolution. Existing metrics often do well in those synthetic datasets, but perform poorly on the realistic blur, e.g., UFRJ database [10, 7]. It is also interesting that Ye et al. studied the relationship between the blurriness of document images and the OCR (optical character recognition) accuracy, and reported that OCR accuracy may not be consistent with human perception [29]. To summarize, few prior arts study the realistic blur from the perspective of perception; the existing metrics are good at synthetic blur among an image series sharing the similar scenes, but perform poorly in practice and change dramatically when scene changes.

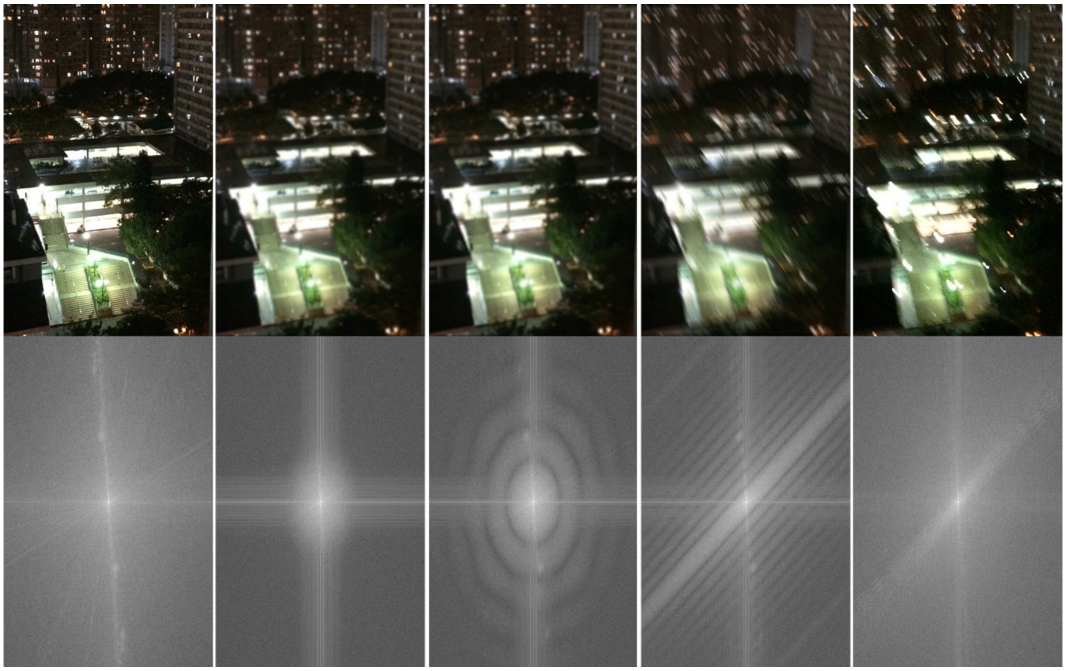

(a) Reference (b) Gaussian blur (c) Disk blur (d) Linear motion blur (e) Realistic blur

Figure 1. Synthetic and realistic blurry images and their log spectrum $\log|\mathcal{F}(I)|$

In fact, realistic blur is challenging to measure because:

1) Realistic blur mainly include out-of-focus blur and motion blur.

Out-of-focus blur generally smooth the edges, but motion blur may generate edges parallel to the motion smear, e.g., light streaks (see Fig. 1(e)). A single feature can hardly predict the hybrid blur.

2) Photos may exhibit diverse scenes, from the smooth, e.g., sky and face, to the rough, e.g., forest and fabrics.

A natural scene may contain steep edges around occlusions or soft edges due to illumination. Simple image statistics often fail to tell a sharp smooth scene from a blur rough scene.

3) Blur pixels do not always degrade photo sharpness.

For example, lens blur can pop out the in-focus objects from the out-of-focus background and make objects distinctively sharper in appearance. Pure average of local sharpness is possibly inconsistent with subjective appreciation.

In this study, we develop a fast metric for practical applications. Keeping this in mind, we learn from realistic data and approximate the solution by using a set of low-level vision features. Contributions of our work include:

1) We design and select a set of features regarding their correlations with various perceptual blur. Those features employ different image statistics and varied pooling strategies to complementarily measure the multiple types of blur in subjective datasets.

2) Such datasets are an exhaustive gathering of current public subjective databases and our own collection of failure cases. Those versatile data may guarantee unbiased data-driven modeling.

3) We formulate a dataset-adaptive logistic regression to co-train our metric on multiple datasets. It bridges the gap between misaligned datasets and makes full use of the adopted data-driven approach.

2. REALISTIC BLUR

In photography, photo blur mainly stems from lens blur and motion blur. Lens blur keeps clear the in-focus objects if any, and yet obscures the out-of-focus things. Motion blur may pervade the whole photo for camera shake or occurs locally on moving objects.

In mathematics, out-of-focus blur is modeled as convolution with a disk kernel, and motion blur as convolution with a trajectory kernel. If the motion happens to be linear uniform during exposure, the kernel evolves to a line. A disk kernel has a spectrum with circularly symmetric sinc waves, while a line kernel exhibits parallel sinc waves in spectrum [12, 16, 28]. If a kernel is **spatially-invariant**, its distinct pattern accumulates and reflects in the image spectrum. Fig. 1 shows the log spectrum of the disk blur (c) and linear motion blur (d) synthesized from the reference (a).

However, the spatially-invariant assumption is too strong for realistic blur. Many factors will violate the assumption.

1) Out-of-focus varies with the object depth, and the motion blur correlates to the velocity of moving objects;

2) Lens geometric distortion differentiates the blur along the radial direction;

3) Nonlinear tone mapping (e.g., Gama correction) in imaging pipeline changes the dependency among pixels which originally takes place at image sensor.

Moreover, after the processing in imaging pipeline including demosaic, denoise and/or resize, the final blur makes toward Gaussian blur to some extent. Fig. 1(e) shows the log spectrum of realistic blurry image, where sinc wave patterns disappear. As a result, realistic blur is often hybrid and hardly described by a simple mathematic model.

3. Perceptual Sharpness Feature

A considerable amount of subjective rating data is available now. It inspires us to use data driven modeling. First of all, we design and select the basic features for each type of blur, including Gaussian blur, out-of-focus blur and motion blur. Then we combine the features with logistic model and train a robust metric on datasets with regression method.

3.1. Maximal gradient (MAG)

Image gradient can indicate the image sharpness, for blur smooths the gradients. For example, Gaussian blurring process is the solution to the diffusion equation of $\frac{\partial I}{\partial t} = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$ in physics. The image intensity can be considered as a flow, that is pushed away by a force equal to the gradient. The steeper the gradient, the faster the pixels diffuse. Considering the directional gradients from the current pixel to its 8-connected neighbors as denoted in Fig. 2, the steepest gradient is defined as $\max_{d=0,1,\ldots,7}|I_d(x,y)|$ by Bahrami [2].

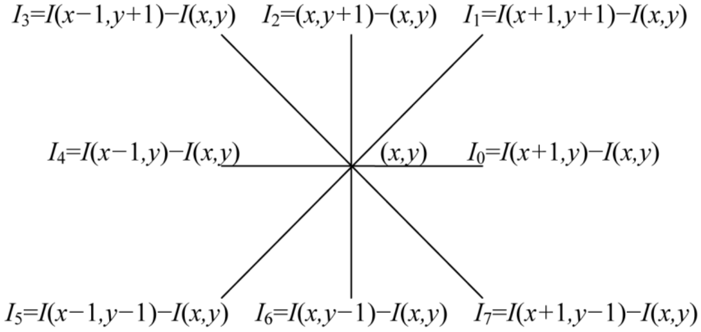

$I_3 = I(x-1,y+1) - I(x,y)$
$I_2 = (x,y+1) - (x,y)$
$I_1 = I(x+1,y+1) - I(x,y)$
$I_4 = I(x-1,y) - I(x,y)$
(x,y)
$I_0 = I(x+1,y) - I(x,y)$
$I_5 = I(x-1,y-1) - I(x,y)$
$I_6 = I(x,y-1) - I(x,y)$
$I_7 = I(x+1,y-1) - I(x,y)$

Figure 2. Notations of directional gradients I_d at pixel (x,y)

Smooth regions always pervade natural images, where steep gradients rather than gentle gradients dominate the perceptual sharpness. So the largest local maximal gradients are taken into account, which often correspond to the structures and textures of an image. As proposed in [2], the global maximal gradient is computed as

$$\mathrm{MAG} = \mathbf{E}\left[\underset{(x,y)}{\max k}\left\{\max_{d=0,1,\ldots,7}|I_d(x,y)|\right\}\right], \qquad (1)$$

where operator $\mathbf{E}[\cdot]$ computes the expectation (i.e., the mean), k is empirically set to 2% and thus the operator max k takes 2% of the largest local maximal gradients. The local maximal gradient maps are shown in the second row of Fig. 3, where 2% of the largest values are marked bright.

3.2. Minimal 2nd order gradients ratio (MGR)

Gaussian blur and out-of-focus blur smooth image isotropically, while motion blur flattens image specially along the motion smears. Such flattening can decrease the 2nd order image gradient in a certain direction. Blurriness in this direction can be estimated by the minimal 2nd order gradient, i.e., $\min_{d=0,1,2,3}|I_d(x,y) + I_{d+4}(x,y)|$.

However, smooth regions also have small 2nd order gradients, and thus may mislead the blurriness measurement above. To overcome the flaw, we, again, select and count the largest values within an image like MAG, as

$$\mathrm{M2G} = \mathbf{E}\left[\underset{(x,y)}{\max k}\left\{\min_{d=0,1,2,3}\left|I_d(x,y) + I_{d+4}(x,y)\right|\right\}\right]$$

The minimal 2nd order gradient maps are shown in the third row of Fig. 3, where the largest values are marked bright.

A heavier motion blur can spread over a wider area, but can be hardly described by the 2nd order gradient which covers only a 3×3 neighbourhoods. To avoid slow convolutions with big kernel size, we resort to multi-scale analysis. To be specific, the ratio of the M2G values between image scales may indicate whether the blurriness propagation terminates at the current stage. Given a sequence of increasing resolutions $\left(I_{2^j}\right)_{j\in Z}$, we have corresponding $\mathrm{M2G}_{2^j}$:

$$\mathbf{E}\left[\underset{(2^j x,\, 2^j y)}{\max k}\left\{\min_{d=0,1,2,3}\left|I_d(2^j x, 2^j y) + I_{d+4}(2^j x, 2^j y)\right|\right\}\right]$$

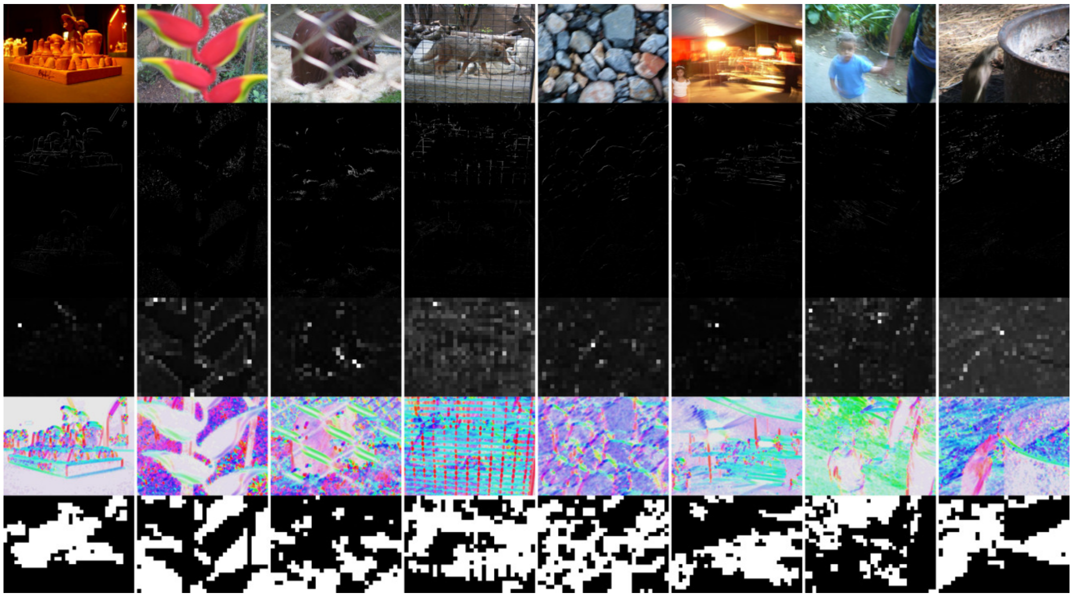


Figure 3. Visualized sharpness feature map of the samples in UFRJ database. The rows, top down, show original images, maximal gradients, minimal 2nd order gradients, gradient kurtosis, doubled gradient angles, and patch-wise MAG (for computing MSG) respectively.

We define the minimal 2nd order gradients ratio (MGR) as:

$$\mathrm{MGR}_1 = \mathrm{M2G}_{2^1}/\mathrm{M2G}_{2^0} \tag{2}$$

$$\mathrm{MGR}_2 = \mathrm{M2G}_{2^2}/\mathrm{M2G}_{2^2} \tag{3}$$

3.3. Average gradient kurtosis (AGK)

Natural images have super-Gaussian distribution with an acute peak and heavy tails in spatial frequency domain. The blur process widens the distribution and thus reduces its peakedness. Actually, the kurtosis for a blurred patch is smaller than that of a sharp one in Fourier [31], DCT (discrete cosine transform) [4], and gradient domain respectively [25]. Shi et al., use local gradient kurtosis to justify the blurriness of patches [25], which inspires us evaluating the sharpness of whole image with the average local gradient kurtosis as

$$\begin{gathered}\mathrm{GK} = \mathbf{E}_{p\in I}\left[\min\{K_p|I_0, K_p|I_2\}\right] \\ \forall p \in I, \forall(x,y) \in p : K_p = \frac{\mathbf{E}\left[\left(I(x,y) - E[I_p]\right)^4\right]}{\mathbf{E}\left[\left(I(x,y) - E[I_p]\right)^2\right]^2}\end{gathered} \tag{4}$$

where image grid I is divided into non-overlapped patches indexed with p, kurtosis operator $K_p[\cdot]$ computes the kurtosis within the p-th patch, and the local gradient kurtosis is the smaller one between the vertical gradient kurtosis and the horizontal gradient kurtosis. The gradient kurtosis maps are shown in the fourth row of Fig. 3.

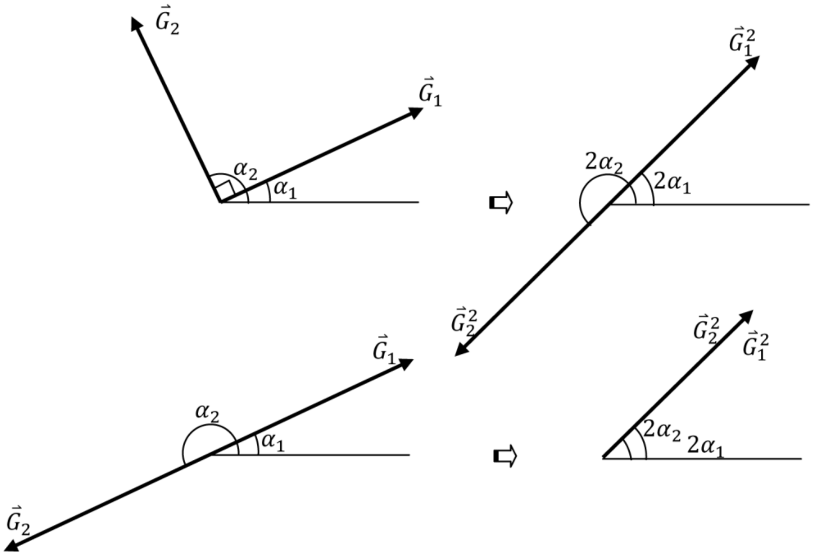


Figure 4. Doubling the vector angle may cancel out the perpendicular vectors (top) and consolidate the opposite vectors (bottom).

3.4. Average angle-doubled gradient

So far, we expect that the maximal gradients can measure the isotropic blur like Gaussian blur and the minimal 2nd order gradients can evaluate the anisotropic blur like simple motion blur. Though blur classification is simplistic in practice, we nevertheless try to identify the blur type. Note that linear motion blur smooths image along and opposite to the motion direction and meanwhile maintain or even enhance the contrast perpendicular to the motion direction. Measuring the coherence of gradient angles should therefore add the same or the opposite directions but cancel out the orthogonal directions. As proposed by Jang et al. [15], doubling and subsequently averaging the gradient vectors can exactly accomplish this goal. Actually as proved in [8], the average doubled gradient angle correlates to the direction, along which the image embodies the minimal high-frequency energy.

The gradient vector field ∇I can be represented by the complex array $I_0 + iI_2$; and doubling arguments (angles in radius) is obtained by squaring the vectors as $\nabla I^2 = I_0^2 - I_2^2 + i2I_0I_2$. Adding the angle-doubled gradients may cancel out the perpendicular gradients and yet consolidate the opposite gradients, as illustrated in Fig. 4. In Fig. 3, the angle-doubled gradient maps are visualized with uniform-lightness color images, with the double angle as hue and the magnitude as saturation. Coherent motion directions leads to concentrated colors. On one hand the energy of average angle-doubled gradient (EDG) indicates the coherence.

$$\left|\overline{\nabla I^2}\right| = \mathbf{E}[I_0(x,y)^2 - I_2(x,y)^2]^2 + \mathbf{E}[2I_0(x,y)I_2(x,y)]^2$$

The bigger the magnitude is, the more directionally coherent the gradients are. On the other hand, the average energy of angle-doubled gradients (ADG), as

$$\begin{aligned}\left|\overline{\nabla I^2}\right| &= \mathbf{E}\left[(I_0(x,y)^2 - I_2(x,y)^2)^2 + \left(2I_0(x,y)I_2(x,y)\right)^2\right] \\ &= \mathbf{E}[(I_0(x,y)^2 + I_2(x,y)^2)^2]\end{aligned} \tag{5}$$

equals to the average squared gradient energy, and reflects the contrast of textures and edges in an image. Chen et al. argue that ADG can identify whether the image (patch) is motion blurred or out-of-focus blurred [8], while Jiang et al. suggest that EDG normalized by ADG can identify blur type more inconsistently with scene change [15]. We on one hand select ADG as a feature of our metric, and on the other hand normalize MDG by ADG to obtain the normalized energy of average angle-doubled gradient (NDG):

$$\left|\widetilde{\nabla I^2}\right| = \left|\overline{\nabla I^2}\right| / \overline{\left|\nabla I^2\right|}$$

To identify heavier blur, we use multi-scale analysis again and define the product of NDGs across image scales as

$$\mathrm{PNDG} = \left|\widetilde{\nabla I^2}\right|_{2^0}\left|\widetilde{\nabla I^2}\right|_{2^1}\left|\widetilde{\nabla I^2}\right|_{2^2}\left|\widetilde{\nabla I^2}\right|_{2^3} \tag{6}$$

3.5. Moment of sharp gradients (MSG)

How a local patch contributes to perceptual global blurriness may depend on where it is. For example, an in-focus subject against out-of-focus background, which is often created in a shallow DOF (depth of field), will look distinctively sharper. Therefore, the distribution of sharp regions and blur regions should be taken into account.

First of all, we classify an image into a shallow DOF or a deep DOF by analyzing the concentration of sharp regions. To this end, we estimate a patch-wise binary sharpness map for an image. During computing the aforementioned MAG, we have located the *k*-largest gradients in an image. A patch that contains more than *T* top-*k*-largest gradients belongs to the sharp patch set, denoted by $p\in P_s$, otherwise the blur patch set, denoted by $p \in P_s$. As a result, concentrated large gradients will cause fewer sharp patches while scattered large gradients will generate more sharp patches. The resultant gradient maps are shown in the bottom row of Fig. 3, where sharp patches are marked bright and blur patches are assigned dark. Accordingly, an image with less than 40% sharp patches is classified as a shallow DOF scene, denoted by $I \in S_s$ otherwise as a deep DOF scene, denoted by $I \in S_d$.

Then, we analyse the distribution of sharp and blur regions. In photography, the subjects are typically framed and composed at the center or at the one third in the image space (known as rule of thirds), where we mark as *composition reference points*. We define *moment arm* as the city block distance from the considered point (x,y) to the nearest composition reference point, as

$$a(x,y) = \min\left\{\left|x-\frac{W}{3}\right|,\left|x-\frac{W}{2}\right|,\left|x-\frac{2W}{3}\right|\right\} + \min\left\{\left|y-\frac{H}{3}\right|,\left|y-\frac{H}{2}\right|,\left|y-\frac{2H}{3}\right|\right\}$$

where *W* and *H* are the width and height of the image. We further define the sharp moments as the average transformed moment arms for all sharp points,

$$\forall (x,y) \in P_s : m_s = \mathbf{E}\left[\frac{1}{1+r\cdot a(x,y)}\right]$$

and the blur moments as the average transformed moment arms for all blur points.

$$\forall (x,y) \in P_b : m_b = \mathbf{E}\left[\frac{1}{1+r\cdot a(x,y)}\right]$$

Finally,we calculate the feature for shallow DOF and deep DOF scene separately:

$$\mathrm{MSG} = \begin{cases} m_s & \text{if } I \in S_s \\ 1-m_b & \text{if } I \in S_d \end{cases} \tag{7}$$

4. ADAPTIVE REGRESSION ON MULTIPLE DATASETS

Given an image set $\{I_i\}$ with the mean opinion scores about perceptual sharpness $\{MOS_i\}$ and a group of selective features $\{\boldsymbol{f}\}$, we look for a metric $q(\boldsymbol{f};\boldsymbol{\beta})$ with parameters $\boldsymbol{\beta}$, to maximize the likelihoods:

$$\max_{\boldsymbol{\beta}} \sum_i \mathcal{L}(q_i(\boldsymbol{f};\boldsymbol{\beta}); MOS_i)$$

If multiple image sets are available, we maximize the total likelihoods among all the sets:

$$\max_{\boldsymbol{\beta}} \sum_j \sum_i \mathcal{L}(q_{i,j}(\boldsymbol{f};\boldsymbol{\beta}); MOS_{i,j}) \tag{8}$$

where the datasets are indexed by *j*.

Subjective opinions about perceptual sharpness are always bounded, like other psychological measurements. That is, opinion scores will approach the lower bound for the worst quality and the upper bound the best quality, which are called the flooring and the ceiling effects.

A key challenge of analyzing multiple datasets is that the mean opinion scores cannot be mixed up straightforwardly. This is because the flooring and ceiling effects rely on the context of the test materials and may vary across datasets. Actually, a human subject can hardly valuate an isolated photo without being demonstrated how the “best” and the “worst” ones look like. The test conditions between the datasets were factually inconsistent here. So the transform from the sharpness measure towards the MOS cannot be fixed since the context is not consistent.

To span the gap, a standard of method is to align the MOSs via the “anchor” samples shared between datasets [3]. However, the datasets here cannot be aligned in that way, for lack of intersect images as anchors. We propose adaptive logistic regression for the problem as following. Logistic modeling is suitable for psychological measurement, since it fuses features and maps them to a bounded interval. A logistic model can be written as

$$q = \frac{1}{1+\exp(\boldsymbol{\beta}^T\cdot \boldsymbol{f} + b)} \tag{9}$$

where $\boldsymbol{\beta}$ is a parameter vector with the equal length as feature vector ***f*** and *b* is a scalar parameter. Traditionally, *b* is often merged into $\boldsymbol{\beta}$ as β_0. However, we isolate *b* and adapt it to each dataset. That is, we use a unique $\boldsymbol{\beta}$ for all datasets and yet $\{b_j\}$ for the datasets indexed with *j*. So Eq.(8) is rewritten as

$$\max_{\beta,b}\sum_{j}\sum_{i}\mathcal{L}\big(q_{i,j}(\boldsymbol{f}:\beta,b_j):MOS_{i,j}\big) \tag{10}$$

We assume the opinion scores as Gaussian distributed variables and therefore instantiate the likelihood as mean squared error. Eq.(10) is derived as

$$\min_{\beta,b}\sum_{j}\sum_{i}\left[MOS_{i,j}-\frac{1}{1+\exp(\beta^{T}\cdot\boldsymbol{f}_{i,j}+b_j)}\right]^2 \tag{11}$$

Parameters b_j controls the shape of the sigmoid logistic curve and adapts the predicted quality scores *q* towards the MOSs for each dataset. They compensate the misaligned flooring and ceiling effects across datasets. It is interesting that *b* does not change the rank order of the predict quality scores for each dataset. Instead, it is $\boldsymbol{\beta}$ that determines the rank order of the predicted quality.

It is straightforward to use the nonlinear regression toolbox of MATLAB to solve $\boldsymbol{\beta}$ with a gradient-descent method. The convergence is usually guaranteed during our ample random tests. Moreover, the regression toolbox also provides the statistical significance test for the feature evaluation, which will be discussed in the next section.

5. EXPERIMENT

5.1. Protocol

We collect nine datasets, including the publicly-available databases and our own database. We extract the subsets of Gaussian blurred images, from the public databases LIVE [24], IVC [18], A57 [6], TID2008 [22], CSIQ [17], VCL_FER [30], and TID2013 [21]. These images are synthesized by convoluting sharp reference images with spatially-invariant Gaussian kernels. UFRJ database [10] and our database, contain realistic blurry photos. In UFRJ database, the photos are captured with digital compact cameras, and wherein the blur is further classified and labeled as out-of-focus, simple motion, complex motion and “other” type. Our database contains the photos snapped with phone cameras and wearable cameras in daily life. Its samples are the failure cases during our many rounds of redesign and retest (refer to the supplementary material for more detail). We believe that those data allow us avoid overfitting on limited and biased data. Both UFRJ and our database keep the JPEG EXIF information intact for each image.

The full datasets contain a total of 1577 blurry images and the associated MOSs. The MOSs represents the ground truth of image sharpness or blurriness, and is used to evaluate the prediction accuracy of metric. In experiment, we normalize all subjective scores to the range [0,1]; a MOS of 0 indicates the worst quality (the blurriest) while a MOS of 1 represents the best quality (the sharpest).

5.2. Evaluation criterion

The model accuracy is evaluated using the Spearman’s rank order correlation coefficient (SROCC) ρ_s between the predicted and the subjective quality score series. ρ_s has a range of $[-1, 1]$: the higher the value, the better the accuracy. Random predictions will achieve a ρ_s value of about 0.

ρ_s evaluates the ordinal match between two score series, and thus remains invariant with any a monotonic mapping of the series. In other words, the accuracy in terms of ρ_s does not rely on any curve-fitting procedure. Such a curve-fitting procedure is, however, inevitable and sensitive for computing other criteria, like Pearson's linear correlation coefficient (LCC) and root-mean-squared error (RMSE).

5.3 Feature evaluation

The selective features $\{f\}$ includes MAG, MGR_1, MGR_2, AGK, ADG, PNDG, MSG as well as EXP (exposure time), as listed in Table 1. The exposure time is recorded in the JPEG EXIF data. For the image without the exposure time information, we set EXP = 0.01 second (at such a shutter speed, the photo just tends not to blur). We compared the proposed features and state-of-the-art metrics over the datasets. The accuracy for each type of blur is plotted in the spider chart of Fig. 5, where the radial axis indicates the correlation ρ_s ranging from −0.4 at the center origin to 1 at the outermost square grid. Four types of blur are counted. For Gaussian blur (the upward direction in Fig. 5), we compute the average ρ_s on the seven Gaussian blur datasets, which cover a total of 687 images. For out-of-focus (rightward), simple motion (downward) and complex motion (leftward) blur, we calculate the ρ_s on the corresponding subsets of UFRJ database, which contain 141, 57 and 62 images respectively.

Table 1. Abbreviations and definitions of features and metrics.

Abbr. (Def.)	**Feature / metric operator [Ref.]**
ADG (5)	Average energy of angle-Doubled Gradients
AGK (4)	Average Gradient Kurtosis
EXP	Exposure time in JPEG EXIF
GRA1	Gaussian derivative [20]
GRA2	Gradient energy [20]
GRA4	Squared gradient [20]
LPC_SI	Local phase congruency sharpness index [14]
MAG (1)	Maximal Gradient
MGR_1 (2)	Minimal 2nd Gradient Ratio at Scale 1 & 2
MGR_2 (3)	Minimal 2nd Gradient Ratio at Scale 3 & 4
MIS9	Vollath's autocorrelation [20]
MSG (7)	Moment of Sharp Gradients
PNDG (6)	Product of Normalized energy of average angle-Doubled Gradient
STA8	Histogram range [20]

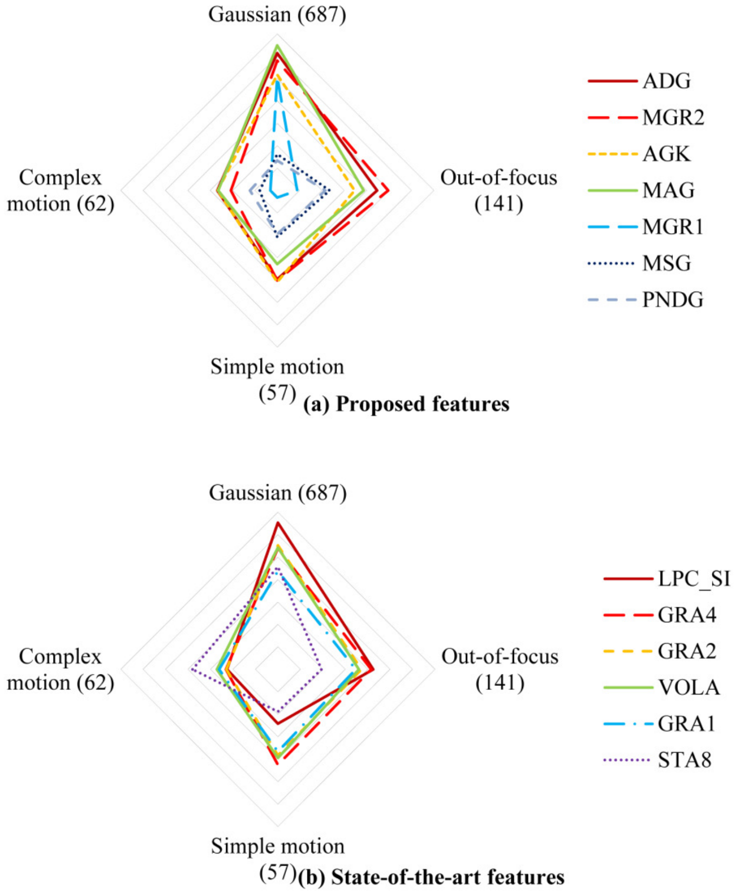
Table 2. Statistical significance of features in regression

Feature f_m	**95% CI of Parm.** β_m	p **value of Parm.** β_m
EXP	0.290±0.042	4.0×10^{-39}
MGR_2	−1.07±0.19	1.2×10^{-28}
MGR_1	−0.363±0.099	1.0×10^{-12}
PNDG	0.0796±0.0277	1.0×10^{-11}
ADG	−0.346±0.142	2.1×10^{-6}
MAG	0.242±0.142	9.0×10^{-4}
AGK	−0.359±0.221	1.5×10^{-3}
MSG	−0.762±0.586	1.1×10^{-2}

These experimental results give us a first impression about the capability of the features. Among the proposed features, ADG has the best overall accuracy on all datasets; MGR_2 ranks the second in overall and is specially good at "simple motion" and "out-of-focus" blur; AGK is generally accurate and does especially well in motion blur; MAG achieves the state-of-the-art accuracy on

Gaussian blur; and the other feature MGR_1, MSG or PNDG alone appears not to correlate with each type of blur very well.

We compare with the state-of-the-art Gaussian blur metric LPC_SI [14], the best focus metric for microscope MIS9 [13], and the recommended focus metrics for OCR, i.e. GRA1, GRA2, GRA4, and STA8 [23] (see their definitions in [20]). Considering the performance of the existing metrics as shown in Fig. 5(b), the Gaussian blur is the easiest to predict, the out-of-focus blur is also easy to handle, and yet the motion blur are much more challenging to measure. This is partly because the Gaussian blur is synthetic and ideally spatially-invariant, the out-of-focus blur here is also nearly spatially-invariant since the images with shallow DOF have been picked out to the “other” type of blur in UFRJ database, but the motion blur here is rarely coincidental with spatially-variant blur. Another reason is that the Gaussian blur images, although abound here, are synthesized and derived from a few reference images, but the images in UFRJ database have more diverse scenes.



(a) Proposed features

(b) State-of-the-art features

Figure 5. Correletion ρ_s between features and single type of perceptual blur, in terms of spider chart

Feature selection not only depends on the performance of using each feature alone, but also relies on that of using feature combinations. The latter point can be evaluated by the statistical significance test. For logistic regression model, the *p* value and the CI (confidence interval) can indicate the significance of the feature variables. On one hand, the smaller the *p* values is, the more confident the corresponding feature is. On the other hand, the 95% CI (confidence interval) of β_m does not cover 0 means that feature f_m is sufficiently confident. In this paper, we omit the process of feature selection but show the statistical significance of the final feature combination in Table 2. Parameters $\beta=\{\beta_m\}$ are obtained by nonlinear regression on the full datasets. All the

95% CIs are apart from 0, so it is confident that the selective features are powerful. It is worth to mention that MGR_1 and PNDG correspond to quite small p values in Table 2, despite low ρ_s values in Fig. 5. It means that MGR_1 and PNDG themselves alone are weak but they are really helpful to the feature combination. Moreover, the image composition related feature MSG also plays a fairly significant role, for a p value of 0.01.

5.4 Metric comparison

With the selective features and adaptive logistic regression, we obtain the metric as Eq. (9). The accuracy of metrics is compared in Fig. 6. In Fig. 6, the radial axis still indicates ρ_s on each dataset ranging from -0.4 at the centre to 1 at the outermost decagon grid. In the upward direction, we compute the weighted average ρ_s on all datasets, which is weighed by the number of images in each dataset.

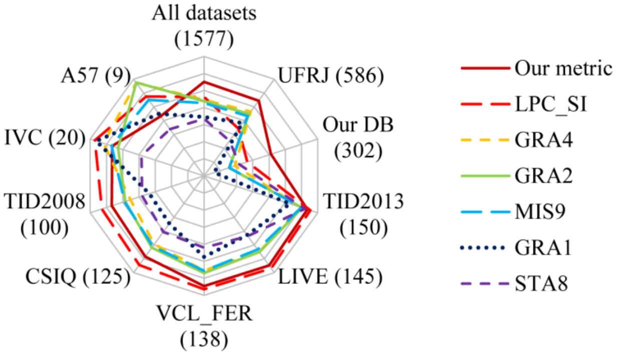


Figure 6. Correlation ρ_s of metrics on each dataset

For a fair comparison, we report the cross validation result of our metric. That is, for each time, we randomly divide each dataset into two segments, 50% for training and the other 50% for testing, find a unique set of parameters $\{\boldsymbol{\beta}\}$ by training, and test it to obtain a set of accuracy on every dataset. Running the procedure for 100 times, we compute the average accuracy.

As a result, our metric achieves the best overall accuracy on all databases. It outperforms other metrics on the realistic photos datasets, UFRJ and our database. UFRJ database is quite challenging. To the best of our knowledge, only two papers disclose the result of their proposal on UFRJ database; the authors of UFRJ database report ρ_s of 0.56±0.04 [10], and Chen et al. claim ρ_s of 0.586 [7]. Our metric attains ρ_s of 0.688±0.052 in cross validation. We use the same features but replace logistic regression with SVR (supporting vector regression), and obtain a ρ_s of 0.631 in cross validation. Our database is even tougher, since it contains the failure cases during our past repetitive trials, such as blurry but high-contrast images (e.g., textural, noisy and overexposure scenes) as well as sharp but low-contrast scenes (e.g., sky, lake and nightscape), as shown in Fig. 7. On our database, LPC_SI only obtains ρ_s of 0.132 and most existing metrics even get a negative ρ_s, which are not better than a random prediction. However, our metric attains ρ_s of 0.424.

For most of the synthetic image datasets, our metric achieves comparable accuracy as the state-of-the art approach LPC_SI. It is inferior to most metrics only on A57 and IVC. These two datasets are too small to guarantee unbiased random divisions in cross validations. Note that the number of samples in each dataset is annotated in the parenthesis in Fig. 6.

6. CONCLUSION

A no-reference sharpness metric is proposed and verified efficient for realistic data. It is nothing special, but comprises a set of nonlinear statistics on image gradients. The assorted of statistics are closely related to various aspects of pooling strategy; the operators maximum, variance, and kurtosis accentuate the steepest gradients, the pixel-wise average, patch-wise average, and pyramid analysis merge gradients in multi-scales, and the moments based on image composition weigh gradients with visual saliency, while the operators minimum and vector mean attribute gradients to the outcome of different blur type. It is important that those statistics combination is "selected"' by a statistic modelling on data, more than a mere handcraft design. Nonetheless, perception on blur involves with high-level vision and goes beyond the proposed statistics. There is still substantial room to improve the measurement by incorporating high-level features.

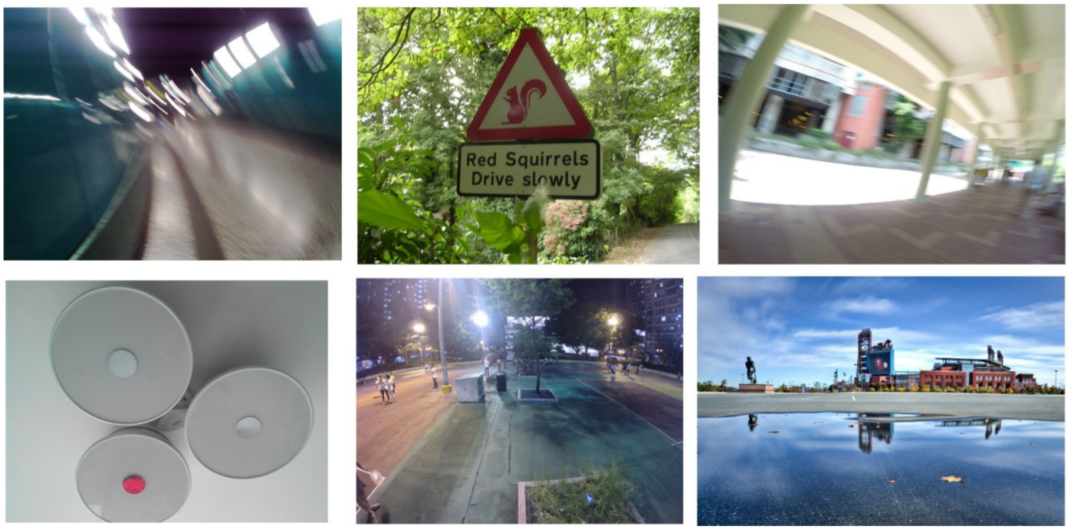


Figure 7 Failure cases of sharpness overesimated (top) and underestimated (bottom) in our database.

REFERENCES

[1] Bae, S.& Durand, F. "Defocus magnification". In Computer Graphics Forum (2007), vol. 26,Wiley Online Library, pp. 571–579.

[2] Bahrami, K. & Kot, A. "A fast approach for no reference image sharpness assessment based on maximum local variation". Signal Processing Letters, IEEE 21, 6 (June 2014), 751–755.

[3] Brill, M. H., Lubin, J., Costa, P., Wolf, S., & Pearson, J. "Accuracy and cross-calibration of video quality metrics: new methods from atis/t1a1". Signal Processing: Image Communication 19, 2 (2004), 101–107. 6

[4] Caviedes, J. & Gurbuz, S. "No-reference sharpness metric based on local edge kurtosis". In Image Processing Proceedings. International Conference on (2002), vol. 3, IEEE, pp. III–53.

[5] Chakpabarti, A., Zickler, T. & Freeman, W. T. "Analyzing spatially-varying blur". In Computer Vision and Pattern Recognition (CVPR), (2010), IEEE, pp. 2512–2519.

[6] Chandler, D. & Hemami, S. A57 database, 2007.

[7] Chen, M.-J. & Bovik, A. C. "No-reference image blur assessment using multiscale gradient". EURASIP Journal on Image and Video Processing 2011, 1 (2011), 1–11.

[8] Chen, X., Yang, J., Wu, Q., Zhao, J. & He, X. “Directional high-pass filter for blurry image analysis”. Signal Processing: Image Communication 27, 7 (2012), 760–771.

[9] Chiang, M.-C. & Boult, T. E. “Local blur estimation and super-resolution”. In IEEE Conference on Computer Vision and Pattern Recognition (1997), IEEE Computer Society, pp. 821–821.

[10] Ciancio, A., Da Costa, A. T., Da Silva, E. A., Said, A., Samadani, R. & Obrador, P. “No-reference blur assessment of digital pictures based on multifeature classifiers”. Image Processing, IEEE Transactions on 20, 1 (2011), 64–75.

[11] Dai, S. & Wu, Y. “Removing partial blur in a single image”. In Computer Vision and Pattern Recognition, CVPR. IEEE Conference on (2009), IEEE, pp. 2544–2551.

[12] Gennery, D. B. “Determination of optical transfer function by inspection of frequency-domain plot”. JOSA 63, 12 (1973), 1571–1577.

[13] Groen, F. C., Young, I. T. & Lightart, G. “A comparison of different focus functions for use in autofocus algorithms”. Cytometry 6, 2 (1985), 81–91.

[14] Hassen, R., Wang, Z. & Salam, M. M. “Image sharpness assessment based on local phase coherence”. Image Processing, IEEE Transactions on 22, 7 (2013), 2798–2810.

[15] Jang, S.-I., Chung, J., Lee, Y., Chung, K., Kim, W. & Lee, C.-W. “A real-time identification method on motion and out of focus blur for a video camera”. Consumer Electronics, IEEE Transactions on 40, 2 (1994), 145–153.

[16] Ji, H. & Liu, C. “Motion blur identification from image gradients. In Computer Vision and Pattern Recognition”, CVPR IEEE Conference on (2008), IEEE, pp. 1–8.

[17] Larson, E. & Chandler, D. Categorical image quality (CSIQ) database 2009.

[18] Le Callet, P., Autrusseau, F., et al. Subjective quality assessment IRCCyN/IVC database.

[19] Liu, R., Li, Z. & Jia, J. “Image partial blur detection and classification”. In Computer Vision and Pattern Recognition, IEEE Conference on (2008), IEEE, pp. 1–8.

[20] Pertuz, S., Puig, D. & Garcia, M. A. “Analysis of focus measure operators for shape from focus”. Pattern Recognition 46, 5 (2013), 1415–1432.

[21] Ponomarenko, N., Ieremeiev, O., Lukin, V., et al. “Color image database TID2013: Peculiarities and preliminary results”. In Visual Information Processing (EUVIP), 4th European Workshop on (2013), IEEE, pp. 106–111.

[22] Ponomarenko N., Lukin, V., Zelensky, A., et al. TID2008 - a database for evaluation of full-reference visual quality assessment metrics. Advances of Modern Radio Electronics 10, 4 (2009), 30–45.

[23] Rustin Ol, M., Chazalon, J. & Ogier, J.-M. “Combining focus measure operators to predict ocr accuracy in mobile-captured document images”. In Document Analysis Systems (DAS), 11th IAPR International Workshop on (2014), IEEE, pp. 181–185.

[24] Sheikh, H. R., Wang, Z., Cormack, L. & Bovik, A. C. Live image quality assessment database release 2, 2005

[25] Shi, J., Xu, L. & Jia, J. “Discriminative blur detection features”. In Computer Vision and Pattern Recognition, IEEE Conference on (2014), IEEE.

[26] Su, B., Lu, S. & Tan, C. L. “Blurred image region detection and classification”. In Proceedings of the 19th ACM international conference on Multimedia (2011), ACM, pp. 1397–1400.

[27] Tang, H., Joshi, N. & Kapoor, A. “Blind image quality assessment using semi supervised rectifier networks”. In Computer Vision and Pattern Recognition, IEEE Conference on (2014), IEEE.

[28] Wu, S., Lin, W., Xie, S., Lu, Z., Ong, E. P. & Yao,S. “Blind blur assessment for vision-based applications”. Journal of Visual Communication and Image Representation 20, 4 (2009), 231–241.

[29] Ye, P., Kumar, J., Kang, L. & Doermann, D. “Real-time no-reference image quality assessment based on filter learning”. In Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on (2013), IEEE, pp. 987–994.

[30] Zari, A., Tatalovi, N., Brajkovi, N., et al. FER image quality assessment database. AUTOMATIKA: 53, 4 (2012), 344–354. 6

[31] Zhang, N. F., Postek, M. T., Larrabee, R. D., et al., “Image sharpness measurement in the scanning electron microscope (part iii)”. Scanning 21, 4 (1999), 246–252.

[32] Zhuo, S. & Sim, T. “Defocus map estimation from a single image”. Pattern Recognition 44, 9 (2011), 1852–1858.

INTENTIONAL BLANK

A BINARY TO RESIDUE CONVERSION USING NEW PROPOSED NON-COPRIME MODULI SET

Mansour Bader[1], Andraws Swidan[2] , Mazin Al-hadidi[3] and Baha Rababah[4]

[1]Department of Computer Engineering, Jordan University, Amman, Jordan
mansoor259@yahoo.com
[2]Department of Computer Engineering, Jordan University, Amman, Jordan
sweidan@ju.edu.jo
[3]Department of Computer Engineering, Al-Balqa'a Applied University, Salt, Jordan
trueamman@yahoo.com
[4]Department of Computer Engineering, University of Portsmouth, Portsmouth, UK
baharababah@yahoo.com

ABSTRACT

Residue Number System is generally supposed to use co-prime moduli set. Non-coprime moduli sets are a field in RNS which is little studied. That's why this work was devoted to them. The resources that discuss non-coprime in RNS are very limited. For the previous reasons, this paper analyses the RNS conversion using suggested non-coprime moduli set.

This paper suggests a new non-coprime moduli set and investigates its performance. The suggested new moduli set has the general representation as {2^n–2, 2^n, 2^n+2}, where n ∈ {2,3,…..,∞}. The calculations among the moduli are done with this n value. These moduli are 2 spaces apart on the numbers line from each other. This range helps in the algorithm's calculations as to be shown.

The proposed non-coprime moduli set is investigated. Conversion algorithm from Binary to Residue is developed. Correctness of the algorithm was obtained through simulation program. Conversion algorithm is implemented.

KEYWORDS

Forward Conversion, Residue Number System, Non-coprime Moduli Set

1. INTRODUCTION

Residue number system (RNS) is a subfield of finite field arithmetic [1]. It is widely used in digital signal processing, image processing, FIR (Finite Impulse Response) filters, and IIR (Infinite Impulse Response) filters because it is a carry-free system and high efficient in addition and multiplication [2]. So, residue number system is used by most applications that need a high degree of concurrency. A lot of researches in computer systems are enthusiastic to go through

Natarajan Meghanathan et al. (Eds) : NETCOM, NCS, WiMoNe, GRAPH-HOC, SPM, CSEIT - 2016
pp. 147– 159, 2016. © CS & IT-CSCP 2016 DOI : 10.5121/csit.2016.61513

residue numbering system because of its characteristics such as, error detection and correction (fault tolerant) [3], modularity, and embedded parallelism.

RNS allows dividing a large number into smaller sub numbers. Numbers are represented by tipples which need less number of bits. The bits can be processed individually and in parallel without carry between them. This improves computation time and simplifies hardware implementation cost.

RNS has also the following advantages over conventional binary number system:

- Reducing the hardware complexity because the system is implemented by designing smaller processing units.
- Improving the speed of operations since all of the tasks are performed in parallel.
- Efficient realization of the various building blocks needed such as adders, multipliers.
- The absence of carry propagation between the modulus channels makes the RNS appealing for building parallel fast processors. This facilitates the realization of high-speed, low-power arithmetic. This advantage is of paramount importance in embedded processors, especially those found in portable devices, for which power consumption is the most critical aspect of the design [4].

Because of these features, computer arithmeticians have historically promoted the RNS for high-speed arithmetic-intensive applications [5].

The rest of this paper is organized as follows. In Section 2, overview of the new algorithm of forward conversion is proposed. Section 3 presents the new non-coprime realization of the proposed forward converter. The hardware implementation of the moduli set is presented in section 4, while the paper is concluded in Section 5.

2. NEW ALGORITHM OF THE FORWARD CONVERSION OVERVIEW

In working with RNS the following three main terminologies are used:

1. Moduli set: defined in terms of relatively prime moduli where the i^{th} modulus presented by m_i and the $\gcd(m_i, m_j) = 1, j \neq i, i = 1,2, \ldots, n$. Numerous moduli sets can be used. The characteristics of RNS based systems depend on the moduli set chosen.
2. Dynamic range (M): this is equal to the product of m_i terms; $M = \prod_{i=1}^{n} m_i$, and denotes the interval of integers that can be represented uniquely in the RNS using the specific moduli set.
3. Residues: to represent any number X in RNS we find $x_i = X \bmod m_i$ for all m_i moduli. The number is represented as $X = x_1, x_2, x_3$.

RNS based processing units are generally composed of: Forward convertor, arithmetic and logic unit (ALU) and a reverse convertor shown in figure 1 [6].

Conversion from Binary to Residue is called forward conversion. This conversion is used in order to process numbers in Residue format, because it is faster and it is easier for human being understood. To use the Residue Number System efficiently one has to interface it with the real world, the numbers should be converted from usual representation either binary or analog to residue representation, this is the first step in using RNS. This step is a very complex and demanding process, which acts as an academic challenge that restrict the use of RNS in many practical applications. Many researches conducted to find the most efficient algorithm, hardware architectures and schemes either using special or arbitrary moduli sets in the implementation of forward convertors in RNS.

In the forward translation of a binary number to its RNS equivalent, one of the most trivial, classical expensive ways is to store all the residues and recall them based on the value of the binary input.

Using the fact that the number can be represented as:

$$X = x_{n-1}\, x_{n-2}\ \ldots x_1 x_0 = \sum_{j=0}^{n-1} x_j 2^j \qquad (1)$$

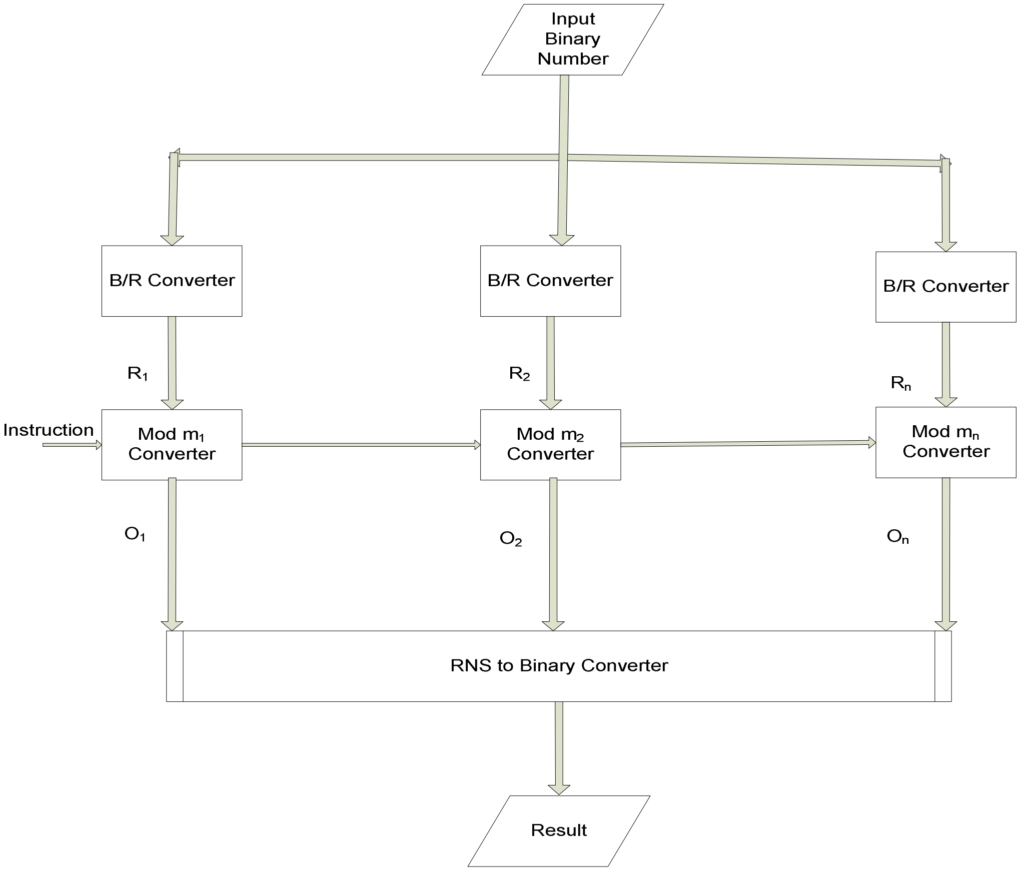


Figure 1. Block diagram of RNS based processing.

It is clear that this is a memory consuming process where you have to store all values in a lookup table that typically consists of ROM, and for complex applications that require large number representation, the size of memory will increase dramatically and thus increasing the cost.

The other implementation is to have special moduli set representation used in the conversion process, consist of three, four, or five bits, the design of these convertors is based on using carry save adder (CSA).

2.1. Non-coprime Moduli Sets Overview

Initially, only RNS with co-prime moduli set was investigated and used. Non-coprime had drawn the attention of research only lately. Thus the knowledge of non-coprime characteristics has been obtained from the co-prime one's as going to be seen through this chapter sections.

It is relatively easy to convert even numbers into their residue numbers representation. Numbers as 2^n, like 16 which is pow(2,4) or in form that we are familiar with now 2^4, and 14 which is 2^4 – 2 and 18 which is 2^4 +2, are not co-primed with each other since there is a common factor between them which is number 2.

The usual way to compute **a mod m** is to take the remainder after integer division. This is straight forward when the operands are within the range of the available divide hardware, but the divide operation is known to be a slow arithmetic operation. Some small microcontrollers have no divide hardware, and it is occasionally necessary to divide very large numbers outside the range that can be done using the available hardware [7].

It can be faster to take the modulus directly than to use the divide instruction when the modulus m is constant, even where there is a hardware divide instruction. Rules of divisibility mentioned in table 1 become even more valuable on machines without a hardware divide instruction or where the numbers involved are out of range.

Table 1. Some of divisibility check rules applied to decimal system.

Divisibility by n	Check
10	The least significant decimal digit is zero.
2	The least significant decimal digit is even.
5	The least significant decimal digit is 0 or 5
3	Sum of the decimal digits is divisible by 3.
9	Sum of the decimal digits is divisible by 9

2.2. Some Math Identities Review

2.2.1. Single General Rule

All divisibility check rules mentioned in table 1 are actually special cases of a single general rule. Given that:

a is represented in number base b

a mod m = ((b mod m)(a/b) + (a mod b)) mod m

In the case of divisibility by 2, 5 and 10 for base 10, the term (b mod m) is zero because 2, 5 and 10 all divide evenly into 10. As a result, the divisibility test simplifies to asking whether (a mod b), that is, the least significant digit of the number, is evenly divisible.

In the case of divisibility by 3 or 9 in base 10, the term (b mod m) is one. As a result, the multiplier for the first term is one. Applying the formula recursively leads to the simple sum of the digits [7].

2.2.2. The Trivial Case: Mod 2, Mod 4, Mod 2^n

Computing modulus for powers of two is trivial on a binary computer, the term (b mod m) is zero, so we just take the modulus by examining the least significant i bits of the binary representation:

$a \bmod 2^i = a \,\&\, (2^i - 1)$

Thus, for a mod 2, we use a & 1, for a mod 4, we use a & 3, and for a mod 8, we use a & 7.

Recall that the **&** operator means logical and. When applied to integers, this computes each bit of the result as the and of the corresponding bits of the operands. For all nonzero positive integers i, the binary representation of 2^i –1 consists of i consecutive one bits, so anding with 2^i –1 preserves the least significant i bits of the operand while forcing all more significant bits to zero [7].

The problem is more interesting when the modulus is not a power of two only.

2.2.3. Mersenne's Number: Mod 3, Mod 7, Mod 2^n-1

In mathematics, a Mersenne prime is a prime number that is one less than a power of two, i.e. 2^n-1. That is, it is a prime number that can be written in the form Mn = 2^n − 1 for some integer n. They are named after Marin Mersenne, a French Minim friar, who studied them in the early 17th century [8].

Consider the problem of computing a mod 3 in binary number system. Note that 4 mod 3 is 1, so:

a mod 3 = ((a/4) + (a mod 4)) mod 3

That is, a mod 3 can be computed from the sum of the digits of the number in base 4. Base 4 is convenient because each base 4 digit of the number consists of 2 bits of the binary represenation; thus a mod 4 can be computed using a & 3 and a / 4 can be computed using a >> 2.

The number 3 is a Mersenne number, that is, one less than a power of two. The property noted above is true of all Mersenne numbers. Thus, we can compute a mod 7 or a mod 15 on a binary computer using:

a mod 7 = ((a/8) + (a mod 8)) mod 7
a mod 15 = ((a/16) + (a mod 16)) mod 15

Recall that a >> b shifts the binary representation of **a** left a total of **b** places. As with logical and, this is a very inexpensive operation on a binary computer, and the effect is the same as dividing a by 2b [7].

In this paper the problem is more interesting when the modulus is in a different shape of a power of two, where it is in Mod (2^n – 2), Mod (2^n + 2) consequently as a new moduli set proposed along with Mod (2^n), that are going to be discussed in the next section.

Our work is done by suggesting the new moduli set {2^n – 2, 2^n , 2^n + 2} and proposing new conversion algorithms upon this new non-coprime moduli set. The next coming section will discuss the background of non-coprime moduli sets.

2.3. Our New Non-coprime Moduli Set Overview

As its name shows "non-coprime" means the non-coprimality among its modulus numbers. Non-coprime moduli sets can be used for error detection and correction purposes [9]. This non-coprimality could be shown in theorems and by examples to prove them too.

Observation 1: 2^k– 2 is not relatively prime to 2^k, where k is a positive natural number.

It is obvious that 1 is not the only prime divisor of 2^k and 2^k– 2, since they are both even and 2^k – 2 is a multiple of 2 (i.e. a number multiplied by 2). Thus, there is a common divisor of them rather than 1, which is number 2 in this case. So 2^k–2 is not relatively prime to 2^k.

Example 1:

Let us take K = 4, in this case the two numbers of the theorem one would be 16 and 14 consequently, and it is obvious that there is a common divisor which is 2 between them when we try to bring them back to their elementary elements.

Observation 2: 2^k+ 2 is not relatively prime to 2^k, where k is a positive natural number.

It is obvious that 1 is not the only prime divisor of 2^k and 2^k+ 2 , since they are both even and 2^k + 2 is a multiple of 2 (i.e. a number multiplied by 2). Thus, there is a common divisor of them other than 1, which is number 2 in this case. So 2^k+2 is not relatively prime to 2^k.

Example2:

Again let us take k = 4, in this case the two numbers of the theorem two would be 16 and 18 consequently, and it is obvious that there is a common divisor which is 2 between them when we try to bring them back to their elementary elements.

2.4. Properties of Non-coprime Moduli Set

2.4.1 Dynamic Range of Our Non-coprime Moduli Set

This property of having a common divisor other than number 1 led to the non-coprime moduli set when gathering the two theorems above. Based on these theorems the moduli set {2n - 2, 2n, 2n + 2} is non-coprime. The dynamic range of co-prime moduli set (M) is equal to $\mathrm{M} = \prod_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{m_i}$. In the non-coprime case it is $\mathrm{M} = (\prod_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{m_i})/4$, it is 1/4 of the size of the co-primed one {2n – 1, 2n, 2n + 1}. This quarter comes from the multiplication of the common divisor (i.e. 2) between its modulus numbers leading to number 4 which can not be multiplied to form the usual (M), thus its size is less than the co-prime one. However it is important to know other characteristics such as its uniqueness and bits representation.

Definition [9]: We define a non-prime moduli set as (ml, ..., mk), where gcd (mi,mj) = l may not be satisfied for some i and j. The least common multiple of the moduli (m1, ..., mk), denoted as: M = lcm(m1, ..., mk), is the dynamic range of (m1, ..., mk) . For any decimal number y∈[O, M – 1), y has a unique representation as (y1, ...,yk), where y = yi mod mi, 0 ≤ yi< mi.

If the residue number (y1, ...,yk) is consistent, the decimal number it represents can be found using the CRT theorem, where M = lcm(m1, ..., mk) [9].

To find M through a formula as the one of the co-prime, we suggest the following formula due to the special non-coprime moduli set:

$$\mathrm{M} = (\prod_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{m_i})/4 \qquad (2)$$

2.4.2 Uniqueness Verification

After dealing with the major and most important part of RNS in the previous section, it is important to discuss the moduli set uniqueness when converted from decimal form (as we see it) into RNS form (as to be implemented). In this section a mathematical proof is going to be presented powered by table 2 to show the uniqueness itself.

As a mathematical proof, we already know that 6, 8 and 10 are not co-primed with each other, in this case the LCM should be used as the previous section showed. Thus we have to return them back into their fundamental factors, for 6 which equal 2*3, for 8 it equals 23 and for 10 it is 2*5, now by taking the non common numbers without repetition and multiplying them we see that:

3*23*5 which equals 120, so M = 120. This is true for all other moduli sets chosen with the form (2n-2 ,2n, 2n +2), and this is the non-coprime moduli set that we have got and found our research on it for thesis level responsibilities, since the pre-moduli value (i.e. 2n -2) and the post moduli one (i.e. 2n +2) are multiples of 2 and are un-coprimed with each other, but they are co-prime with each other without it (as in 3 and 5 for the example above, when neglecting the multiplication of them with number 2). This fact is also true for all other moduli sets used; this is because they are 2 numbers relatively co-primed with each other and they are always odd.

The uniqueness interval was verified by simulation program, the results of a program using the moduli set (6,8,10) was simulated, at which each iteration is repeated every 6*8*10/4 times, which is in this case the number 120 (that is 'M' for the moduli set itself), and thus the uniqueness is guaranteed under the non-coprime moduli set within its dynamic range.

3. FORWARD CONVERSION ALGORITHM

This section has three sub sections of it as the number of the modulus numbers forming the moduli set are three.

3.1 Pre-modulus Algorithm Implementation

The binary representation led to this part's discovery, since its values are prime after dividing it by 2. We will give you how it works in words first then the algorithm of calculation

$|X|2^n$ -2 can be represented in figure 1.

The idea of its binary cutting circles around the common factor which is number 2 in this case, as the previous sections showed that after dividing the pre number by the common factor the result is $2^{n-1} – 1$. Since the pre value has a $2^n – 2$ shape, then the cut of its binary representation would be in (**n** cut at the beginning as the co-prime algorithm did, but for the rest part it is taken (**n-1**) each time) starting from LSB again.

3.2 Middle-modulus Part Algorithm

No change is done on it, so we refer you to [15] – [17] where obtaining the residue of X with respect to modulus 2^n is the easiest operation.

3.3 Post-modulus Algorithm Implementation

The binary representation led to its working algorithm, since its value is prime after dividing it by 2. Again we will give you how it works in words first then we will put them in the flowchart of figure 2 to easily understand it.

The idea of its binary cutting circles around the common factor which is number 2 in this case, as the pre-modulus section showed. Here dividing the post number by the common factor the result is 2^n-1 + 1, and since the post value has a $2^n + 2$ shape, then the cut of its binary representation would be in (n cut at the beginning as the co-prime algorithm did, but for the rest parts it is taken (n-1) each time) starting from LSB again as the pre-modulus did, some examples will show how this is done.

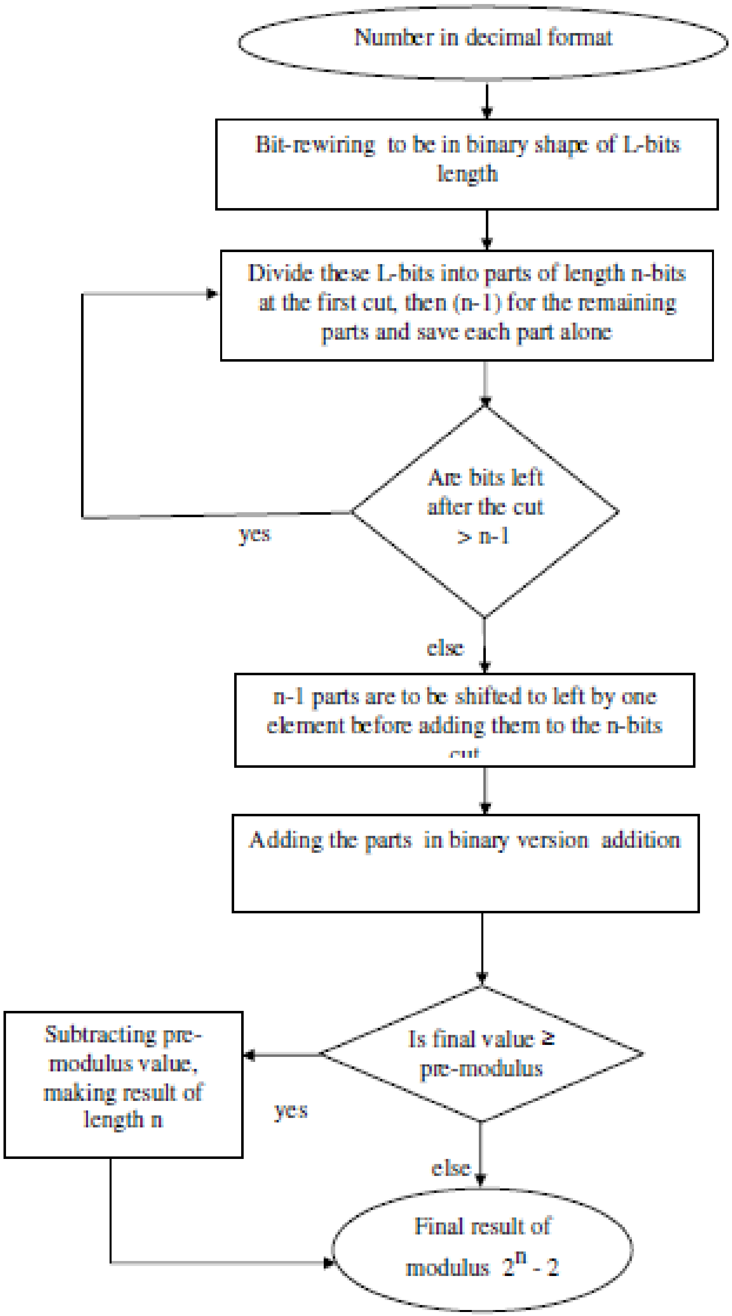

Number in decimal format
Bit-rewiring to be in binary shape of L-bits length
Divide these L-bits into parts of length n-bits at the first cut, then (n-1) for the remaining parts and save each part alone
Are bits left after the cut > n-1
yes
else
n-1 parts are to be shifted to left by one element before adding them to the n-bits cut
Adding the parts in binary version addition
Is final value ≥ pre-modulus
yes
Subtracting pre-modulus value, making result of length n
else
Final result of modulus 2^n - 2

Figure 1. Computing the residue with respect to 2^n-2

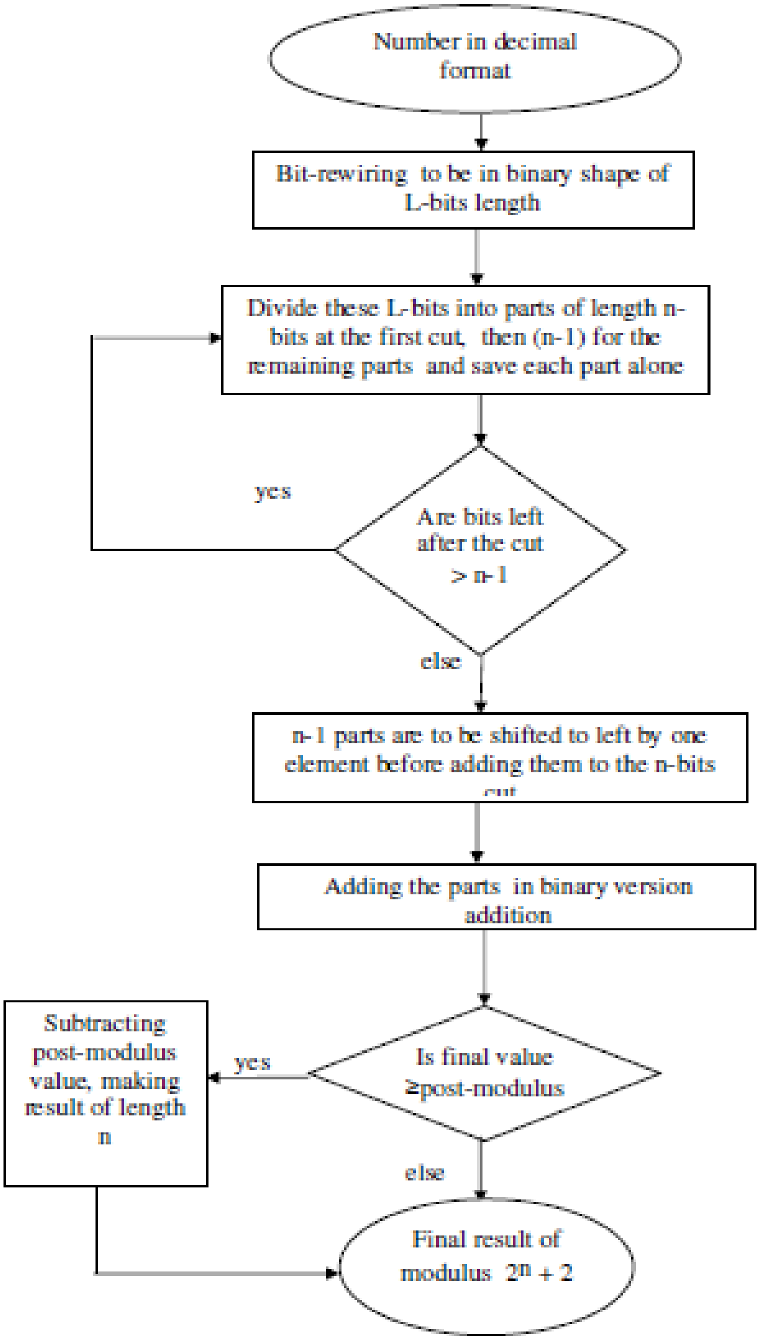

Number in decimal format
Bit-rewiring to be in binary shape of L-bits length
Divide these L-bits into parts of length n-bits at the first cut, then (n-1) for the remaining parts and save each part alone
yes
Are bits left after the cut > n-1
else
n-1 parts are to be shifted to left by one element before adding them to the n-bits cut
Adding the parts in binary version addition
Subtracting post-modulus value, making result of length n
yes
Is final value ≥post-modulus
else
Final result of modulus 2n + 2

Figure 2. Computing the residue with respect to $2^n + 2$

4. ALGORITH IMPLEMENTATION

In this section the implementation of the new non-coprime moduli set converter's implementation is done through hardware block diagrams. As described in the previous section, the converter has

3 stages of computing the modulus; they are (pre-modulus '$2^n - 2$', middle modulus '2^n' and post-modulus '$2^n + 2$'). So this section also consists of three hardware implementation blocks of them.

4.1. Pre-modulus Hardware Implementation

From the proposed algorithm presented for the pre-modulus calculation discussed in the flowchart of figure 1, we can come with this block diagram for the hardware design of it in figure 3.

Notice that the start for cutting is from LSB side to the MSB. The DR for any n value is calculated as was shown in formula (2), thus any number inside the range of M has 3 part cuts -at most-. For example let n = 5, M = 15*32*17 = 8160, so the numbers inside the DR are {0,1,2,……,8159}. However 8159 is represented in binary form as (1111111011111) which consist of 13 digits and it is equal to 5 + 4 + 4 as the cuts presented by its algorithm showed. This is true for all n values, so we will name the first n-bits cut A, the second (n-1) bits cut B and the final part of (n-1) bits is C.

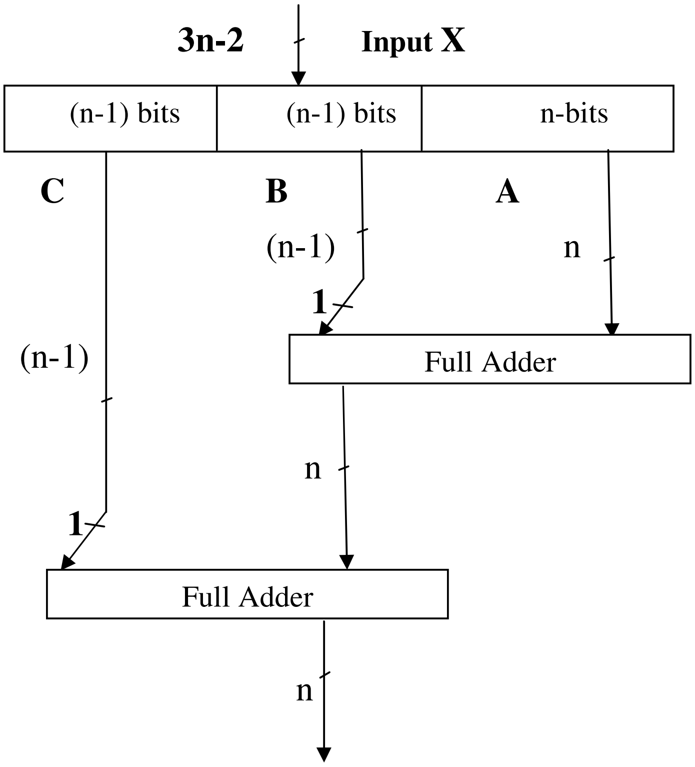


Figure 3. Calculation of Pre-modulus ($2^n - 2$).

4.2 Middle-modulus Hardware Implementation

No change is done on it, so we refer you to [15]-[17]. Obtaining the residue of X with respect to modulus 2^n is the easiest operation. Block diagram of it is shown in figure 4.

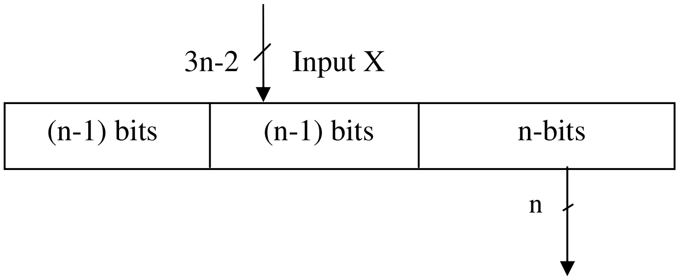


Figure 4. Calculation of Middle-modulus (2^n).

4.3 Post-modulus Hardware Implementation

From the proposed algorithm presented for the post-modulus (2^n + 2) calculation discussed in the flowchart of figure 2, we can come with this block diagram for the hardware design of it in figure 5.

Notice that the start for cutting is from LSB side to the MSB. The DR for any n value is calculated as was shown in formula (2), thus any number inside the range of M has 3 part cuts -at most-. For example let n = 5, M = 15*32*17 = 8160, so the numbers inside the DR are {0,1,2,……,8159}. However 8159 is represented in binary form as (1111111011111) which consists of 13 digits and it is equal to 5 + 4 + 4 as the cuts presented by its algorithm showed. This is true for all n values, so we will name the first n-bits cut A, the second (n-1) bits cut B and the final part of (n-1) bits is C.

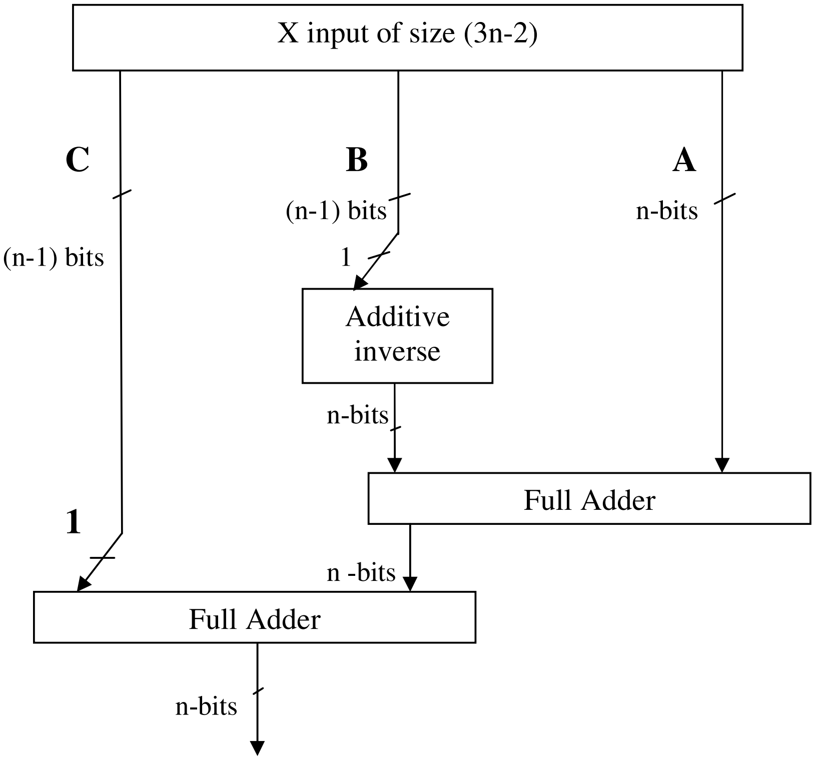


Figure 5. Calculation of Post-modulus (2^n + 2).

5. CONCLUSIONS

A new non-coprime moduli set has been proposed. A general formula for the dynamic range was derived. Algorithm of the special non-coprime moduli set has been suggested. The uniqueness for the new special non-coprime moduli set just as the co-prime one's among DR has been verified.

This research revealed that non-coprime moduli set may be suitable for wide variety of cases not limited to co-prime only.

ACKNOWLEDGEMENTS

The authors would like to thank everyone.

REFERENCES

[1] Neha Singh, (2008), "An overview of Residue Number System". National Seminar on Devices, Circuits & Communication.

[2] Chaves, R., and Sousa, L. (2007) "Improving residue number system multiplication with more balanced moduli sets and enhanced modular arithmetic structures". Computers &Digital Techniques IET, Volume 1, Issue 5, pages 472-480.

[3] Modiri, Samira, Movaghar, and Barati (2012) "Study of error control capability for the new moduli set {22n+ 1+ 2n-1, 22n+ 1-1, 2n-1, 23n, 23n+ 1-1}". Journal of Advanced Computer Science & Technology, Vol. 1, Pages: 176-186.

[4] Abdelfattah (2011) “Data Conversion in Residue Number System”. McGill University.

[5] Bajard and Plantard(2004) "RNS bases and conversions". Proceedings of Advanced Signal Processing Algorithms, Architectures, and Implementations XIV SPIE, Vol. 5559,Page:61-75.

[6] William A. Chren, Jr., (2005), "Residue Number System Arithmetic Circuits With Built-in Self Test". United States Patent 6886123 B1.

[7] http://homepage.cs.uiowa.edu /~jones/bcd/mod.shtml, last accessed on October, 2015.

[8] https://en.wikipedia.org/wiki/Mersenne_prime, last accessed on October, 2015.

[9] Y. Wang.(1998) "New Chinese Remainder Theorems". In proceedings of the Thirty Second Asilomar Conference on Signals, Systems and Computers, Pages: 165-171.

[10] Vidhyalakshmi.M,Prof.Satyabama . (2014) “Design and Implementation of Efficient Binary to Residue Converter Using Moduli Method”. International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Special Issue 1, March 2014.

[11] Shende, Radha, and Zode.(2012) "Efficient design 2k− 1 binary to residue converter." On proceedings of International Conference IEEE, Devices, Circuits and Systems (ICDCS), Pages: 482 – 485.

[12] Omondi, Amos, and Premkuar.(2007) “Residue number systems: theory and implementation”. Imperial College Press.

[13] Mohan,and Ananda(2002) "Residue number systems: algorithms and architectures". Springer, vol.1, pages:1-268.

[14] Jameii, Mahdi, Taghipour, and Azad (2011) “Using both Binary and Residue Representations for Achieving Fast Converters in RNS”.Journal of advances in computer research, Pages: 91-104.

[15] Hiasat, and Sweidan(2003) "Residue number system to binary converter for the moduli set (2n– 1, 2n, 2 n+ 1)". Journal of systems architecture,Vol.49, Pages: 53-58.

[16] Bajard and Plantard(2004) "RNS bases and conversions". Proceedings of Advanced Signal Processing Algorithms, Architectures, and Implementations XIV SPIE, Vol. 5559,Page:61-75.

[17] Ricardo Chaves, and Leonel Sousa, (2004), "{2n + 1, 2n+k, 2n − 1} : A New RNS Moduli Set Extension". Euromicro Symposium Conference: Digital System Design.

[18] Taleshmekaeil, D.K., and Mousavi, A. (2010), "The use of Residue Number System for improving the Digital Image Processing". IEEE 10th International Conference on Signal Processing (ICSP), pages 775-780.

[19] Neha Singh, (2008), "An overview of Residue Number System". National Seminar on Devices, Circuits & Communication.

AUTHORS

Mansour Bader holds a MSc in computer engineering and networks, University of Jordan, Jordan, 2016. BSc Computer Engineering, Al-Balqa Applied University, Jordan, 2008. He is a technical support engineer of computer networks at computer center of Al-Balqa Applied University for 8 years and a half.

Dr. Andraws I. Swidan was born in Al-Karak Jordan in 1954. He received his diploma in Computer Engineering (with honours) and Ph.D. in Computer Engineering from LETI Ulianov Lenin, Sanct Peterburg (Leningrad), Russia in 1979 and 1982 respectively. He Joined the Electrical Engineering Department at the University of Jordan in 1983 and was one of the founders of the Computer Engineering Department at the University of Jordan in 1999. Since then he is a professor at the department. He is also an Adjunct Professor with the ECE department of the McGill University, Montreal, Canada. He holds several technical certifications among which the CISSP. He is an IEEE member, Jordanian Engineering Association member Quebec College of engineers member. He is a Canada Professional Engineer (The province of Quebec). He was member of several national and international scientific committees. He holds several patents and tens of publications.
His main areas of research interest are: computer arithmetic, computer security, encryption algorithms.

Mazin Al-hadidi PhD. in Engineering Science (Computing Machines, Systems and Networks), Institute of Problems of Simulation in Power Engineering Academy of Science, Ukraine/Kiev .1990-1994, with grade Excellent.Bachelor and Master Degrees in Engineering (Computer and intellectual systems and networks) Kiev Institute of Civil Aviation Engineers, as a governmental scholarship, Soviet Union / Kiev, 1984-1990, with grade very good.General Secondary 12 Years Certificate in the Science branch, Jordan/Al-Salt, 1984, with grade very good.

Baha Rababah has a MSc of computer network administration and management (with Distinction), University of Portsmouth, UK, 2015. BSc Computer Engineering, Al-Balqa Applied University, Jordan, 2010. He is a lecturer of computer networks subjects in Keys Training and Solution, Irbid, Jordan.

INTENTIONAL BLANK

ALBAY EMERGENCY RESPONSE AND REPORT TOOL (ALERRT)

Elmer Figuracion[1], Thelma Palaoag[1], Dennis Ignacio[1] and Mary Jane Doblon[1]

[1]College of Information Technology and Computer Science, University of the Cordilleras, Baguio City, Philippines, 2600
ecfiguracion@yahoo.com
tpalaoag@gmail.com
denzignacio@yahoo.com
jhainedoblon@yahoo.com

ABSTRACT

Resilient public alert and warning tools are essential to save lives and protect property during times of national, regional, and local emergencies. Nowadays, immediate emergency alerts became one of the priority in both national and local government. The Provincial Government of Albay is geared towards becoming the most liveable province of the Philippines, which means that it would be known for good education, good healthcare and good environment where people are healthy, happy, employed and lives to their full potential. To achieve this goal, disaster risk reduction and climate change adaptation must be anchored well so as to move to its destination of shared socioeconomic advancement. Supporting this vision, this study focuses on the design and development of a mobile based Albay Emergency Reporting and Response Tool (ALERRT). It is a mobile based resilient form of emergency alert notification that aids the concerned citizens of any emergencies, accidents and concerns that require immediate response from the government sector concerned.

In order to materialize this project, Featured-driven (FDD) methodology was used. Likewise, opinions from different strata of society were solicited along the areas on social awareness, readiness to respond and willingness to use a mobile application to report an incident or emergencies. With 92.5% of the respondent who is willing to report an incidents or emergencies, this application paved way to a better agency response and levering the people to use IT solutions to become resilient in times of emergencies. It can be used to report on emergencies ranging from fires to typhoon-related incidents, vehicular accidents with casualties, health-related concerns (e.g. unidentified person suffering from heart attack), community related incidents & concerns, and any other occurrences which requires immediate and concrete response from the concerned agencies. Having such resilient form of emergency alert notification like ALERRT is deemed necessary for a disaster prone places like Albay.

KEYWORDS

Emergency response, Disaster management, Incident report, Incident mapping, Mobile Application

1. INTRODUCTION

All throughout the year, there could never be a chance that we could be spared from any disasters or occurrences like floods and storms which are the most frequently occurring hazards. Aside

Natarajan Meghanathan et al. (Eds) : NETCOM, NCS, WiMoNe, GRAPH-HOC, SPM, CSEIT - 2016
pp. 161– 177, 2016. © CS & IT-CSCP 2016 DOI : 10.5121/csit.2016.61514

from the mentioned natural calamities, the country also experiences human-induced disasters brought about or influenced by political or socio-economic factors, among others. Violence, traffic hazards, road accidents, broken pipes or electric wires for instance has caused public anxiety, lost of lives, destruction of properties, and living discomforts.

This mobile application- **Albay Emergency Reporting and Response Tool (ALERRT)** is a tool that seeks to encourage the people to become proactive members of the community by increasing their awareness thereby improving resilience and decreasing vulnerabilities. This will provide the citizens to have an easy means of reporting any incidents (emergencies, accidents or concerns) requiring response from any local or national government units, allow citizens to have detailed documentation of the event (image/video capture), allow concerned government sector to act based on reported scenario, and citizens can track down government actions (i.e. action taken on the reported vehicular accident). Furthermore, the system can resolve several issues which include but not limited to slow response of concerned government agencies, poor participation and responsiveness from the community, and unresolved cases of incidents.

1.1. Survey Questionnaires

In order to gather enough data for the system, opinions from different strata of society were solicited through a questionnaire. The respondents include 10 High School students, 10 College students, 10 Professionals, and 10 members of the households whose age bracket is reflected in Table 1; they were randomly chosen by the respondents. The survey questionnaire includes questions on their social awareness, readiness to respond, and also their willingness to use a mobile application in reporting incidents.

Table 1. Respondent's Demographics

Age Bracket	**High School**	**College**	**Professional**	**Household**
16 below	10			
17- 24 yrs.old		10	4	
25- 34 yrs.old			5	3
35-44 yrs.old				5
45-54 yrs.old				2
55 yrs.old above			1	

The result of the survey done through questionnaire revealed that 37 respondents or 92.5% are willing to respond to incidents. See table 2 below.

Table 2. Willingness to Report Emergency Situations

Educational Attainment/Status	**Yes**	**No**
High School	10	0
College	9	1
Professional	10	0
Household	8	2
TOTAL	**37**	**3**

Social awareness was also asked to the respondents, this includes their knowledge of the location of police stations, clinics and hospitals, and barangay halls within their vicinity. Results show that 82.5% of the respondents know the location of police stations, 95% in hospitals and clinics, and 90% in barangay halls nearby.

However, most of the respondents are not aware of the emergency hotlines of police, hospitals, fire dept, etc. in their area. Table 3 reveals the result.

Table 3. Awareness of Emergency Hotlines

Educational Attainment/Status	**Yes**	**No**
High School	1	9
College	0	10
Professional	2	8
Household	3	7
TOTAL	**6**	**34**

The survey likewise revealed that majority of the respondents are smart phone users, and almost all are interested to use a mobile application in reporting an incident; and when asked to rank the features that they would like to see in the application, 57.5% would like an app that would send a text message to report the incident (see table 4 below).

Table 4. Features of the mobile app to be made available

Mobile App Feature	**High School**	**College**	**Professional**	**Household**	**Total**	**Rank**
Text report of incident	5	8	6	4	23	1
Documentation of event(image/video)	4	5	6	2	17	2
Allow concerned sector to act based on the report scenario	1	3	3	1	8	4
Allow viewing of feedback, response, and action taken for the incident	2	5	4	3	14	3

1.2. Interview

An interview was conducted to a personnel of Albay Health Emergency Management (AHEM) and Bureau of Fire Protection (BFP) so as to be acquainted on how emergency responses are done by their agencies.

An interview was conducted to the following personnel of two agencies:

1. **Mr. Eduard E. Gandul Jr**.- a registered Nurse and midwife, certified medical responder for almost 5 years, and an emergency medical technician. He is currently connected with Bicol Regional Training and Teaching Hospital (BRTTH-HEMS)/ Albay Health Emergency Management (AHEM).

2. **Gelacio Molato Jr., BFP** - assigned at Bureau of Fire Protection Tabaco City.

Questions include the procedure when receiving reports of incidents (fire or any emergency-related incidents), emergency measures, and their suggestions to the study being conducted. According to them, they follow a certain protocol in reporting and receiving incident reports from various sources; however, the very common source is through telephone calls. In receiving reports, they follow the “NOIPOITOI” format which stands for: Name of Incident-Place of Incident-Time of Incident. The communication officer receives the information reported by the caller and transmits it to the rescuers/ agencies concerned.

The data collected from the interview aided the researchers in conceptualizing the format of the report needed so as to post in the wall of the various agencies linked to the application.

2. REVIEW OF RELATED LITERATURE

The “New Media” has been in many ways assisted the dissemination of any news, information and data to the people. Whether it is about, health, disaster or emergency situation, it has in many ways made the life of the people accessible, easy, and updated. The following studies present the role of social media and or wireless reporting system in public health, disaster management, and crime management.

Dong (2015) in his capstone study on *“Social Media in Public Health Organizations: A Case Study of Social Media Use in the Minnesota Department of Health”* mentioned that mobile communication or Smart phone penetration brings a new range of possibilities for public health promotion, as the demographic is getting more and more comfortable with mobile news feeding. He further mentioned that the MDH (Minnesota Department of Health) mainly uses Twitter to disseminate news and update about the agency’s work and events; uses Facebook to personalize the organization by employing images and storytelling in content; and uses YouTube for public education and to support public health campaign work.[1] Related to Dong’s study is Owen, Daniel's (2013) thesis entitled *“Citizen Photojournalism: Motivations for Photographing a Natural Disaster and Sharing the Photos on the Web”* shared that the citizen photojournalists inform an audience with their pictures and then they interact with a community with their pictures, they would like to inform the audience about the event of any disaster. Informing an audience coincides with sharing the photos online because it is the easiest way to publish their photos to inform people. Social media Web sites such as Flickr, Facebook and Twitter are hosts for the online communities in which the citizen photojournalists belong.[2]

Barbeau (2007) in his paper “*Wireless Emergency Reporting System*” presented a method of providing emergency related information to and from a centralized location over a wireless network. The method utilizes cellular phones in emergency communications and entails two embodiments that employ location aware technologies, in portable form, in security applications. One embodiment serves as a modern high-tech “neighborhood watch,” enabling law enforcement access to the many “eyes and ears” of the public simultaneously via available cell phones. Cell phones with embedded digital cameras allow the instant capture and remote submission of suspicious circumstances to law enforcement through pictures or video.[3]

Mobile Technology has aided the reporting of various incidents which aims to resolve the sloppy communication and action of the agencies concerned. The following studies ranging from disaster response, health related issues, crime related incident, and others, cited several functionalities that will prove the beneficiality of a reporting and response tool.

Fajardo, et.al (2010) developed the "*A Mobile Disaster Management System Using the Android Technology*" or simple called MyDisasterDroid to determine the optimum route along different geographical locations that the volunteers and rescuers need to take in order to serve the most number of people and provide maximum coverage of the area in the shortest possible time.[4]

Tamboli, et. al (2013) in his study "*Incident Reporting System Using GIS*" presented a paper about integrated application-software which will be used to report an incident or accident immediately and also keep the log of activities which in turn helping public and the authorities to deal with problems and emergencies. Among the applications of the system include a notification of any incident to concerning department, thus the concerned authorities can respond quickly and in an efficient way to solve the problem; it is also used to navigate the response team in minimum possible time; as it keeps the log of activities so it can also be used to maintain log of incidents for further investigation, and this application will help to handle any kind of incident which requires help in a hassle free manner and will also analyze the incident to be reduced. In summary, this application will provide a communication medium for the public to indicate to the respective authorities about the emergencies or incidents identified. This is a very useful real-time application for time-critical incidents.[5]

Eguchi, R. (2008) is his prepared paper entitled "*The Application of Remote Sensing Technologies for Disaster Management*" focused on the integration of remote sensing technologies in all aspects of disaster management, i.e., disaster preparedness, mitigation, response and recovery. In order to demonstrate their efficacy in these four areas, cases histories and examples from recent disasters, including the Marmara, Turkey earthquake, the Bam, Iran earthquake, and the Indian Ocean earthquake and tsunami are discussed. Finally, the paper ends with a view towards the future. What new developments can be expected in technology development and implementation, and what future challenges must be overcome to realize broader application of these technologies in future disasters.[6]

A case study presented by Gupta, Preeti, et.al (2011) dealt on "*Disaster Management in Flash Floods in Leh*" was based on the authors' own experience of managing a natural disaster caused by the flash floods. The paper presents a firsthand description of a disaster and its prompt management. The data was collected from the records of the district civil administration, the civil hospital, and the Army Hospital, Leh. The approach used was both quantitative as well as qualitative. It included data collection from the primary sources of the district collectorate, interviews with the district civil administration and army officials who organized rescue operations, restoration of communication, and transport, mass casualty management, and informal discussions with local residents. In here, the researchers emphasized the importance of readiness not just in the health sector, disaster management sector, but also with the communication and transmission aspect.[7]

K Nakajima, et. al (2016) developed "*A Web-Based Incident Reporting System and Multidisciplinary Collaborative Projects for Patient Safety in a Japanese Hospital*" which is currently recognized as a useful tool for patient safety in individual hospitals as well as at the national level. The use of a computerized system in limited settings such as intensive care units succeeded in involving physicians in reporting to a greater extent, where they accounted for more than 20% of the total number of reports. Reported incidents have led professional groups to take action, including making recommendations for restrictions on storage areas for high risk drugs and the establishment of a Department of Clinical Engineering for the centralized management of medical devices; as well as reporting incidents and a faster response to problems brought to light as a result of such reporting—which were the barriers that we faced in introducing the patient safety programs, can be resolved by the web-based incident reporting system which streamlines the process of reporting and information sharing.[8]

In the study entitled "*Use of Mobile Phones in an Emergency Reporting System for Infectious Disease Surveillance*" conducted by Yang, C. et.al (2010), the result indicates that the mobile phone reporting system helped restore the reporting capacity of health-care agencies in earthquake-affected areas. The drop in the number of cases reported might have been caused by two factors: the rate of unreported cases increased because doctors were flooded with patients after the earthquake; and the occurrence of infectious disease in some areas was possibly lower than in past years due to the stringent disease prevention. Last, whenever possible, mobile phones with global positioning system (GPS) capacity should be used. The reporting system can be programmed to attach coordinate data to each text message automatically. This could help us track the disease in a spatial resolution higher than the township level. And control measures adopted by Chinese authorities after the earthquake.[9]

Yunus (2006) in his thesis "*Web Based Multi-Participant Spatial Data Entry in Crime Mapping*" designed a crime mapping system in order to produce data along with a web based data entry methodology. The emphasis of the study is to convey some new improvements for effective and accurate geocoding of point based crime incidents, offenders and victim's data. The study establishes a server side Web architecture that provides map visualization.In summary, this study can be a prototype for online crime mapping in tactical crime analysis.[10]

3. THE SYSTEM

3.1. Conceptual Model

To fully understand the main functions of the system, a use case diagram was used to illustrate the main activities of the stakeholders as shown in Figure 1.

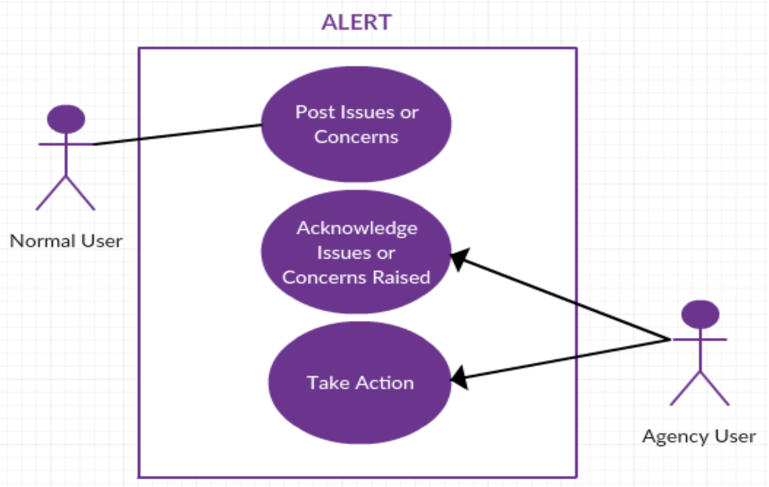


Figure 1. ALERRT System Overview

In general, the ALERRT system involves two important users as shown in figure 1: 1. Normal User 2. Agency User. The normal user is the one involved in raising issues and concerns that requires the agency's attention. Once a certain issues/concerns are raised, then it is the responsibility of the agency user to acknowledge the post and take proper action. In summary, the objective is to let the preferred agency know the issues/concerns so that proper action can be carried over by the concerned agency.

3.2. Database Design

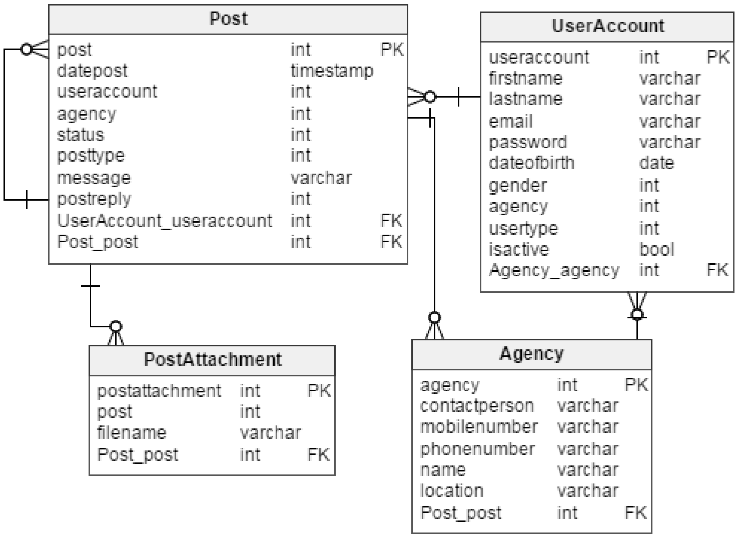


Figure 2. ER Diagram of ALERRT

There are four tables used in the ALERRT system as depicted in Figure 2. Below is the summary.

1. UserAccount - holds the records for ALERRT users. User's could be of the following types: a. normal user b. agency user c. admin user. A Normal user is limited in posting issues/concerns,; agency users are entitled to respond and take action to any issues/concerns raised by the normal user; and lastly the admin user is capable of maintaining the agencies and user accounts. The UserAccount table is linked to Agency table if the specific user is a representative of an agency.

2. Agency - holds records of registered agencies within the ALERRT system. Information in this table includes the contact person, mobile and phone number and the name of the agency and its location.

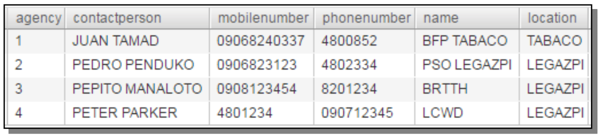
3. Posts - where all issues/concerns raised by user's or any public announcements by agencies are recorded. Posts are linked to the user who created the post, but is optionally connected to an agency. Only those posts which require the agency's attention are connected to agency. These kinds of posts are maintained in the status information for monitoring purposes (e.g. to track down what happens to the issues/concerns raised and how the agency respond to it). Lastly, each post is linked to itself; this signifies the replies made to a certain post.

4. PostAttachment - A post can optionally have picture attachments. This table handles the physical location of the image on the server for a certain post.

3.3. Data Instances

Succeeding figures below show sample data instances for ALERRT tables.

A. Agency



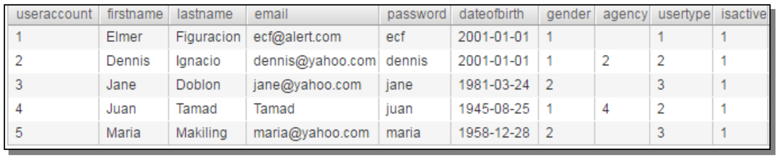
agency	contactperson	mobilenumber	phonenumber	name	location
1	JUAN TAMAD	09068240337	4800852	BFP TABACO	TABACO
2	PEDRO PENDUKO	0906823123	4802334	PSO LEGAZPI	LEGAZPI
3	PEPITO MANALOTO	0908123454	8201234	BRTTH	LEGAZPI
4	PETER PARKER	4801234	090712345	LCWD	LEGAZPI

Figure 3.1 Agency's Sample Data

The AGENCY table holds the basic information intended for agencies as depicted on Figure 3.1. Below is the summary of the fields used:

Fields	Description
Agency	Primary Key
ContactPerson	The point person for the agency
MobileNumber	Agency's mobile number
PhoneNumber	Agency's phone number
Name	Name of the agency
Location	City Location of the Agency

B. UserAccount



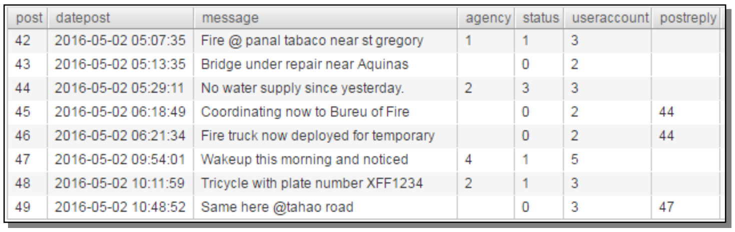
useraccount	firstname	lastname	email	password	dateofbirth	gender	agency	usertype	isactive
1	Elmer	Figuracion	ecf@alert.com	ecf	2001-01-01	1		1	1
2	Dennis	Ignacio	dennis@yahoo.com	dennis	2001-01-01	1	2	2	1
3	Jane	Doblon	jane@yahoo.com	jane	1981-03-24	2		3	1
4	Juan	Tamad	Tamad	juan	1945-08-25	1	4	2	1
5	Maria	Makiling	maria@yahoo.com	maria	1958-12-28	2		3	1

Figure 3.2 UserAccount' Sample Data

The USERACCOUNT table holds the record of all the ALERRT's users. Below is the summarized field:

Fields	Description
Firstname	User's firstname
Lastname	User's lastname
Email	Email Address used for login to ALERT
Password	Password used for login to ALERT
DateOfBirth	User's date of birth
Gender	Gender 1. Male 2. Female
Agency	Assigned with Agency PK when user is linked to an agency, otherwise it can be NULL
UserType	Type of user 1. Admin User 2. Agency User 3. Normal User
IsActive	Signifies if user is active or not

C. Post



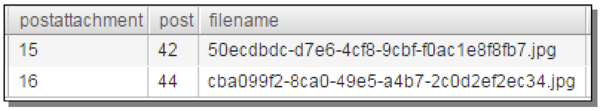
post	datepost	message	agency	status	useraccount	postreply
42	2016-05-02 05:07:35	Fire @ panal tabaco near st gregory	1	1	3	
43	2016-05-02 05:13:35	Bridge under repair near Aquinas		0	2	
44	2016-05-02 05:29:11	No water supply since yesterday.	2	3	3	
45	2016-05-02 06:18:49	Coordinating now to Bureu of Fire		0	2	44
46	2016-05-02 06:21:34	Fire truck now deployed for temporary		0	2	44
47	2016-05-02 09:54:01	Wakeup this morning and noticed	4	1	5	
48	2016-05-02 10:11:59	Tricycle with plate number XFF1234	2	1	3	
49	2016-05-02 10:48:52	Same here @tahao road		0	3	47

Figure 3.3 POST's Sample Data

POSTS table records any issues/concerns raised and its replies by ALERRT's users. Below is the Summarized definition of the POST fields.

Fields	Description
Post	Primary Key
DatePost	Date/Time of post
Message	Text Description of the post
Agency	Intended agency. Linked to Agency table
Status	Post Status which has the following values 1. Public Information (when post is for general public such as announcements, advisory, etc) 2. Reported(means the post is intended to a specific agency and its now flag as reported) 3. Acknowledge (when post is received by agency and raised issues/concerned is now being examined) 4. Action Taken (when posts is already taking care of)
UserAccount	The user who created the post
PostReply	A reply to an existing post

D. PostAttachment



postattachment	post	filename
15	42	50ecdbdc-d7e6-4cf8-9cbf-f0ac1e8f8fb7.jpg
16	44	cba099f2-8ca0-49e5-a4b7-2c0d2ef2ec34.jpg

Figure 4. POSTATTACHMENT's Sample Data

The POSTATTACHMENT table records all the images attached to a post. Below is the field definition:

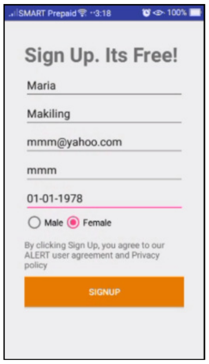
Fields	Description
PostAttachment	Primary Key
Post	Foreign Key to POST table
Filename	File location of the image

3.4. Sample Screenshots

The ALERRT comprises three main users: normal user, agency user and an admin user. Each has its own purpose in maintaining the flow of information that goes within the ALERRT system. Below summarizes each user and their functions in the system.

A. Normal User

1. Sign Up



Sign Up. Its Free!

Maria

Makiling

mmm@yahoo.com

mmm

01-01-1978

○ Male ● Female

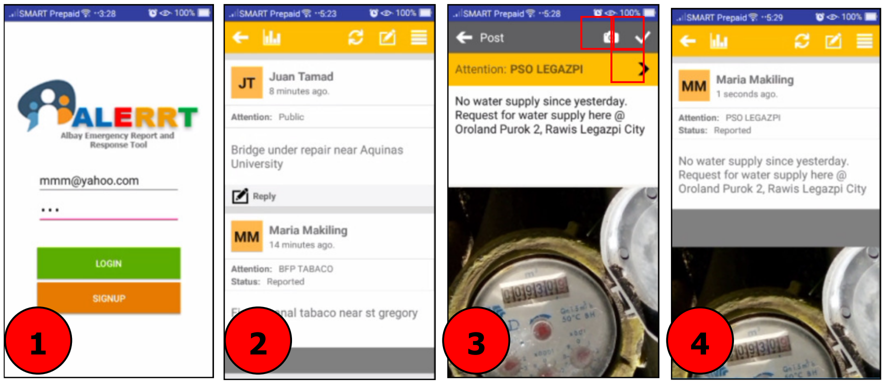
By clicking Sign Up, you agree to our ALERT user agreement and Privacy policy

Figure 5. Sign Up

Signing up to ALERRT is the first step required to be involved in the system. Figure 5 shows the required information when signing up. The most important piece of information here is the email address. This is unique from the entire system and user won't be able to register if email is already in use.

2. Post Issues/Concerns

To post an issues or a concern, users need to be logged in first (1). Next is to press the write post button on top level highlighted on (2). This will take the user to (3) a screen. A post is by default intended for public information, but if the post requires agency attention, then pressing the > besides "Attention: <Agency Name>" which will be taken to another screen where user can choose an agency. To attach an image to a post, the camera button lets the user choose an existing image from the phone. User can then describe the issues/concerns by typing some text and hitting the tick mark button to finally commit the post. This will lead to (4) where post written is available on the public wall.



mmm@yahoo.com

•••

JT Juan Tamad
8 minutes ago.

Attention: Public

Bridge under repair near Aquinas University

MM Maria Makiling
14 minutes ago.

Attention: BFP TABACO
Status: Reported

Fi nal tabaco near st gregory

Attention: PSO LEGAZPI

No water supply since yesterday. Request for water supply here @ Oroland Purok 2, Rawis Legazpi City

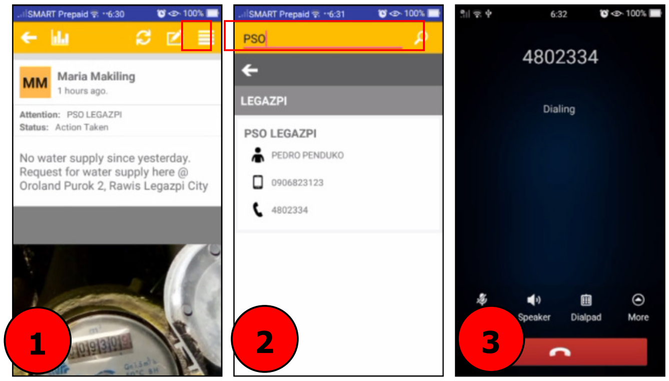
MM Maria Makiling
1 seconds ago.

Attention: PSO LEGAZPI
Status: Reported

No water supply since yesterday. Request for water supply here @ Oroland Purok 2, Rawis Legazpi City

Figure 6. Post Issues and Concerns

3. Call Agency


1

Maria Makiling
1 hours ago.
Attention: PSO LEGAZPI
Status: Action Taken
No water supply since yesterday. Request for water supply here @ Oroland Purok 2, Rawis Legazpi City

2

PSO
LEGAZPI
PSO LEGAZPI
PEDRO PENDUKO
0906823123
4802334

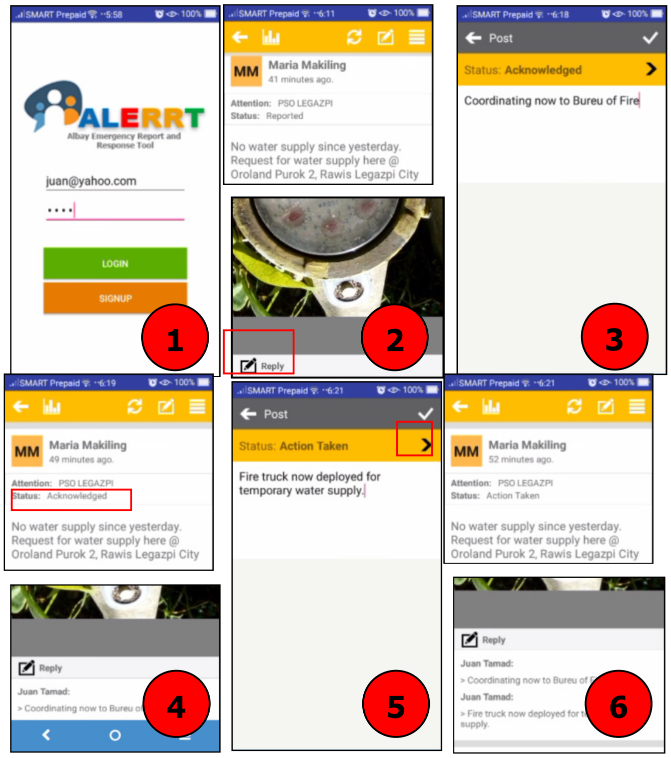
3

4802334
Dialing

Figure 7. Contact Agency

ALERRT manages lists of registered agencies together with their respective contact numbers. This gives the capacity for all ALERRT users to contact the agency's representative directly. To invoke a call, from step (1) on the figure, tap the list icon. This will take every user to a screen showing all the registered agencies. Next is to search for the target agency as highlighted in step (2). To initiate a call, tap the mobile/phone icon which will open up (3) screen showing an ongoing call.

B. Agency User


1

ALERRT
Albay Emergency Report and Response Tool
juan@yahoo.com
••••
LOGIN
SIGNUP

2

Maria Makiling
41 minutes ago.
Attention: PSO LEGAZPI
Status: Reported
No water supply since yesterday. Request for water supply here @ Oroland Purok 2, Rawis Legazpi City
Reply

3

Post
Status: Acknowledged
Coordinating now to Bureu of Fire

4

Maria Makiling
49 minutes ago.
Attention: PSO LEGAZPI
Status: Acknowledged
No water supply since yesterday. Request for water supply here @ Oroland Purok 2, Rawis Legazpi City
Reply
Juan Tamad:
> Coordinating now to Bureu o

5

Post
Status: Action Taken
Fire truck now deployed for temporary water supply.

6

Maria Makiling
52 minutes ago.
Attention: PSO LEGAZPI
Status: Action Taken
No water supply since yesterday. Request for water supply here @ Oroland Purok 2, Rawis Legazpi City
Reply
Juan Tamad:
> Coordinating now to Bureu of F
Juan Tamad:
> Fire truck now deployed for t supply.

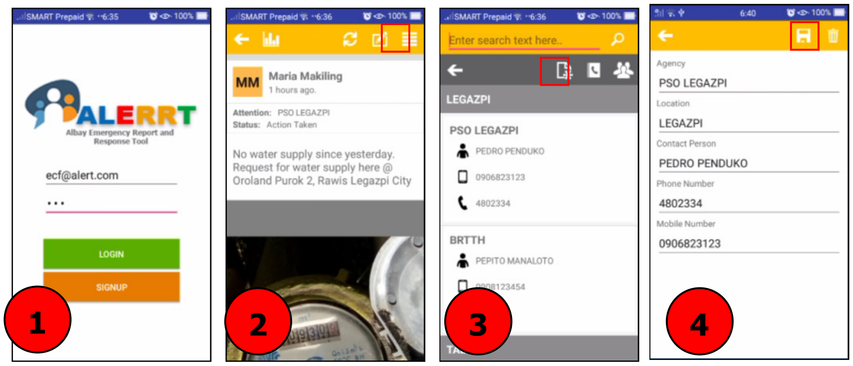
Figure 8. Agency User Respond to Post

The Agency user's main purpose on ALERRT system is to track down issues/concerns raised against its agency and to take action based from post. To do this, agency user needs to be logged as depicted on (1). Next is to find the post that requires attention (2) and tapping the reply button. On (3) screen, the > button on the right side of "Status:<Status Name>" can be set to "Acknowledged". Preferably, this is the status that needs to be set by the agency once a post is received. This indicates that the post is being taken care of and the issues/concerns raised are now being examined. Screen (4) shows that the reply is reflected on the original post and the post status is now changed to "Acknowledged". Lastly, once the post is resolved, the agency user on (5) can now set the status to "Action Taken" including the text summary of the action being done. This is then reflected on the original post as show on screen (6).

C. Admin User

The Admin user maintains the list of registered agencies within the system as well as assigns a registered user to an agency. The screens below show the details of these functions.

1. Register a new Agency



1

ecf@alert.com

•••

2

Maria Makiling
1 hours ago.

Attention: PSO LEGAZPI
Status: Action Taken

No water supply since yesterday. Request for water supply here @ Oroland Purok 2, Rawis Legazpi City

3

Enter search text here..

LEGAZPI

PSO LEGAZPI
PEDRO PENDUKO
0906823123
4802334

BRTTH
PEPITO MANALOTO
0908123454

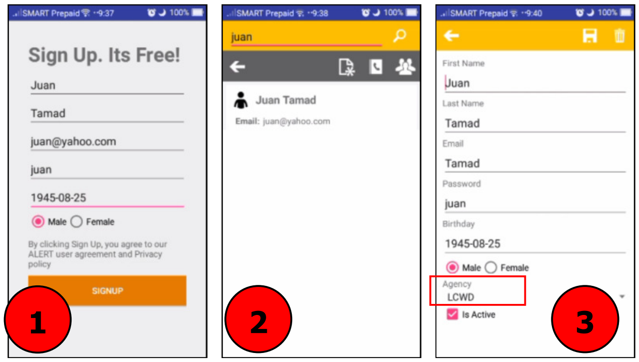
4

Agency: PSO LEGAZPI
Location: LEGAZPI
Contact Person: PEDRO PENDUKO
Phone Number: 4802334
Mobile Number: 0906823123

Figure 9. Register a new agency

After being logged in as shown on step 1, an admin user can then tap the list button (highlighted on step 2) to access the list of agencies and users. By default the list of agencies are listed. Now to register a new agency, tap the new button on step 2. This will open up the step 4 asking for agency information. Finally, by tapping the save button on step 4 it commits the new agency record.

2. Assign user as agency representative



1

Sign Up. Its Free!

Juan
Tamad
juan@yahoo.com
juan
1945-08-25
● Male ○ Female

By clicking Sign Up, you agree to our ALERT user agreement and Privacy policy

2

juan

Juan Tamad
Email: juan@yahoo.com

3

First Name: Juan
Last Name: Tamad
Email: Tamad
Password: juan
Birthday: 1945-08-25
● Male ○ Female
Agency: LCWD
☑ Is Active

Figure 10. Assign user as agency representative

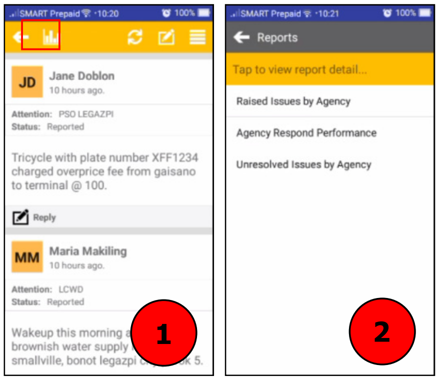
Every agency needs to have a representative, which is called the agency user. This user needs to register first as normal user (1); and to become an agency user, he/she needs to contact the admin user to flag its account as agency representative. Step 2 shows the screen of admin user where it filters a certain user. Tapping the target user will show screen 3, where user can then assign the preferred agency.

4. DATA OPTIMIZATION

The key concept of this study is about data optimization- that is getting information from a post and carries a necessary action to resolve the issues. Data optimization is an important aspect in database management in particular and in data warehouse management in general. It is most commonly known to be a non-specific technique used by several applications in fetching data from a data sources so that the data could use in data view tools and applications such as those used in statistical reporting.[11]

Any information posted by the user is significant in understanding the most common issues and concerns raised by a user, so as to aid the agencies for better understanding on what steps to be taken- to limit, if not to eradicate such concerns. Information carried on the post is of advantage to the agency to understand future concerns and to respond to it smoothly, the next time it happens again. The data collected from the post can be of great use in establishing connection to the constituents, raise awareness of event happening on the environment, and objectively encourage citizens to become proactive in common issues around them. The ALERRT system encourages not only the agency to become a good responder but also raises awareness to its users to become concerned citizens who are willing to assist when needed.

To aid the analysis of data, the ALERRT has three major reports that provide the overview of the data collected. As shown in Figure 11 below, the report can be accessed by clicking the chart icon on screen 1.



JD Jane Doblon
10 hours ago.
Attention: PSO LEGAZPI
Status: Reported

Tricycle with plate number XFF1234 charged overprice fee from gaisano to terminal @ 100.

MM Maria Makiling
10 hours ago.
Attention: LCWD
Status: Reported

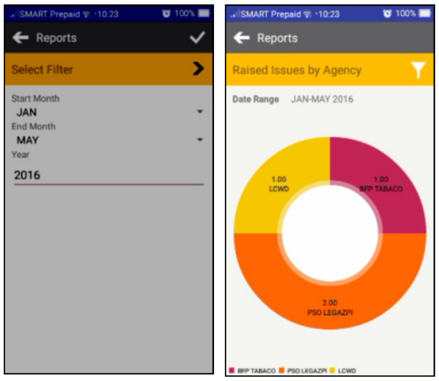
Wakeup this morning a brownish water supply smallville, bonot legazpi k 5.

Tap to view report detail...

- Raised Issues by Agency
- Agency Respond Performance
- Unresolved Issues by Agency

Figure 11. ALERT's reports

1. Raised Issues by Agency



Reports

Select Filter

Start Month
JAN

End Month
MAY

Year
2016

Reports

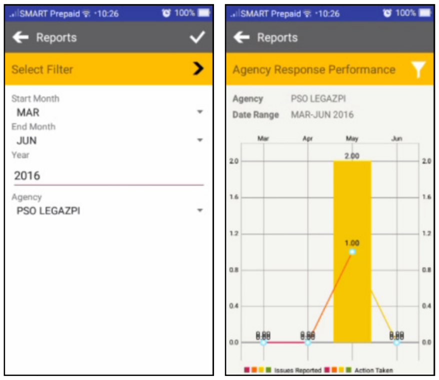
Raised Issues by Agency

Date Range JAN-MAY 2016

Figure 12. Report - Raised Issues by Agency

Raised Issues by Agency shows the total summary of issues categorized by agencies. This can be filtered out by start month to end month and year. This provides an statistics on which agencies are getting/receiving the most number of issues/concerns.

2. Agency Respond Performance



Reports

Select Filter

Start Month
MAR

End Month
JUN

Year
2016

Agency
PSO LEGAZPI

Reports

Agency Response Performance

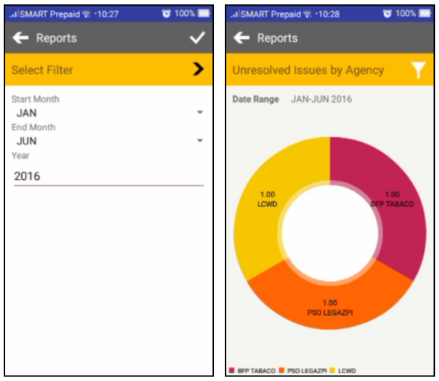
Agency PSO LEGAZPI

Date Range MAR-JUN 2016

Figure 13. Report - Agency Respond Performance

This report returns the agency’s performance in responding to posts. The report is filtered by start month-to-end month, year and by agency. It reflects the total number of issues/concerns raised represented by bar chart and the total resolved posts represented by a line graph. This gives an idea on how the agency performed in dealing with problems. The key here is, it reflects the overall performance of an agency in relation to giving resolution to posts initiated by users.

3. Unresolved Issues By Agency



Start Month
JAN
End Month
JUN
Year
2016

Date Range JAN-JUN 2016

Figure 13. Report - Unresolved issues by Agency

This report shows the summary of unresolved issues by agencies. The data can be filtered by start month-to-end month and by year. Basically, when an agency is listed here with higher total, this means that the agency is not performing well and can be assumed that it failed to resolve any issues raised to them.

These features of the application found value on the data gathered from the respondents and use it to generate certain statistical data with the purpose of evaluating the job performance of agencies. By analyzing the data gathered, an evaluation or assessment tool has been made up and a forecast of possible occurrences can be extracted for future planning.

5. CONCLUSIONS

Bridging communication between the constituent and the concerned agency is the main goal of the study. It seeks to provide ways on how a certain concern can be brought forward directly to intended agencies. This paved way to a better agency response; a better understanding of common concerns raised; and birth of an active, concerned, and vigilant community.

There are few items that the authors wish to incorporate in this study that can be used by future researchers in order to improve this research. These are:

1. Categorized Post

There should be a way in classifying or categorizing where a certain post belongs. This will give detailed statistics on the data that can be analyzed from ALERRT system which can be used by the agency in implementing proper planning in responding to such concerns.

2. ALERRT Web

This is a web version of ALERRT to reach other users who do not have mobile devices and also to provide multiple ways on interacting to ALERRT system.

Lastly, below are the several researches that the authors wish to accomplish in the future. These are systems that can be connected to ALERRT to provide more information and public service to its users.

1. Flood Control System

This is using an arduino device that automatically feeds data to ALERRT system to notify which bridges/highways/rivers are flooded.

2. Fire Control System

Using the ALERRT System, this envisions to automatically communicate with any Fire Department/Stations whenever fire happens to any establishment. Using an arduino device, this will prompt the department of any fire occurrences that need immediate attention.

ACKNOWLEDGEMENTS

This research paper is will not be possible without the help and support of our family.

Thank you also to all the UC-CATC MIT family for this endeavour and most of all to Almighty God for His guidance. To God be all the glory.

REFERENCES

[1] Fajardo, Jovilyn Therese, et.al (2010, June). A Mobile Disaster Management System Using the Android Technology. Retrieved from citeseerx.ist.psu.edu/viewdoc/ doi=10.1.1.458.3601

[2] K Nakajima, et. al. (2016, February). A web-based incident reporting system and Multidisciplinary collaborative projects for patient safety in a Japanese hospital. Downloaded from http://qualitysafety.bmj.com/.

[3] Asif S. Tamboli, et.al (2013, February). Incident Reporting System Using GIS. Retrieved from www.estij.org/papers/vol3no12013/10vol3no1.pdf.

[4] Eguchi, R. (2008, October). The Application of Remote Sensing Technologies for Disaster Management. Retrieved from www.iitk.ac.in/nicee/wcee/article/14_K004.pdf.

[5] Yang,C. et. al.(2010, December). Use of Mobile Phones in an Emergency Reporting system for infectious disease surveillance after the Sichuan earthquake in China. Retrieved from https://depts.washington.edu/einet/symposium/PRC031210.pdf.

[6] Barbeau,Sean et.al(2007, February). Wireless Emergency-Reporting System. Retrieved from www.research.usf.edu/absolute-news/templates/?a=2513&z=1.

[7] Gupta, Preeti, et. al (2011, November). Disaster Management in Flash Floods in Leh: A Case Study. Retrieved from www.ifrc.org/Global/Publications/disasters/dref/cs-India.pdf.

[8] Dong, Chuqing (2015, July). Social Media in Public Health Organizations: A Case Study of Social Media Use in the Minnesota Department of Health. Retrieved from http://www.freefullpdf.com/#gsc.tab=0&gsc.q=social%20media&gsc.sort=&gsc.ref

[9] Owen, Daniel (2013, May). Citizen Photojournalism: Motivations for Photographing a Natural Disaster and Sharing the Photos on the Web". Retrieved from https://etd.ohiolink.edu/rws_etd/document/get/akron1362739905/inline

[10] Yunus Emre Aydin (2006, May). Web Based Multi-Participant Spatial Data Entry in Crime Mapping". Retrieved from https://etd.lib.metu.edu.tr/upload/12607250/index.pdf

[11] http://www.merkleinc.com/what-we-do/data-analytics/data-solutions/data-optimization#.Vyox9tJ96t8

AUTHORS

Elmer C. Figuracion is an IT Professional, an Instructor, and a freelance software developer. He finished his Bachelor's degree in Computer Science at Polytechnic Institute of Tabaco and is now taking Master in Information Technology at the University of Cordilleras, Baguio City.

Thelma D. Palaoag is a graduate of Doctor in Information Technology at the University of the Cordilleras. She is currently a faculty of the College of Information Technology and Computer Science of the same university.

Dennis Ignacio is a BSIT graduate of University of Nueva Caceres, Naga City. He is currently an IT Instructor at La Consolacion College of Daet and also a student of University of the Cordilleras taking his Master in Information Technology.

Mary Jane Doblon holds her BSCS degree at the University of Caceres, Naga City. She is an IT Instructor at La Consolacion College of Daet and likewise a student of University of the Cordilleras taking up Master in Information Technology.

INTENTIONAL BLANK

AN EFFICIENT RECOVERY SCHEME FOR BUFFER-BASED B-TREE INDEXES ON FLASH MEMORY

VanPhi Ho, Seung-Joo Jeong, Dong-Joo Park

School of Computer Science and Engineering,
Soongsil University
Seoul, Korea
{hvphi, qkaxhf1007, djpark }@ssu.ac.kr

ABSTRACT

Recently, flash memory has been widely used because of its advantages such as fast access speed, nonvolatile, low power consumption. However, erase-before-write characteristic causes the B-tree implementation on flash memory to be inefficient because it generates many flash operations. To address this problem, variants of buffer-based B-tree index have been proposed for flash memory which can reduce a number of write operations. Since these B-trees use a main-memory resident index buffer to temporarily store newly created index units, their data may be lost if a system crash occurs. This study introduces a novel recovery scheme for the buffer-based B-tree indexes on flash memory, called ERS. ERS can minimize the risk of losing data by deploying logging and recovery policies. The experimental results show that ERS yields a good performance and helps the buffer-based B-tree indexes improve the reliability.

KEYWORDS

B-tree index, flash-aware index, flash memory.

1. INTRODUCTION

Flash memory [1-2] has been widely used because it has many positive features such as high-speed access, low power consumption, small size and high reliability. Besides these advantages, it has some downsides including erase-before-write, limited life cycle. In order to access data on the flash memory as accessing hard disk drives, hosts need to use a firmware module named Flash Translation Layer (FTL) [1]. FTL translates the logical address to physical address between the host and flash memory. The FTL has two main features: address mapping and garbage collection. Many FTL algorithms have been used to confine the limitation of physical characteristics and enhance the performance of flash memory.

By using FTL algorithms, the performance of a flash memory has been improved. However, implementing B-tree index directly on flash memory may not be efficient because the erase-before-write characteristic. Updating B-tree nodes causes the overwrite operations on flash memory occur frequently. To address these problems, variants of B-tree index have been proposed for flash memory. Among these B-tree variants, there are some B-trees using a main-memory resident index buffer (called buffer-based B-tree for short) to temporarily store newly created index units in order to reduce the number of flash operations. However, using the main-memory resident index buffer causes the index data in the buffer may be lost when a sudden power-off occurs.

Natarajan Meghanathan et al. (Eds) : NETCOM, NCS, WiMoNe, GRAPH-HOC, SPM, CSEIT - 2016
pp. 179– 186, 2016. © CS & IT-CSCP 2016 DOI : 10.5121/csit.2016.61515

This paper present a new recovery method for the buffer-based B-tree indexes on Flash Memory, called ERS. Whenever the index units are inserted into the index buffer, the ERS backs up the index units to an area called logger which is located in flash memory to avoid losing its data. If the system reboots after a crash, the backed up index units in the logger are read back to the index buffer. ERS could minimize the loss of data.

The experimental results indicate that our proposed scheme achieves a good performance and it helps the buffer-based B-tree indexes improve the reliability.

The paper is organized as follows: Section 2 reviews background and related works. The design of ERS and its operations are presented in Section 3. Section 4 experimentally evaluates the efficiency of ERS, and finally, Section 5 concludes the paper.

2. BACKGROUND AND RELATED WORKS

Flash memory is a storage device whose data is nonvolatile. It is widely used nowadays because its strong points. Different from a traditional hard disk drive, flash memory is consisted of a number of NAND flash memory arrays, a controller, and an SRAM. NAND flash memory arrays are organized in many blocks. Each block contains a fixed number of pages (e.g. 32, 64). A page is the smallest unit of read and write operations while the block is the smallest unit of erase operations. Similar to the hard disk drive, flash memory supports all basic operations: read, write and erase. The read operation is the fastest one, that is about 10 times faster than a write operation. The erase operation is very time-consuming, which takes about 2ms. The erase operation is over 10 times slower than a write operation. Also, as mentioned above, the main drawback characteristic of NAND flash memory is that it has erase-before-write architecture. Moreover, the life cycle of flash memory is limited. The number of erase cycles for a block is bounded about 100,000 times. Therefore, frequent erasing of some particular locations may deteriorate both the overall performance and lifetime of the flash memory. Since flash memory owns these physical characteristics, it requires an intermediate module called Flash Translation Layer (FTL) for translating the address mapping, managing and controlling data. By using FTL, the general performance of flash memory is improved and quickly deploy disk-based applications without any modifications.

B-tree index [3] is a data structure which is popularly used in many file systems and database management systems because of quickly access capability. However, implementing B-tree directly on flash memory may suffer from degradation of the efficiency of B-tree index as well as the lifecycle of flash memory because of the erase-before-write limitation of flash memory.

To address these problems, variants of B-tree have been proposed for flash memory. Wu et al. presented BFTL [4], the first B-tree variant. BFTL is consist of a node translation table and a reservation buffer. Every newly created index unit which reflects the inserted, deleted or modified records is temporarily stored in the reservation buffer. When the reservation buffer is full, all index units in the buffer are flushed to flash memory in FIFO order by an internal operation of BFTL called commit. Since some index units for the same node may be written in various pages, a node translation table is used. The node translation table collects all index units and maintain the information of the pages having the index units of the same B-tree node. As a result, BFTL reduces the number of flash operations. However, many read operations is needed to access a B-tree node because the index units of one node may be scattered on many different flash pages. Moreover, since the buffer is a volatile storage, its data may be lost when a sudden power-off occurs leading to the stored B-tree in the flash memory may be an unstable structure. And then, it may yield serious problems when managing a number of data.

In order to solve the drawbacks of BFTL, a new index buffer management scheme named IBSF [5] was proposed. The main idea of IBSF is to store all index units associated with a B-tree node onto one page, so IBSF does not need the node translation table. Similar to BFTL, IBSF temporarily stores newly created index units into the index buffer. When flushing records from the index buffer to flash memory, IBSF selects victim index units by identifying the records to be inserted into the same logical node. This prevents them from spreading across several flash pages. Thus, IBSF reduces the search overhead of BFTL. However, due to the fact that there are a lot of index units still in the index buffer of IBSF, its data may be lost when a sudden power-off occurs similarly to BFTL.

Later on, a write-optimized B-tree layer for NAND Flash memory (WOBF) [6] was proposed. Basically, WOBF inherits the advantages of BFTL and IBSF. It employs the index buffer and the node translation table used in BFTL and the commit policy of IBSF. Its performance is improved by sorting all the index units in the index buffer before performing commit operations. Sorting all the index units prevents the index units belonging to the same node from being scattered over many pages. This reduces the number of read operations when building a logical node. Nevertheless, similar to BFTL and IBSF, WOBF still suffers from losing data when a sudden power-off occurs because the index buffer is volatile.

Summary, the above buffer-based B-tree indexes reduce the number of writes when building a B-tree by exploiting the main memory index buffer. However, since the index buffer is a volatile storage, their data may be lost when a sudden power-off occurs leading to the stored B-tree index in the flash memory may be an unstable structure. Therefore, it may yield serious problems when managing a number of data.

3. The Design and Implementation of ERS

3.1. The design of ERS

This section presents a novel recovery scheme for the buffer-based B-tree indexes on flash memory, called ERS, to efficiently restore the data of buffer-based B-trees when a sudden power failure occurs. Its objective is to significantly reduce the risk of losing data of buffer-based B-tree indexes when a B-tree is built. In order to achieve the aforementioned goal, we maintain a flash-memory resident logger which uses some blocks of flash memory to temporarily stores newly created index units whenever the newly created index units are inserted into the index buffer. Figure 1 shows the architecture of ERS comprising an index buffer, buffer-based B-tree policies, a logger and a recovery module.

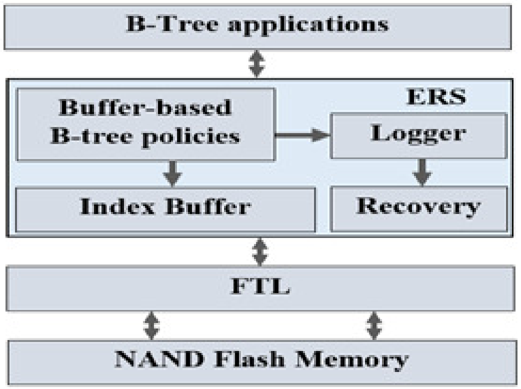


Figure 1. Overall architecture of ERS

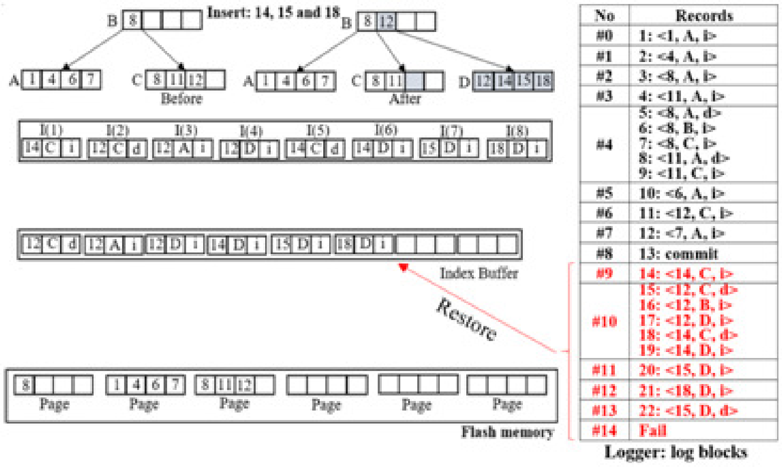
The buffer-based B-tree policies module basically manages the index unit in the index buffer according to the buffer-based B-tree algorithms. The logger is located in flash memory to avoid losing its data. It uses some blocks of flash memory to store all newly generated index units. These blocks are called log blocks. The logger adopts the logging mechanism [7] for the recovery so that it sequentially records the newly created index units whenever a B-tree node is modified. Since writing data sequentially into the logger, the overhead of this writing is relatively small [8-9]. When a commit operation is performed successfully (e.g. all the index units are written onto flash memory), a completion commit sign is set in the logger. The recovery is triggered when the system restarts after a crash. It detects and eliminates incompatibilities by restoring the state of the system just before the crash took place. By using the recovery module, ERS ensures the durability of all the data before the crash.

3.2. The implementation of ERS

3.2.1. Logging policy

For recovering data after the system crash, BMS writes the newly created index units to the logger sequentially. When a record is inserted into B-tree or deleted from B-tree, one or more index units are created to reflect the insertion/deletion. After inserting the newly created index units into the index buffer, ERS backups these index units to the logger simultaneously. The newly created index units are temporarily stored in the index buffer based on the policies of the buffer-based B-tree indexes. Owing to the limitation of the index buffer size, all index units in the index buffer are flushed to the flash memory when the index buffer is full by using the commit operation. When a commit operation is performed successfully, a commit record is created in the logger to denote that all data is written to flash memory. The commit record is a checkpoint that denotes a commit operation is finished successfully. Since the size of an index unit is much smaller than that of a flash page, all the index units related to one record are written to one page of the log blocks. This reduces the number of write operations onto the logger and saves a lot of space of the log blocks.

Figure 2 presents an example of the logging policy. Supposing that a buffer-based B-tree is built by inserting 10 records having key values as 1, 4, 8, 11, 12, 6, 7, 14, 15 and 18.



No	Records
#0	1: <1, A, i>
#1	2: <4, A, i>
#2	3: <8, A, i>
#3	4: <11, A, i>
#4	5: <8, A, d> 6: <8, B, i> 7: <8, C, i> 8: <11, A, d> 9: <11, C, i>
#5	10: <6, A, i>
#6	11: <12, C, i>
#7	12: <7, A, i>
#8	13: commit
#9	14: <14, C, i>
#10	15: <12, C, d> 16: <12, B, i> 17: <12, D, i> 18: <14, C, d> 19: <14, D, i>
#11	20: <15, D, i>
#12	21: <18, D, i>
#13	22: <15, D, d>
#14	Fail

Figure 2. Logging and recovery policy

To build the B-tree, the index units are created to reflect the insertions or deletions and then they are inserted into the index buffer. Simultaneously, these index units are written to the logger (e.g. page #0 to page #7 in the first block of log blocks). In this example, the index buffer is already full when index unit <14, C, i> is created. At this time, all index units in the index buffer are written to flash memory and then a commit record is set in the logger (e.g. stored in page #8 of the first log block). After vacating the index buffer, the newly generating index units are inserted into the index buffer continuously according to the insertion and deletion policies. Additionally, since the size of the logger is limited, the logger eventually is fulfilled. Therefore, some log blocks in the logger should be erased timely to vacate space.

3.2.2. Recovery policy

For reliability and compatibility of buffer-based B-trees, ERS performs the recovery policy when the system is rebooted after the crash. The recovery policy is described as follows: First, ERS finds the last commit record in the logger because all index units are written into flash memory successfully before the commit record is created. Then it redoes operations by restoring all records after the last commit records in the logger to the index buffer. At this point, ERS operates normally by applying insertion, deletion and commit policies.

As shown in figure 2, supposing that the crash occurs after the record having key value 15 is deleted in node D (e.g. record 22 <15, D, d> in the figure). According to the recovery policy, ERS searches the last commit record in the logger. In this example, the last commit is the 13th record (page #8) in the logger. After getting the last commit record, ERS redoes operations by reading log records from 14th to 22th and inserts them into the index buffer. In this example, if there two more index units are inserted into the index buffer, a commit operation will be performed because the index buffer is full. So that, all the data that have not been written into flash memory is recovered successfully.

Processing recovery operation under this order allows ERS to reduce the number of write operations and garbage collections of the buffer-based B-trees. By using logging and recovery policies, ERS is much more reliable and compatible than the buffer-based B-trees. In addition, flash memory cards are sensitive to electrostatic discharge (ESD) damage, which can occur when electronic cards or components are handled improperly, results in complete or intermittent failures. Therefore, deploying ERS will be good in practical systems.

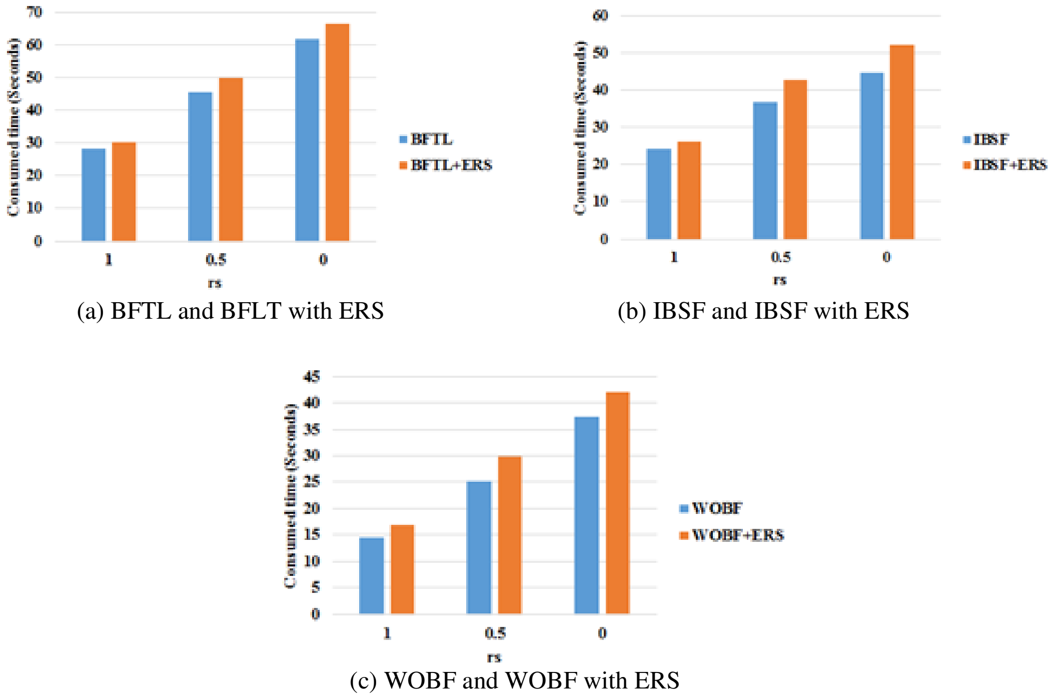
4. PERFORMANCE EVALUATION

This section shows the experimental results achieved by applying the proposed ERS and compares its performance to that of the buffer-based B-trees. All variant B-trees were performed on a NAND flash simulator which might be able to count the internal flash operations (read/write/erase). This simulator was configured for 64MB SLC NAND flash memory with 528 byte page size and 16 Kbyte block size. Every node of B-trees had 64 entries, each of which contained 4 byte integer key to search and a 4 byte pointer to point to the child node. The size index buffers are fixed as 64, and the index keys were unique integers in the range of 1 – 100,000. The performance of the buffer-based B-trees and ERS were assessed in terms of performance metrics: the average time to build B-trees and the recovery time. In order to control the key value distribution, a ratio called rs (ratio of the key sequence) was used. If the ratio was equal to 1, the key values were in ascending order. However, if the rs was equal to 0, the key values were randomly generated.

4.1. Performance of the B-trees creation

In this section, we assess the performance of ERS based on time consumption when building B-trees. Figure 3 presents the consumed time when constructing the buffer-based B-trees by inserting 100000 records. Overall, ERS yields about 8.2-11.3% overheads compared to those of the buffer-based B-trees on average. Concretely, ERS yields 8.19% overheads compared to that of BFTL, 10.56% overheads compared to that of IBSF and 11.32% overheads compared to that of

WOBF. The reason for these overheads is that ERS writes and manages the log record to the logger whenever a B-tree node is updated. However, the gap of their performance is smaller than we expected because the log records are sequentially written and does not yield the overwrite operation. Especially, the gap is about 9.23% when key values are fully sequential order.



(a) BFTL and BFLT with ERS

(b) IBSF and IBSF with ERS

(c) WOBF and WOBF with ERS

Figure 3. The Consumed time when inserting 100000 records

4.2. Performance of Recovery

Figure 4 presents the consumed time of recovery data when the systems reboot after crashes which randomly occurred. It can be seen that ERS quickly recover in all cases of rs. On average, it takes about 0.41 to 0.44 seconds to recover all the data which have not been written before the crashes occur. ERS consumes about 0.44 seconds for BFTL, 0.41 seconds for IBSF, and 0.42 seconds for WOBF to recover the data loss.

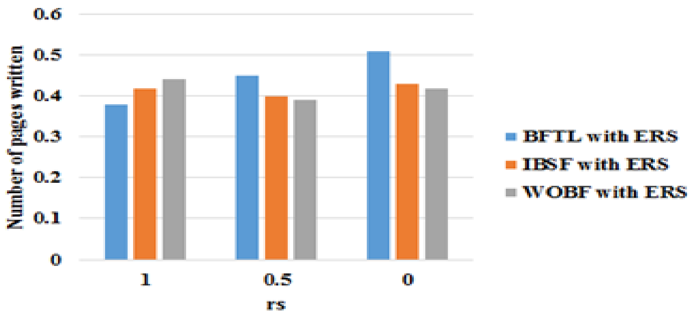


Figure 4. The recovery time

In fact, the consumed time for restoring the data does not depend on rs. Instead, it depends on the size of index buffer because ERS sets a commit record in the logger whenever the index buffer is successfully committed. This means if the size of index buffer is big, the number of records which needs to be recovered after a crash is large resulting in lots of time consumed for restoring the data.

Through these experiments, we can see that the performance of ERS is quite good. Besides that, it is much more reliable because it helps the buffer-based B-tree quickly recovers the data after a crash. Therefore, it will be good in practical systems.

5. CONCLUSION

Flash memory and B-tree index structure are widely used for embedded systems, personal computers, and large-scale server systems. Due to hardware restrictions, the performance of flash memory could significantly deteriorate when directly implementing B-tree. To solve this issue, many buffer-based B-tree index variants have been proposed for flash memory in order to reduce the number of flash operations. However, these B-tree indexes suffer from the risk of losing data when a sudden power-off occurs. In this study, we proposed a new recovery scheme for Buffer-based B-tree indexes on flash memory. The proposed system can minimize the loss of data by deploying logging and recovery policies. The experimental results show that ERS yields a good performance and helps the buffer-based B-tree indexes improve the reliability.

ACKNOWLEDGEMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1D1A1A01056593)

REFERENCES

[1] Shinde Pratibha et al. “Efficient Flash Translation layer for Flash Memory,” International Journal of Scientific and Research Publications, Volume 3, Issue 4, April 2013

[2] E.gal, S. Toledo, “Algorithms and data structures for flash memory,” ACM Computing surveys 37, 2005, pp138-163

[3] D. S. Batory, “B+-Trees and Indexed Sequential Files: A Performance Comparison,” Proceeding of Special Interest Group on Management of Data, 1981, pp. 30-39.

[4] Chin-Hsien Wu et al. “An Efficient B-Tree Layer Implementation for Flash Memory Storage Systems,” ACM Transactions on Embedded Computing Systems, Vol. 6, No. 3, Article 19, 2007

[5] Hyun-Seob Lee and Dong-Ho Lee, “An Efficient Index Buffer Management Scheme for Implementing a B-Tree on NAND Flash Memory,” Data & Knowledge Engineering, vol. 69, no.9, 2010, pp. 901-916.

[6] Xiaona Gong et al. “A Write-Optimized B-Tree Layer for NAND Flash,” Proceeding of the 7th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM), pp.1-4, 2011

[7] J. S. M. Verhofstad, "Recovery Techniques for Database Systems," ACM Computing Surveys, vol. 10, pp. 167-195, 1978.

[8] Drew et al., “A Comparison of File System Workloads,” Proceedings of the 6th USENIX Conference on File and Storage Technologies, 2000, pp. 41-54.

[9] Andrew W Leung et al., “Measurement and Analysis of Large-Scale Network File System Workloads,” Proceedings of the 6th USENIX Conference on File and Storage Technologies, 2008, pp. 213-226

AUTHORS

VanPhi Ho received his BS and MS degrees in the Computer Science Department at Da Nang University in October 2004 and October 2009, respectively. He is currently a Ph.D. student in the School of Computing at Soongsil University. His research interests include flash memory-based DBMSs and database systems.

Dong-Joo Park received his BS and MS degrees in the Computer Engineering Department at Seoul National University in February 1995 and February 1997, respectively, a Ph.D. in School of CS&E from Seoul National University in August 2001. He is currently an associate professor in the School of Computing at Soongsil University. His research interests include flash memory-based DBMSs, multimedia databases, and database systems.

Seung-Joo Jeong is currently a Master's course student in the School of Computing at Soongsil University. His research interest include flash memory-based DBMSs and database systems.

LIVE VIRTUAL MACHINE MIGRATION USING SHADOW PAGING IN CLOUD COMPUTING

SangWook Han[1] and HwaMin Lee[2]

[1]Department of Computer Science &Engineering, Soonchunhyang University,
Asan, South Korea
sanguk@sch.ac.kr
[2]Department of Computer Software Engineering, Soonchunhyang University,
Asan, South Korea
leehm@sch.ac.kr

ABSTRACT

Cloud Computing shares computing resources to execute application. Cloud systems provide high-specification resources in the form of services, leading to the provision of user convenience and greater ease for personal-computer users; however, expansions of the cloud-system service necessitate a corresponding enhancement of the technology that is used for server-resource management. In this paper, by monitoring the resources of a cloud server, we sought to identify the causes of server overload and degradation, followed by the running of a dynamic-page-migration mechanism. According to this process, we designed the proposed migration architecture for the minimization of user inconvenience.

KEYWORDS

Live Migration, Shadow Paging, Dynamic Page Migration, Cloud Computing

1. INTRODUCTION

Cloud computing has garnered the spotlight over recent years in the field of computing. Cloud computing is tailored to the needs of the user, and its services are provided regardless of the location of the user and the devices that are used. The specifications that are required by a user or a corporation can be attained via WAN or Internet, in a home or office, in cafes, or on public transport; that is, you can use the service from any location [1].

Cloud computing can be divided into the following types: SaaS, PaaS, and IaaS; a computer user's choice will depend on the form that is needed. Regarding PaaS and SaaS, a computing environment must be provided for the user because the provision of hardware is too expensive in these cases; therefore, server visualization is used to reduce the cost [4]. Visualization involves the installation of a virtual machine monitor as systemic software that manages the virtualized operating systems on a single hardware unit, thereby supporting the virtual machine that is provided to the user, and the software is in the desired form for the installation of an operating system [1]. In terms of cloud computing, virtualization can be considered an application of the

Natarajan Meghanathan et al. (Eds) : NETCOM, NCS, WiMoNe, GRAPH-HOC, SPM, CSEIT - 2016
pp. 187– 193, 2016. © CS & IT-CSCP 2016 DOI : 10.5121/csit.2016.61516

operating system. From the perspective of administrators, virtualization means that it is possible to deliver a service through the provision of an infrastructure and platform.

Although cloud computing is advantageous, however, two significant problems have emerged in relation to the use of virtual operating systems. First, hardware utilization is rapidly increased due to the frequent occurrence of page faults, leading to thrashing; and second, when a hardware failure occurs in the virtual machine, all of the corresponding services are interrupted. Resource management has become increasingly important because of these problems, and hardware-related research studies are in progress to develop measures that minimize user inconvenience without interrupting services. The purpose of this paper is the improvement of the efficiency of existing migration methods.

This paper is organized as follows: chapter 2 looks at the existing content regarding migration; chapter 3 proposes a mechanism for an efficient migration process through the use of monitoring and the dynamic-paging-migration technique; and in chapter 4, the conclusion provides a direction for future research initiatives.

2. RELATED WORKS

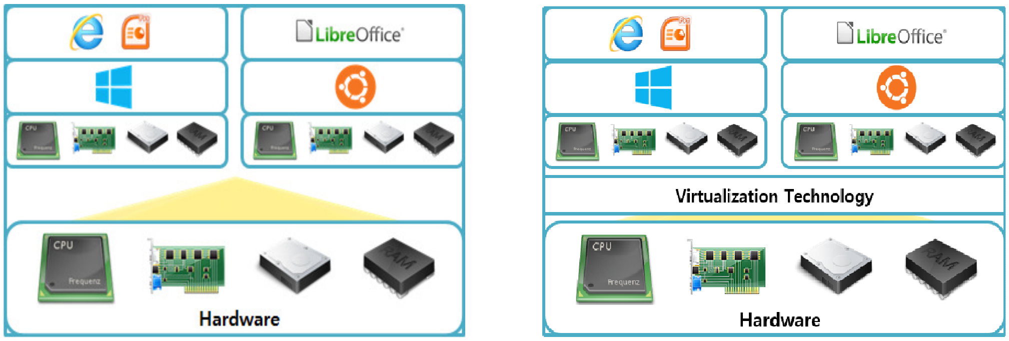
2.1. Monitoring

The shapes and sizes of the clouds of cloud computing are widely variable. Because cloud users are difficult to manage personally, automation is required, and this necessitates an interaction with the surroundings. Monitoring can be seen as one of the best techniques for the configuration of a cloud system; by utilizing a monitoring system, it is possible for a user to perform the following actions, among others:

- Auto VM provisioning
- Auto scaling
- Auto service provisioning
- High availability
- Deploy management

2.2. Hypervisor

A hypervisor is the tool base for virtualization. Regarding commercial clouds, there is a variety of hypervisors that is divided into full virtualization and para-virtualization, depending on the virtualization method [2].



(a) Full virtualization (b) Para virtualization

Figure 1. Types of Virtualization

Full virtualization requires the virtualization of all of the hardware of a system, while the guest operating system remains unchanged, and the advantage is that it can be applied for a variety of operating systems.

As the name suggests, para-virtualization is applicable for a system wherein the hardware has not been fully virtualized; therefore, a guest operating system does not control the hardware.

2.3. Shadow Paging

Shadow Paging (Dynamic page migration) is a transition technique whereby a page is transmitted through the use of the physical memory space of a shadow page [1]. For the performance of a migration (the service runs on a virtual server), the information of the changed pages needs to be saved in the shadow page before the transferal to a newly allocated space can occur; therefore, this technique makes it possible to prevent service failure. Shadow paging involves the creation and saving of a physical memory space for the storage of the information of a page that has been changed by a user, and the newly created space is referred to as the "shadow page." Dynamic page migration comprises a mobile data system, whereby small amounts of memory space are progressively allocated to the target machine; by using the shadow page, mere seconds of service downtime occur, meaning that the user does not experience any inconvenience. The corresponding details are provided in chapter 3.2.

2.4. Orchestration

Orchestration is the management of resources whereby arrangement and alignment are automated. In terms of cloud-computing services, the following actions should be functional, among others: the issuance of an authentication key, the creation of a network check, and the creation of security rules. Orchestration cannot, however, be directly engaged each time an administrator creates an instance. A template-based engine can be used for the easy automation of this process that enables the deployment of infrastructure [2].

3. VM LIVE MIGRATION TECHNIQUE

3.1. Monitoring-system configuration

The previous monitoring system (Openstack ceilometer) offers the resources that need to be deployed in a cloud and enables a user to monitor the statuses of the resources for a performance assessment; furthermore, the program offers visibility and insight through the monitoring of the resource state of the dispersed cloud system [2]. The previous monitoring system, however, should change the internal source code according to systemic needs; therefore, we propose a monitoring system in this paper that can address this limitation. As shown in Table 1, the essential element of the monitoring system was selected.

Using the proposed monitoring system, it is possible to obtain a value for dynamic page migration by monitoring the information shown in Table 1. In addition, the system can be used for the determination of the necessary information such as the page that needs to be sent when the information shown in Table 1 is obtained through the migration mechanism.

Table 1. Information required for monitoring

Edit page frequently
Network transfer rate
Page size
Total size of virtual-machine memory

3.2. Shadow-paging migration method

If a server overload or degradation is detected through monitoring, the migration process is commenced. The corresponding migration-operation process is shown in Figure 3 and Table 2.

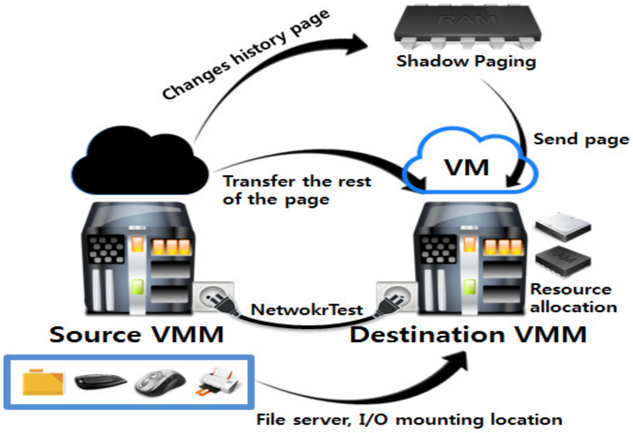


Figure 2. Proposed Shadow-paging Technique

A step-by-step description is shown in Table 2, as follows:

Table 2. Shadow-paging step-by-step instructions

1. Initialization phase : - Target-machine selection - Socket opening
2. Reservation phase : - Connection of socket - Allocation of a resource to the target machine
3. Iterative pre-copy phase : - Storage of the changed information in the shadow page
4. Stop-and-copy phase: - Sending of the rest of the page
5. Commitment phase : - Mounting of the file server and I/O device to the target machine
6. Activation phase: - Service activation

3.3. Full configuration for a virtualization migration

In this paper, the flow of the proposed migration method is shown in Figure 4. If an overload or deterioration occurs in the server then the allocation of resources occurs quickly for the determination of the migration.

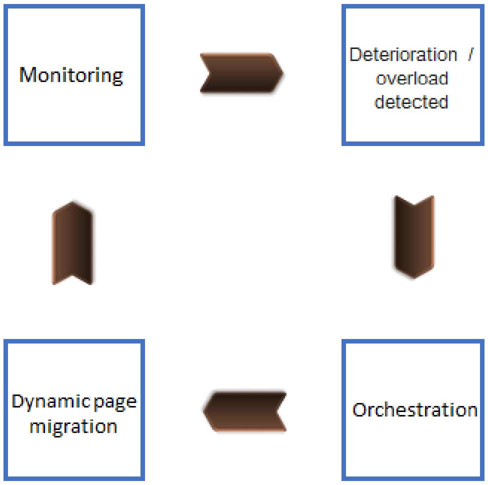


Figure 3. Proposed-migration-structure flow

The proposed monitoring system checks the status of the server, the Target-machine Selection Algorithm selects the target machine, and an automated orchestration process allocates the resources. By using the dynamic-page technique, the migration is performed.

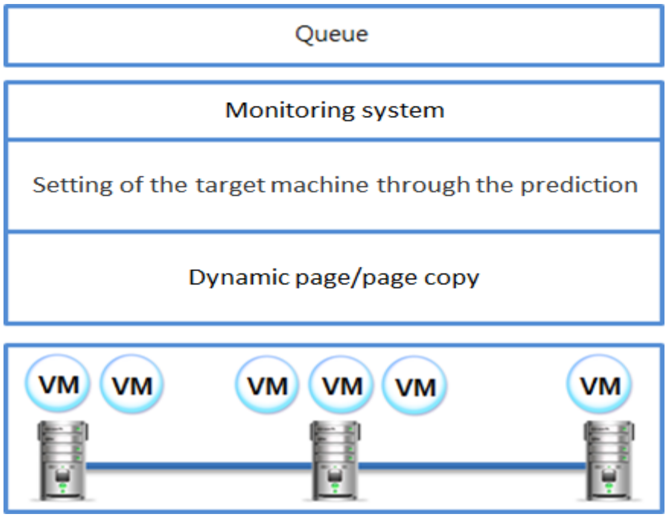


Figure 4. Proposed system structure

4. CONCLUSION

The development of cloud computing has led to many changes for both individuals and companies over recent years. Companies that provide cloud computing are therefore obliged to provide the best services to their customers and the companies in receipt of their services; furthermore, in addition to ensuring the effective management of their resources, they need to actively conduct research regarding migration. The method that is proposed in this research paper monitors resources to identify any deterioration or overloading regarding a server, and it performs migration in real time; provided that a target machine is nominated, this proposed method can be used for a server migration while the services are maintained by shadow paging. Further study is required regarding effective algorithms for shadow-paging migration; accordingly, this paper's objective is the development of a faster and more-precise migration technique for the harmonization of target-algorithm selection and the dynamic-page-migration algorithm.

ACKNOWLEDGEMENTS

This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2016-R0992-16-1006) supervised by the IITP(Institute for Information & communications Technology Promotion).

REFERENCES

[1] Kapil, D., Pilli, E. S., & Joshi, R. C. (2013) Live virtual machine migration techniques: Survey and research challenges, In 2013 IEEE 3rd International Advance Computing Conference, pp. 963-969.

[2] Son, A. Y., & Huh, E. N. (2016) Migration Method for Seamless Service in Cloud Computing: Survey and Research Challenges, In 2016 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA), pp. 404-409.

[3] Wood, T., Ramakrishnan, K. K., Shenoy, P., & Van der Merwe, J. (2011) CloudNet: dynamic pooling of cloud resources by live WAN migration of virtual machines, In ACM Sigplan Notices, Vol. 46, No. 7, pp. 121-132.

[4] Kim, B., (2013) “An Efficient Method for Live Virtual Machine Migration Time Using Prediction of Pre-copy Phase", Master’s Thesis, Seogang University, Korea.

[5] Du, Y., Yu, H., Shi, G., Chen, J., & Zheng, W. (2010) Microwiper: efficient memory propagation in live migration of virtual machines. In 2010 39th International Conference on Parallel Processing, pp. 141-149.

[6] Ahmad, R. W., Gani, A., Hamid, S. H. A., Shiraz, M., Yousafzai, A., & Xia, F. (2015). A survey on virtual machine migration and server consolidation frameworks for cloud data centers. Journal of Network and Computer Applications, 52, 11-25.

[7] Zheng, J., Ng, T., Sripanidkulchai, K., & Liu, Z. (2013). Pacer: A progress management system for live virtual machine migration in cloud computing.IEEE transactions on network and service management, 10(4), 369-382.

[8] Nathan, S., Bellur, U., & Kulkarni, P. (2015, August). Towards a comprehensive performance model of virtual machine live migration. In Proceedings of the Sixth ACM Symposium on Cloud Computing (pp. 288-301). ACM.

[9] Sun, G., Liao, D., Anand, V., Zhao, D., & Yu, H. (2016). A new technique for efficient live migration of multiple virtual machines. Future Generation Computer Systems, 55, 74-86.

[10] Jackson, K., Bunch, C., & Sigler, E. (2015). OpenStack cloud computing cookbook. Packt Publishing Ltd.

AUTHORS

Sangwook Han

He is master course student at Dept. of Computer Software Engineering in Soonchunhyang University. His main research interests are cloud and mobile computing.

HwaMin Lee

She is a Professor at Dept. of Computer Software Engineering in Soonchunhyang University. Her main research interests are cloud and mobile computing, Internet of Things, and IT convergence.

INTENTIONAL BLANK

AN EMPIRICAL STUDY OF USING CLOUD-BASED SERVICES IN CAPSTONE PROJECT DEVELOPMENT

Zhiguang Xu

Department of Computer Science, Valdosta State University,
Valdosta, GA, USA
zxu@valdosta.edu

ABSTRACT

Cloud computing is gaining prominence and popularity in three important forms: Software as a Service, Platform as a Service, and Infrastructure as a Service. In this paper, we will present an empirical study of how these cloud-based services were used in an undergraduate Computer Science capstone class to enable agile and effective development, testing, and deployment of sophisticated software systems, facilitate team collaborations among students, and ease the project assessment and grading tasks for teachers. Especially, in this class, students and teachers could leverage time, talent, and resources collaboratively and distributedly on his/her own schedule, from his/her convenient location, and using heterogeneous programming platforms thanks to such a completely All-In-Cloud environment, which eliminated the necessity of spending valuable development time on local setup, configuration, and maintenance, streamlined version control and group management, and greatly increased the collective productivity of student groups. Despite of the relatively steep learning curve in the beginning of the semester, all nine groups of students benefitted tremendously from such an All-In-Cloud experience and eight of them completed their substantial software projects successfully. This paper is concluded with a vision on expanding and standardizing the adoption of the Cloud ecosystem in other Computer Science classes in the future.

KEYWORDS

SPI, Cloud Computing, Software Development, Capstone Project, Computer Science Education

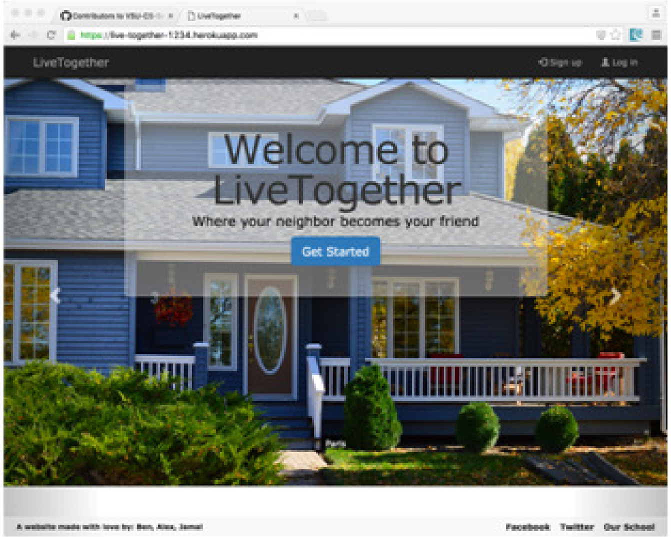
1. INTRODUCTION

CS 4900, Senior Seminar, is a project-driven course designed to provide senior capstone experiences for graduating Computer Science (CS) majors at Valdosta State University (VSU). While oftentimes, students produce impressive applications, much of their efforts centre only on the task of coding such applications itself with rare, if any, concerns about how it fits into a much larger enterprise-level picture, how to factor in the real-world software production parameters such as bottom-line economics, control of data, security, compatibility, etc., and ultimately, the strategy to respond to the rapid migration toward cloud computing as the framework for most modern applications in today's industry [3]. Without a well-architected exposure to the cloud and an easy-to-follow procedure to take advantage of the services provided by it, such a disconnection between the academia and industry results in a great deal of students' valuable time being spent only on their local computers performing tedious installation and configuration of the software

Natarajan Meghanathan et al. (Eds) : NETCOM, NCS, WiMoNe, GRAPH-HOC, SPM, CSEIT - 2016
pp. 195– 209, 2016. © CS & IT-CSCP 2016 DOI : 10.5121/csit.2016.61517

development environment and the quality test of their software products. Moreover, as the complexity of the applications they build scales up, it is a challenging undertaking for the students to conduct efficient and effective team collaborations when they work distributedly and for the instructor to help the students during the semester and evaluate individual student's performance at the end of it.

In spring of 2016, we addressed this problem head on in Senior Seminar. Thirty students in this class formed nine groups to build a full-fledged Web server application completely in cloud. The application was called FriendsNextDoor, a private social network service that allows users to connect with people who live in their neighbourhood, and it was mainly written in Ruby on Rails. CS faculty at VSU including the instructor of the class attended the project demonstrations at the end of the semester. They concluded that eight out of nine student groups completed their projects with an "excellent" overall quality. Homepage of one of such projects, LiveTogether as they named it, is show in Figure 1 below. In addition, the end-of-semester survey indicated that 86% of students strongly agreed that "the all-in-cloud programming ecosystem used in the capstone project gave them a great exposure to what it means to work in a professional context".



Welcome to LiveTogether

Where your neighbor becomes your friend

A website made with love by: Ben, Alex, Jamal

Figure 1. Homepage of a Student Capstone Project

SPI is an acronym for three popular IT paradigms under the umbrella notion of cloud computing: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). In Section 2, we briefly present how SPI has evolved in the recent years, in particular, the opportunities and challenges they brought to the Computer Science education. Then, in section 3, we outline the requirements of the capstone project that students were asked to develop in the Senior Seminar class. Then we present a big picture along with one detailed configuration example of how the SPI ecosystem was actually exploited to fulfil such requirements with an emphasis on the All-In-Cloud (AIC) programming settings. This is followed by section 4, where seven specific SPI services/products are selected to elaborate the procedures of integrating them into our capstone projects in technological details. In Section 5, we evaluate the final products of the capstone projects based on the assessments from both the students' and CS faculty's perspectives. Finally, in Section 6, we summarize the paper with conclusions and a vision on improving the AIC-SPI model and the expansion of adopting it in other CS classes in the future.

2. COMPARING TRADITIONAL IT AND CLOUD-BASED SPI

National Institute of Standards and Technology (NIST) defines Cloud Computing as Internet-based technology, which offers computational resources such as large storage capacity, high network bandwidth, and vast processing power via a computer network and delivers flexible, scalable, and on-demand services to the end-users [1]. In comparison to traditional IT, such resources and services are classified into three "SPI" layers as shown in Figure 2 below, each serving different purposes and tasks. Infrastructure as a Service (IaaS) delivers physical or (more often) virtual machines and other resources such as servers, storage, and networking connectivity, for example, Amazon Simple Storage Service (S3). The Platform as a Service (PaaS) layer acts as a container enabling the users to modify and develop their platform and deploy their applications. Typical example at this layer is Heroku. The most abstract layer is Software as a Service (SaaS), which allows users to run hosted applications on the Cloud and use them through a web browser remotely, e.g. Cloud9, an online IDE with full Ubuntu workspace. Besides others, all three examples mentioned above were heavily used in our Senior Seminar capstone projects.

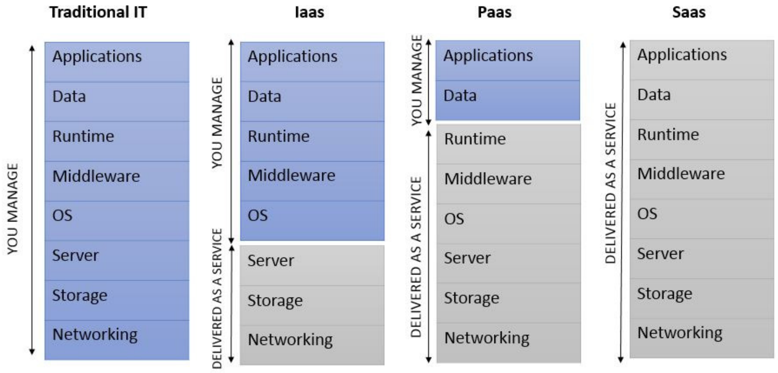


Figure 2. Traditional IT vs. SPI

Cloud computing in less than a decade has gone from utility computing with nearly unlimited on-demand scalability at very attractive prices, to the industrialization of IT that has been pioneered by born-to-the-cloud companies like Amazon, Google, and Salesforce, to the total transformation to digital business [2] as nicely summarized in a recently released report State of the Market: Enterprise Cloud 2016 from Verizon Communications Inc. – "As cloud increasingly becomes the norm … it is not enough to think cloud first. You need to think cloud only."

In the context of CS education that is inherently and tremendously impacted by the Cloud, technologically speaking, the advantages of SaaS, PaaS, and IaaS solutions are simplicity of integration (most of the time, students need only a web browser), reduced cost (the code/data center resides within the cloud rather than local), and scalability (resources are dynamically and transparently allocated in response to the fluctuation in data and computation sizes). Among others, there are two major challenges of adopting SPI – data security and "lock-in" to the products of a particular provider [3, 4]. During the development of the capstone projects in senior seminar, we encountered both of such issues, which will be discussed more in sections 4 and 5.

3. USE CLOUD-BASED SPI IN CAPSTONE PROJECT DEVELOPMENT

3.1. Capstone Project Requirements

Consider FriendsNextDoor a modern, more attractive, more secure, and more versatile version of a community email list service or Yahoo Groups, the popular message board, where regular homeowners, community leads, Home Owner Associations, local businesses, public agencies, and other neighbourhood constituencies can post neighbourhood news, offer items for sale, seek/provide babysitting opportunities, lend/borrow yard tools, ask for help finding lost pets, or organize a block party, just to name a few. Although it was an open ended project, the following list of features were required:

- For any potential user, signing up to the web site should be subject to some sort of verifications, e.g. location/address, invitation, approval by the community leads, or a combination of them.

- Users are of different roles with different levels of privileges/responsibilities. Such roles at least include regular users, community leads, local businesses, public agencies, and system admin.

- Users sharing common interests can form, join, and leave groups.

- Two types of message/post boards: public board and group-specific private board. And optionally, a real-time chat system.

- Events section where event organizers and participants can share information (e.g. through calendar, map, photo album, etc.)

- Community leads should be elected through a point system. They cannot be simply appointed.

- Optionally, a RSS feed publish/subscribe system to share information among multiple neighborhoods.

3.2. A High-level View

In order to provide a high-level view and a global picture of how the AIC-SPI ecosystem fits in our capstone projects, a visualization of the most popularly used SPI services in today's software ecosystem that is custom-rendered on https://modeanalytics.com/ is displayed in Figure 3 below.

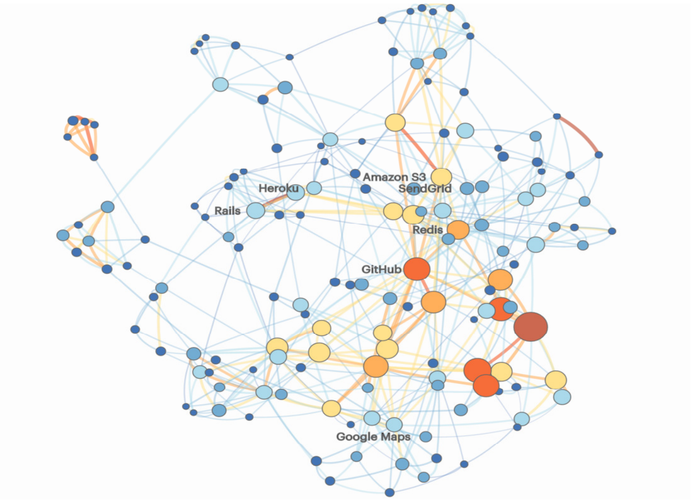


Figure 3. The SPI Ecosystem

Dots represent software services/products, and the lines between them represent companies using both services they connect. The bigger and redder the dot, the more companies use the service. The thicker and redder the line, the more companies use two services together. Particularly, those labeled dots in Figure 3 were the SPI services/products that were actually chosen to be used in the FriendsNextDoor projects by our Senior Seminar students themselves. Such services belonging to different layers in the AIC-SPI model as shown in Figure 4 were integrated into FriendsNextDoor through respective Rails "gems" (i.e. third-party libraries). As an example, one of such gems named fog for Amazon S3 is going to be discussed in the next section to show how such integration was configured at some technically low level.

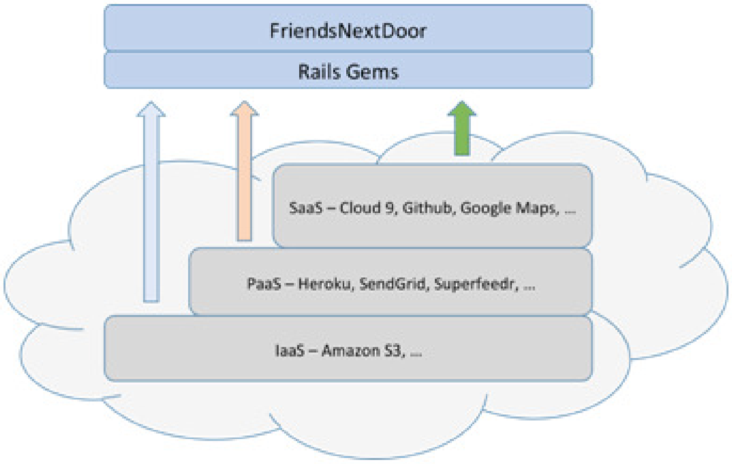

FriendsNextDoor
Rails Gems
SaaS – Cloud 9, Github, Google Maps, ...
PaaS – Heroku, SendGrid, Superfeedr, ...
IaaS – Amazon S3, ...

Figure 4. The SPI Layers

3.3. An Example of SPI Configuration at Low-level

Rails leverages a wide set of third-party libraries, most of which are released in the form of a "gem". A few selected Rails gems commonly used in our FriendsNextDoor application along with some brief descriptions are listed in Table 1 below.

Table 1. Selected Gems used in FriendsNextDoor

Gem Name	Description
fog	A powerful cloud services gem
jquery-rails	A JavaScript library
bcrypt	Securing password
bootstrap-sass	A sass-powered version of Bootstrap
devise	User authentication
pundit	User authorization
sqlite3	DBMS for SQLite
pg	DBMS for PostgreSQL
carrierwave	Uploading files
gmaps4rails	Google Maps solution for Rails
simple_calendar	A calendar render
figaro	A Heroku-friendly Rails app configuration tool
websocket-rails	Websocket support in Rails
forem	A Rails forum engine

In particular, the first gem on the list, fog, was used to switch the file uploader in our project (built upon another gem carrierwave) from using regular file storage that suffers from suboptimal efficiency to using Amazon S3, one of the most popular cloud storage providers. Sensitive information such as access keys and passwords can be configured in an automatic and secure fashion via yet another gem of figaro.

Specifically, thanks to the fact that in the current version of fog, all fog providers are getting separated into meta-sub-gems (e.g. for-aws, fog-google, and fog-vsphere, etc.) to lower the load time and dependency count, the only fog gem that we actually needed to install was fog-aws. The configuration of the fog gem turned out to be very simple and straightforward. It only involved two files `config/initializers/carrierwave.rb`:

```
if Rails.env.production?
  CarrierWave.configure do |config|
    config.fog_credentials = {
      # Configuration for Amazon S3
      :provider              => 'AWS',
      :aws_access_key_id    => ENV['S3_ACCESS_KEY'],
      :aws_secret_access_key => ENV['S3_SECRET_KEY']
    }
    config.fog_directory     =  ENV['S3_BUCKET']
  end
end
```

and `app/uploaders/picture_uploader.rb`:

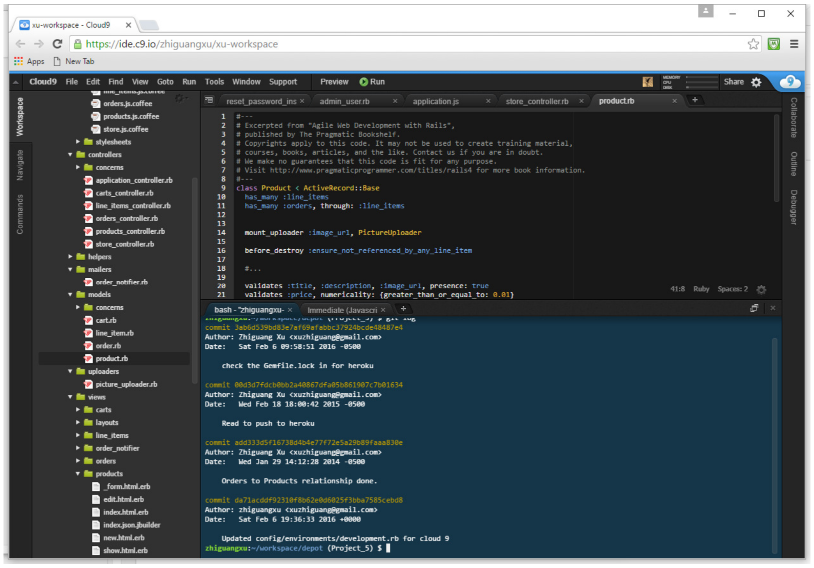
```
# Choose what kind of storage to use:
if Rails.env.production?
    storage :fog
else
    storage :file
end
```

Other SPI services/products were integrated into the FriendsNextDoor application in a similar fashion.

4. AIC-SPI IN ACTION

4.1. Cloud9 – Online IDE with Full Ubuntu Workspace

Cloud9 as shown in Figure 5 provides a hosted development environment to which you get access from a web browser. A cloud based IDE has been on everyone's mind since the concept of asynchronous web applications started taking a hold among developers as well as CS educators. However, it is Cloud9 as a key component in our AIC-SPI ecosystem that makes it truly "all in cloud". Unlike Linux virtual private servers from sites such as Linode, the Cloud9 service is free and the system is maintained for us so we will spend more time as a developer and less time as a sysadmin.



```ruby
#---
# Excerpted from "Agile Web Development with Rails",
# published by The Pragmatic Bookshelf.
# Copyrights apply to this code. It may not be used to create training material,
# courses, books, articles, and the like. Contact us if you are in doubt.
# We make no guarantees that this code is fit for any purpose.
# Visit http://www.pragmaticprogrammer.com/titles/rails4 for more book information.
#---
class Product < ActiveRecord::Base
  has_many :line_items
  has_many :orders, through: :line_items


  mount_uploader :image_url, PictureUploader

  before_destroy :ensure_not_referenced_by_any_line_item

  #...

  validates :title, :description, :image_url, presence: true
  validates :price, numericality: {greater_than_or_equal_to: 0.01}
```

```
zhiguangxu:~/workspace/depot (Project_5) $ git log
commit 3ab6d539bd83e7af69afabbc37924bcde48487e4
Author: Zhiguang Xu <xuzhiguang@gmail.com>
Date:   Sat Feb 6 09:58:51 2016 -0500

    check the Gemfile.lock in for heroku

commit 00d3d7fdcb0bb2a40867dfa05b861907c7b01634
Author: Zhiguang Xu <xuzhiguang@gmail.com>
Date:   Wed Feb 18 18:00:42 2015 -0500

    Read to push to heroku

commit add333d5f16738d4b4e77f72e5a29b89faaa830e
Author: Zhiguang Xu <xuzhiguang@gmail.com>
Date:   Wed Jan 29 14:12:28 2014 -0500

    Orders to Products relationship done.

commit da71acddf92310f8b62e0d6025f3bba7585cebd8
Author: zhiguangxu <xuzhiguang@gmail.com>
Date:   Sat Feb 6 19:36:33 2016 +0000

    Updated config/environments/development.rb for cloud 9
zhiguangxu:~/workspace/depot (Project_5) $
```

Figure 5. Cloud 9 IDE

In senior seminar, the following features that Cloud9 provides are highlighted:

- Cloud9 offers complete and seamless Git integration with services like Github and Bitbucket. Students were able to login to Cloud9 using their Github accounts, pull and push code in between, deploy applications to online host such as Heroku, and collaborate with teammates through live coding.

- The virtualized backends allowed students to spin up instances of servers (which is always up to date and compatible with the latest Rails) when load increased.

- Cloud9 was a lifesaver (as students in senior seminar called it) for PC users since Rails and most of the related SPI services/products are not necessarily Windows-friendly.

- The Ubuntu terminal access to command line functionalities was particularly appropriate for some students who wanted to learn UNIX but might be overwhelmed by the effort required to install and configure Ubuntu Linux.

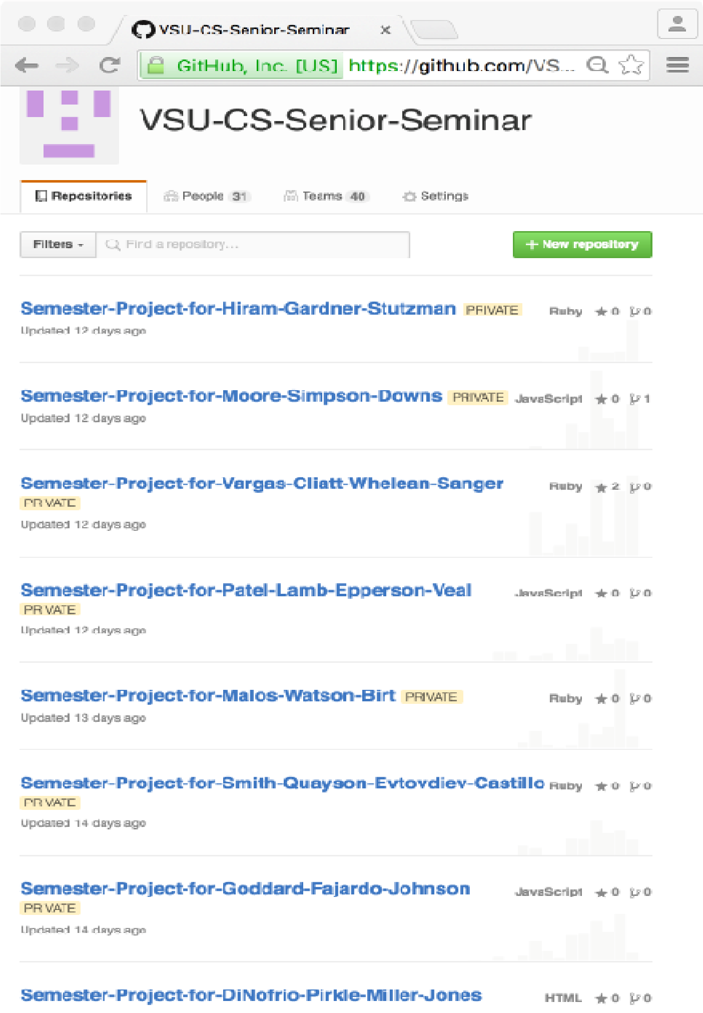
4.2. Github – Git Repository Hosting Service

Github was chosen as the online Git repository-hosting site for senior seminar due to the following reasons:

- Free private repositories, thanks to Github's educational program that allow students and teacher to securely access their projects from anywhere at any time [5]. We created an Organization "VSU-CS-Senior-Seminar" on Github that has 31 members (30 students and 1 teacher) and 40 private repositories (30 for student individual projects, 9 for group projects, and 1 for the teacher, see Figure 6).

- Secure source code backup in the Cloud (The proved important when one student's laptop crashed in the middle of the semester and it was his backups on Github that saved his project).

- Seamless integration with Cloud9.



Semester-Project-for-Hiram-Gardner-Stutzman PRIVATE Ruby ★ 0 ⑂ 0
Updated 12 days ago

Semester-Project-for-Moore-Simpson-Downs PRIVATE JavaScript ★ 0 ⑂ 1
Updated 12 days ago

Semester-Project-for-Vargas-Cliatt-Whelean-Sanger Ruby ★ 2 ⑂ 0
PRIVATE
Updated 12 days ago

Semester-Project-for-Patel-Lamb-Epperson-Veal JavaScript ★ 0 ⑂ 0
PRIVATE
Updated 12 days ago

Semester-Project-for-Malos-Watson-Birt PRIVATE Ruby ★ 0 ⑂ 0
Updated 13 days ago

Semester-Project-for-Smith-Quayson-Evtovdiev-Castillo Ruby ★ 0 ⑂ 0
PRIVATE
Updated 14 days ago

Semester-Project-for-Goddard-Fajardo-Johnson JavaScript ★ 0 ⑂ 0
PRIVATE
Updated 14 days ago

Semester-Project-for-DiNofrio-Pirkle-Miller-Jones HTML ★ 0 ⑂ 0

Figure 6. Repositories on Github

- Clean and fast submission and grading of projects, especially when their sizes went beyond megabytes.

- Rich tools for administrating student groups, visualizing students' contributions to their group projects, archiving projects for future course assessments, and much more [5]. Grading individual student based on his/her contributions to a group programming project like the ones we had in senior seminar has always been a challenge for the CS teachers. Thanks to the graphs provided by Github where each group member's commits are chronically tracked as shown in Figure 7, along with other traditional project deliverables such as group report, workloads could be distributed more balanced among group members, "free riders" could be identified more early-on, and grades could be assigned more fair and reasonable.

- Source code and tutorials of most of the third-party libraries (i.e. Rails gems) are housed on Github.

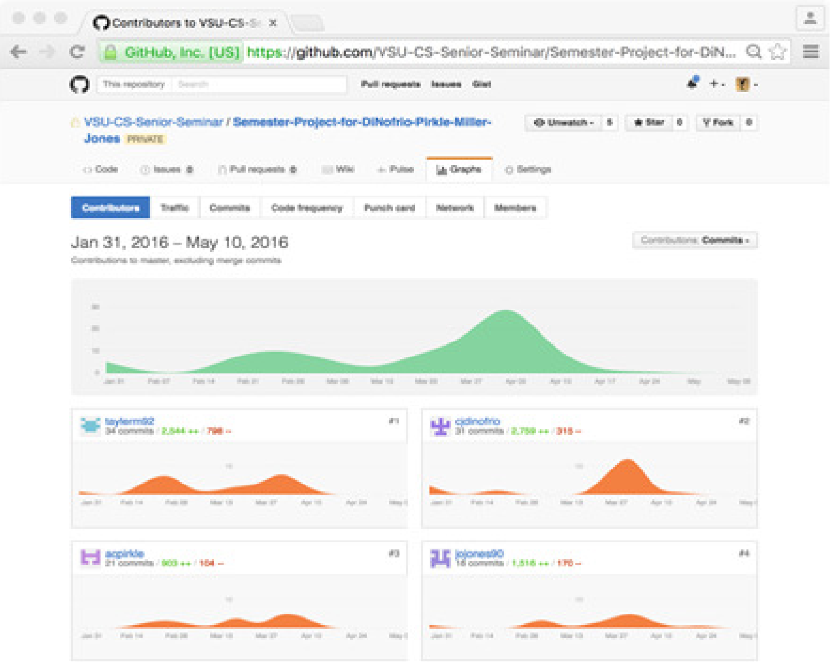


Figure 7. Commits Graph from Individual Students

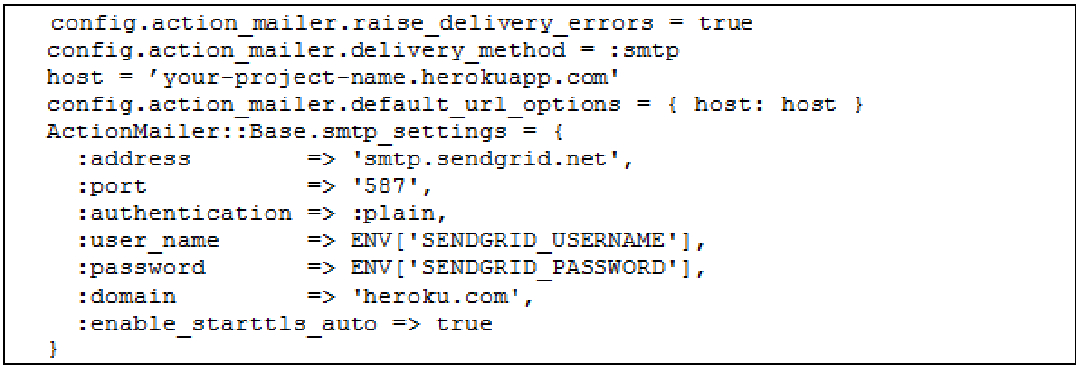
4.3. AWS S3 – Cloud Storage

The file/image uploader via a third-party gem named “carrierwave” is good enough for preliminary development and testing on Cloud9, however, it uses the local file system (public/uploads) for storing the files/images, which isn’t a good and efficient practice and produces lots of headaches in production on Heroku. Therefore, we used one of the most popular and well-supported cloud storage, Amazon Simple Storage Service (S3) to store files/images separately from our application server. Details on the integration of Amazon S3 into our applications were discussed in section 3.1 above.

4.4. Sendgrid – Email Delivery and Management System

Instead of using a private email service provider such as Microsoft Outlook and Gmail which would cause problems of being locked in to these specific products, we outsourced the email delivery and management to SendGrid, an in-cloud transactional email service provider, to send potentially mass emails. In addition to better throughput when handling large volume of emails, other benefits of using SendGrid include better deliverability, professional level reliability, and transparent performance analytics.

SendGrid makes the email configuration fairly easy. The file below (`config/environments/production.rb`) needs to be updated for running the project on Heroku. Similar to what was discussed for Amazon S3, environment variables SENDGRID_USERNAME and SENDGRID_PASSWORD can be encrypted using the Figaro gem for better security purposes.



```
config.action_mailer.raise_delivery_errors = true
config.action_mailer.delivery_method = :smtp
host = ’your-project-name.herokuapp.com'
config.action_mailer.default_url_options = { host: host }
ActionMailer::Base.smtp_settings = {
  :address        => 'smtp.sendgrid.net',
  :port           => '587',
  :authentication => :plain,
  :user_name      => ENV['SENDGRID_USERNAME'],
  :password       => ENV['SENDGRID_PASSWORD'],
  :domain         => 'heroku.com',
  :enable_starttls_auto => true
}
```

4.5. Google Maps – Location Based Services from Google

Several groups implemented a location based user verification system via Google Maps API by showing a visual verification of a user’s neighbourhood as a superimposition of the neighbourhood’s outline on the map (Figure 8).

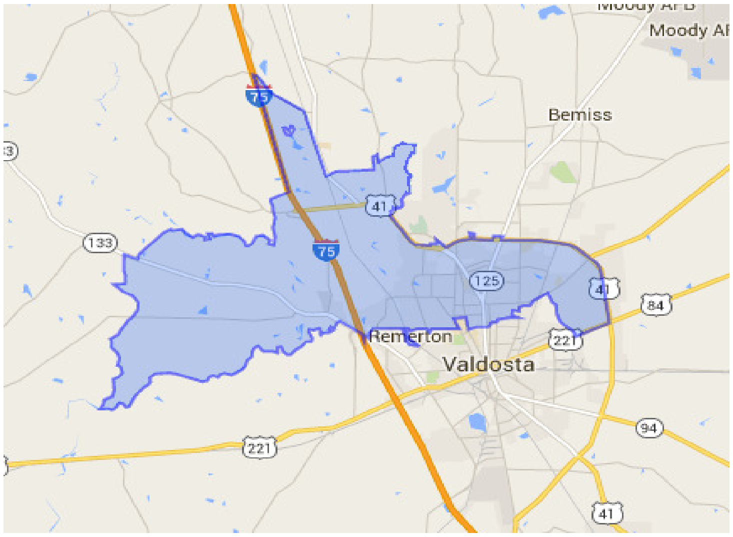


Figure 8. Visual Verification of Neighborhood

Each outline was stored in a kml file that stored Google Maps compliant polygons in xml syntax. The complexity came when accommodating every zip code in the United States. These approximately 30,000 kml files were parsed from a single large file (totalling 1.4 million lines of code) using a small Java program. Each kml file was then independently hosted providing their FriendsNextDoor application with embedded Google Maps based services without inducing a prohibitively long delay in load or draw times (Figure 9).

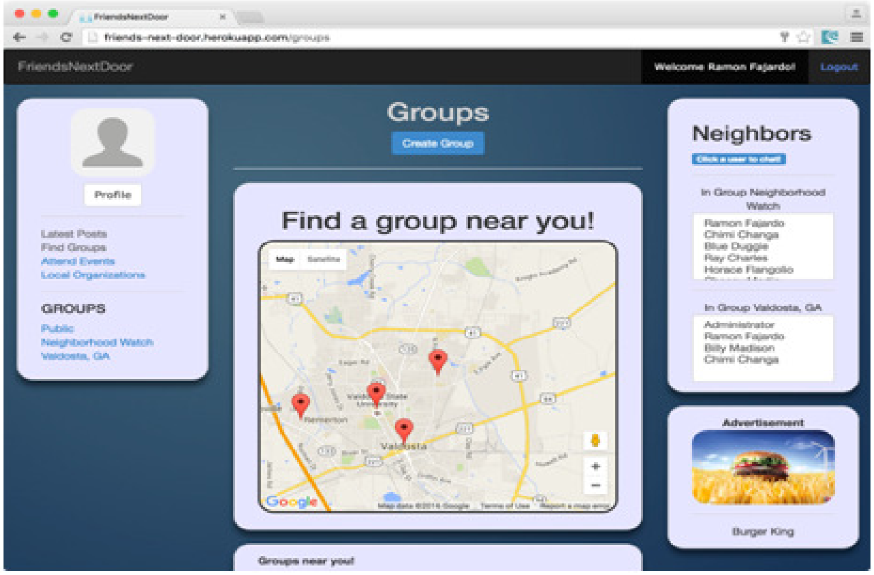


Figure 9. Location Based Service in a Student Project

4.6. Superfeedr – Hub for RSS Feed Publication and Subscription

In order to facilitate cross-site information sharing and dissemination, a Rich Site Summary (RSS) feed publication/subscription system was built in the FriendsNextDoor site developed by one group of three students as their honours option project.

Pubsubhubbub (PuSH) is an open protocol for distributing content from RSS publishers to subscribers that improves upon the RSS specification by adding an intermediate server to serve as a "hub" server at https://superfeedr.com which assists both the subscriber and the publisher with RSS communications.

Specifically, the hub eases the load of the publisher by establishing a subscription system and handing all subscription requests to the publisher's feed, reducing the number of calls to access the publisher's feed to only the occasional calls by the hub to fetch the RSS page check it for updates. The hub also eases the burdens of the subscribers by not requiring them to have to frequently request for any potential updates. Instead, subscribers can subscribe to a publisher's feed by sending an HTTP POST request to the hub containing the publisher's feed URL and a URL for the subscriber's webhook – a URL that routes to a callback function running on the subscriber's server, telling it to process that HTTP request body in a specific way. The webhook will "listen" for updates from the hub and store any updates it receives in the subscriber's database. In this way, when a publisher makes changes to their RSS feed, the hub will "push" notifications containing the new RSS page to all of the subscribers, rather than having them constantly probe for new information (Figure 10).

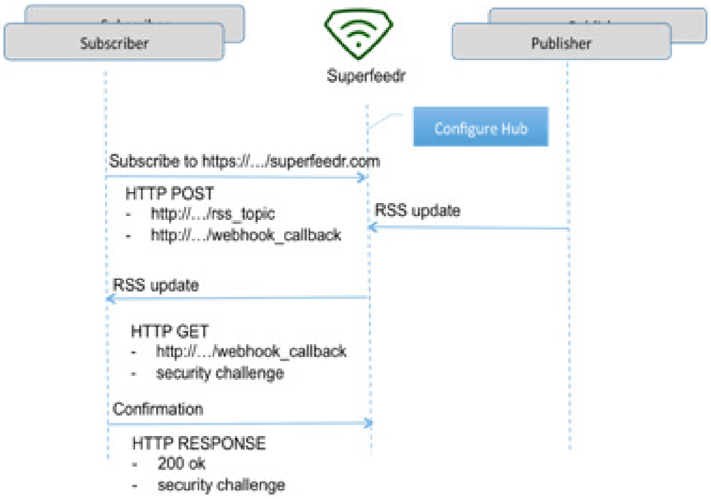


Figure 10. PuSH and the Hub Server

4.7. Heroku – Cloud PaaS

After developing a Rails application locally on Cloud9, the next step is to deploy it online. Heroku is a popular PaaS hosting service that is free to start using. Typically, in a Ubuntu terminal that is built into Cloud9, the deployment process involves running a sequence of command-line commands and scripts such as git-push the source code of the application to Heroku, install relevant gems and their dependencies, create and pre-seed database tables with certain data, configure and encrypt environmental variables, and fire up separate servers for various purposes. In Figure 11, three of such servers in addition to the Rails server are shown: a Solr server for site-wise searches, a Faye server for live chattings, and a Redis server for real-time RSS feed updates.

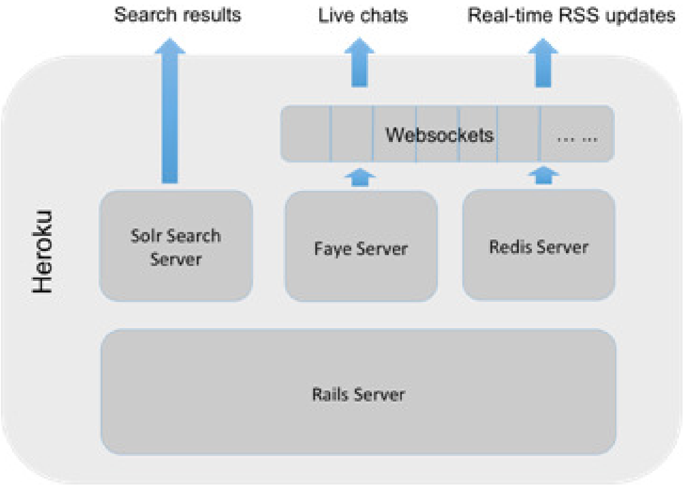


Figure 11. Rails and Other Servers Running on Heroku

Specifically, as a continuation discussion of the RSS feed feature presented in section 4.6 above, a major issues faced in the implementation of that feature was live pushing of updated feed data once it was received by the subscriber, parsed, and stored in the subscriber's database. Although each subscriber could receive updates from the publisher's feed, each browser client could not see any changes in their web page until they refreshed the page. This issue was solved using Web Sockets, which provide a connection for communication between the subscriber server and the browser clients currently viewing a page from the subscriber server, allowing the subscriber to send updated data that it received from the hub to each browser client. However, as the number of clients accessing the server increases, the amount of processing needed to be done by the Web Socket increases, slowing down access times substantially. To aid in this process, we used Redis, an in-memory hash, to store subscriber-client connection information, significantly reducing the read/write times for the Web Socket and allowing it to continue functioning quickly for a larger number of clients.

5. PROJECT ASSESSMENT

5.1. Faculty Assessment

After attending the final project demonstrations given by the students in senior seminar, CS faculty at VSU concluded that eight out of nine (or 89%) student groups completed their capstone projects with an "excellent" overall quality (see the detailed faculty assessments in Table 2). These projects not only met all the requirements listed in section 3.1 with elegant and professionally looking user interfaces, but also provided quite a few additional features such as live chatting, full-fledged searching, RSS based cross-site information sharing, and location based services, etc.

Table 2. Faculty Assessment

Assessment Item	**Agree**
Students have designed, implemented, and evaluated a computer-based system, process, component, or program to meet desired needs	87%
Students have demonstrated ability to use current techniques, skills, and tools necessary for computer practice	91%
Students have applied mathematical foundations, algorithmic principles, and computer science theory in the modelling and design of computer-based systems in a way that demonstrates comprehension of the trade-offs involved in design choices	90%

5.2. Student Assessment

The end-of-semester survey completed by all students in senior seminar indicated that 86% of them strongly agreed that "the all-in-cloud programming ecosystem used in the capstone project gave them a great exposure to what it means to work in a professional context" and all of them strongly agreed or agreed that "overall speaking, working in the capstone project better prepared them for a career in software development". Additionally, the survey asked students to vote for two services/products provided in the SPI ecosystem that benefited them the most. The result is shown in Figure 12 below where the number of student votes for a specific service/product is marked as its "popularity".

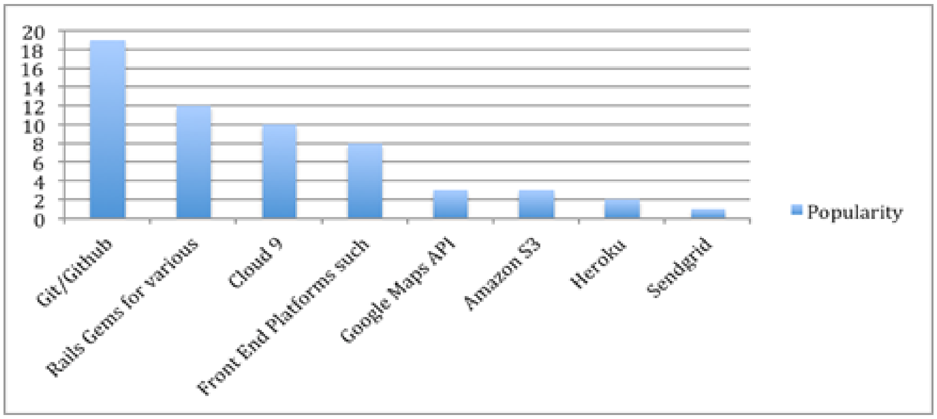


Figure 12. Cloud-based Services/Products that benefited the students the most

5.3. Issues

There were two major issues that were brought to our attention while we appreciated the agility and effectiveness of the project development in senior seminar thanks to the AIC-SPI ecosystem – security and inadequate testing.

Data and code security cloud computing in general is one of the top issues that everyone concerns. We are not exceptional. Two months after the instructor created buckets on Amazon S3 for storing images for his own Rails projects, hackers comprised his Amazon access keys. Fortunately, Amazon caught unauthorized activities and resolved the issue. Further investigations indicated that these access keys were leaked on Github where accidentally, they were made accessible to the public for a short period of time. As an alarm, the vulnerability of the AIC-SPI ecosystem was emphasized to the students in senior seminar before it had further widely spread.

A widely cited 2002 study prepared for NIST reported that 50 percent of software development budgets go to testing [6]. However, flaws in software still cost the U.S. economy $59.5 billion annually. Unfortunately, due to time constraints, in senior seminar, we didn't spend as much time as we should on software testing including unit testing, although it is such an important built-in component in Rails.

6. Conclusion and Future Work

Our experience in developing capstone application with the AIC-SPI ecosystem has been very positive. We have seen senior students in the capstone class voluntarily and comfortably use such an ecosystem in other projects. SPI and cloud computing in general give them unique opportunities and exposures to collaborative, distributed, and real-world practices that are prevalent in today's software development industry and community. The experience and competitive skills gained in CS 4900 will scale with students and enable them to collaborate with their peers, contribute to open source software projects, and eventually transfer their new knowledge to the work environment in the future. It also streamlines the teacher's tasks of grading student projects and giving lectures.

The proposed AIC-SPI ecosystem takes advantage of the intimate relationship that exists between the cloud technologies and CS courses [1]. Due to this unique attribute, it can penetrate into all layers of the Cloud and provide meaningful assistance for students in a wider range of CS classes.

Therefore, future works include expanding the adoption of the AIC-SPI ecosystem in other Computer Science classes where students' programming skills are emphasized.

As instructors in computer science departments we are preparing people to develop software. If testing is 50% of the effort, we are not properly preparing our students if we do not include software testing in the curriculum. Therefore, another area to work on in the future is to invest more time in teaching students software testing.

REFERENCES

[1] H. Rajaei and A. Aldakheel, "Cloud Computing in Computer Science and Engineering Education," in 2012 American Society for Engineering Education Annual Conference, San Antonio, TX, June 17-20, 2012.

[2] P. Kumar, S. Kommareddy, and N. Rani, "Effective Ways Cloud Computing Can Contribute to Education Success," Advanced Computing: An International Journal (ACIJ), Vol.4, No. 4, July 2013.

[3] R. Roggio, "Cloud Computing for Capstone Software Development Courses," 2011 Information Systems Educators Conference (ISECON) Proceedings v28-n1692, Wilmington, North Carolina, Nov 2011.

[4] O. Akin, F. Matthew, and D. Comfort, "The Impact and Challenges of Cloud Computing Adoption on Public Universities in Southwestern Nigeria," International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 5, No. 8, 2014.

[5] Z. Xu, "Distributed Student Software Project Management with Git," the 7th International Multi-Conference on Computing in the Global Information Technology (ICCGI 2012), p.p. 159-164, ISBN 978-1-61208-202-8, Venice, Italy, June 24-29, 2012.

[6] G. Tassey, "The Economic Impacts of Inadequate Infrastracture for Software Testing", NIST Planning Report 02-3, May 2002.

INTENTIONAL BLANK

WI-FI FINGERPRINT-BASED APPROACH TO SECURING THE CONNECTED VEHICLE AGAINST WIRELESS ATTACK

Hyeokchan Kwon, Sokjoon Lee and Byung-ho Chung

Electronics and Telecommunications Research Institute,
218 Gajeong-ro, Yoseong-gu, Daejeon, Republic of KOREA
{hckwon, junny, cbh}@etri.re.kr

ABSTRACT

In this paper, we present wifi fingerprint-based approach to securing the connected vehicle against wireless attack. In current connected vehicles such as Tesla EV, Mitsubishi outlander PHEV etc., there is a wi-fi access point on the vehicle to connect to the mobile device which has telematics apps installed. And generally the wi-fi access point is managed by the head unit system in the vehicle. Currently, the headunit in the vehicle utilizes white-list that contain MAC addresses of the pre-registered (i.e authorized) device. However, the white-list based mechanism cannot detect the device that forges its MAC address with authorized one. This paper presents security mechanism to detect rogue telematics device that has a spoofed (i.e, forged) MAC by analysing wi-fi fingerprint. We generate wi-fi fingerprint by analysing radio frequency features such as error vector magnitude (EVM), frequency offset, I/Q offset, sync correlation and so on. And we also utilizing distance information for improving detection ratio. The prototype of the proposed mechanism is implemented in this work, and we provide experimental results.

KEYWORDS

Connected Vehicle Security, Wireless Attack, Wi-Fi Fingerprint, Telematics Device Authentication

1. INTRODUCTION

Recently, various telematics apps for diagnostic the car, setting the configuration, locating the car, locking it remotely etc. are exist and they use wireless network such as wi-fi, Bluetooth etc. to connect to the vehicle. In current connected vehicles such as Tesla EV, Mitsubishi outlander PHEV etc., there is a wi-fi access point on the vehicle to connect to the mobile device which has telematics apps installed. And generally the wi-fi access point is managed by the head unit system in the vehicle.

In recent years, some hacking accidents have occurred with connected vehicles providing wi-fi access. For example, in this year, Mitsubishi outlander PHEV was hacked [1] by cracking wi-fi PSK (Pre-shared key) and analysing binary protocol using MITM (Man in the middle attack). In this case, the hackers were able to disable the theft alarm, unlock the car, turn the light, pop the window/jimmy, turns on pre-heating, pre-cooling and so on. For another example, the Tesla EV was also hacked [2] by using malicious wi-fi hotspot which is connected to a car's web browser. In this case, the hackers were able to remotely unlock the door, take over control of the dashboard

Natarajan Meghanathan et al. (Eds) : NETCOM, NCS, WiMoNe, GRAPH-HOC, SPM, CSEIT - 2016
pp. 211– 217, 2016. © CS & IT-CSCP 2016 DOI : 10.5121/csit.2016.61518

computer screen, open the door, move the seats and activate the indicators and windscreen wipers, as well as fold in the wing mirrors while the vehicle was in motion. And they were also able to take remote control of Tesla's brakes and door locks from 12 miles away.

In this paper, we present wi-fi fingerprint-based approach to securing the connected vehicle against wireless attack. Currently, with regard to wi-fi access, the head unit check MAC address of the device in order to determine whether the device is authorized or not. To do this, head unit use white-list which consists of MAC addresses of the pre-registered device. However, the white-list based mechanism cannot detect the device that forges its MAC address with authorized one. This paper presents security mechanism to detect rogue device that has a spoofed (i.e, forged) MAC by analysing wi-fi fingerprint. We generate wi-fi fingerprint by analysing radio frequency features such as error vector magnitude (EVM), frequency offset, I/Q offset, sync correlation and so on. And we also utilizing distance information for improving detection ratio. The prototype of the proposed mechanism with considering EVM as a radio frequency feature is implemented in this work, and we provide experimental results. So far, security research regard to security vulnerability/threat on connected vehicle has not been conducted.

The rest of the paper is organized as follows. The wi-fi fingerprint-based telematics device authentication mechanism and experiment result is described in section 2. Finally, conclusion is given in section 3.

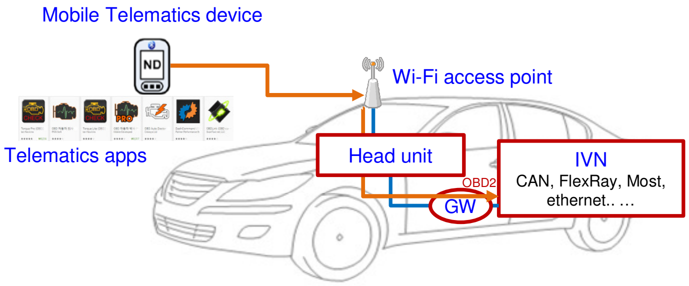


Figure 1. Wi-fi based connected vehicle

2. WI-FI FINGERPRINT-BASED TELEMATICS DEVICE AUTHENTICATION MECHANISM FOR CONNECTED VEHICLE

In order to authenticate telematics device, we utilizes the radio frequency features and distance information (i.e. RSSI). In the wi-fi fingerprint generation phase, the wi-fi access point on the vehicle collects and analysis wi-fi signals of the device by moving the physical location of the device and generates wi-fi fingerprint. In the verification phase, it estimates the distance from device to vehicle by using RSSI value, and it analyses wi-fi fingerprint data with the best nearby radio frequency features in current relative distance with the vehicle. Figure 2 shows the overall architecture of this mechanism.

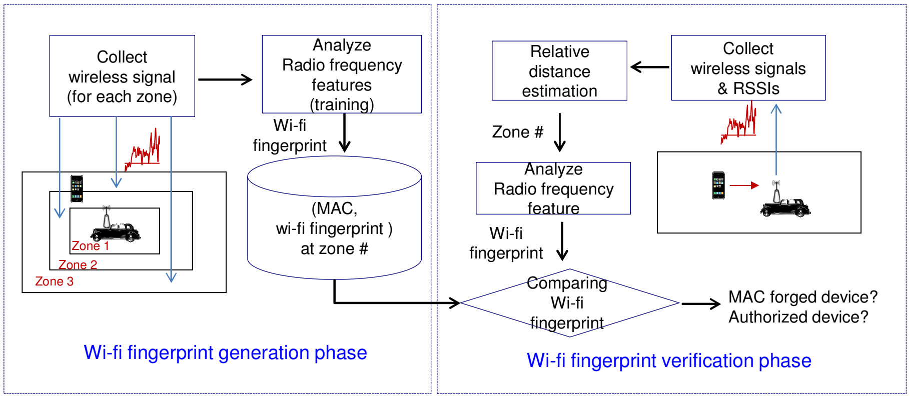

Collect wireless signal (for each zone)
Analyze Radio frequency features (training)
Wi-fi fingerprint
(MAC, wi-fi fingerprint) at zone #
Zone 1
Zone 2
Zone 3
Wi-fi fingerprint generation phase
Relative distance estimation
Collect wireless signals & RSSIs
Zone #
Analyze Radio frequency feature
Wi-fi fingerprint
Comparing Wi-fi fingerprint
MAC forged device? Authorized device?
Wi-fi fingerprint verification phase

Figure 2. Overall process of wi-fi fingerprint–based MAC verification for telematics device

2.1. Wi-fi fingerprint generation mechanism

In the Wi-fi fingerprint generation phase, the head unit registers MAC address of the authorized device. And it determines a wireless signal collection zone, and moves wireless device to the measurement point and generates wireless wi-fi signals. The head unit collects wi-fi signals from the authorized device, and then analyse them by machine learning algorithm such as K-NN, SVM and so on, and creates wi-fi fingerprint of the authorized device. In this paper, we use K-NNDD (K-Nearest Neighbour Data Description) [5] for training radio frequency of authorized devices. The relative distance and RF fingerprint information is stored in wi-fi fingerprint database.

In this paper, we applied error vector magnitude (EVM) as a RF feature. EVM is a vector magnitude difference between an ideal reference signal and measured signal. Figure 3 shows the concept of error vector magnitude (EVM) and mathematical formula for deriving EVM value.

$$EVM = \sqrt{Err_I{}^2 + Err_Q{}^2} \quad \text{, where} \begin{cases} Err_I = I_{reference} \cdot I_{measured} \\ Err_Q = Q_{reference} \cdot Q_{measured} \end{cases} \quad \text{, (formula 1)}$$

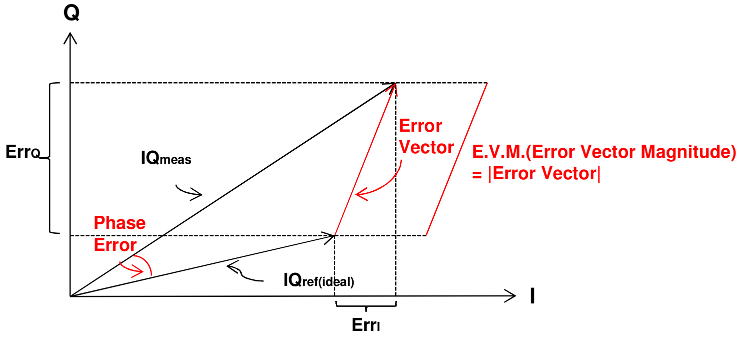


Figure 3. The concept of the Error Vector Magnitude

EVM is calculated by comparing the difference between measured signals with an ideal reference signals for determining the error vector. The EVM value is the root mean square value of the error vector over time at the instants of the symbol clock transitions. There are various reasons of mismatching measured signal with reference ideal signal such as hardware impairment, channel characteristics, noise at the receiver and modulation error. By using modulation error, we can identify particular wireless devices with different manufacturer or different wifi-chipset or even the same manufacturer/wifi-chipset.

2.2. Wi-fi fingerprint verification mechanism

In the Wi-fi fingerprint verification phase, the head unit collect and analyse RF signals of the device and extracts MAC address, RSSI and radio frequency features. And then it estimates the relative distance from the device to the vehicle by analysing the RSSI value. To calculate distance from RSSI values we used the following formula:

$$RSSI = -12.5Ln(d) - 36.25$$

The correction factors are derived through iterative experiments by minimizing the difference from the value by distance estimation algorithm with real distance. The notation d in this formula is a distance. The head unit then selects the radio frequency features having the highest P(radio frequency features | *d*) of the MAC of the device. P(radio frequency features | *d*) means a probability of radio frequency features in a given zone. Then it creates radio frequency signature from the radio frequency features by using machine learning algorithm. Then it determines whether the device having cloned MAC by comparing the wi-fi signature of the device with the device having same MAC in the database. In this paper, we use K-NN(K-Nearest Neighbor) algorithm for comparing measured radio frequency signature with reference radio frequency signature.

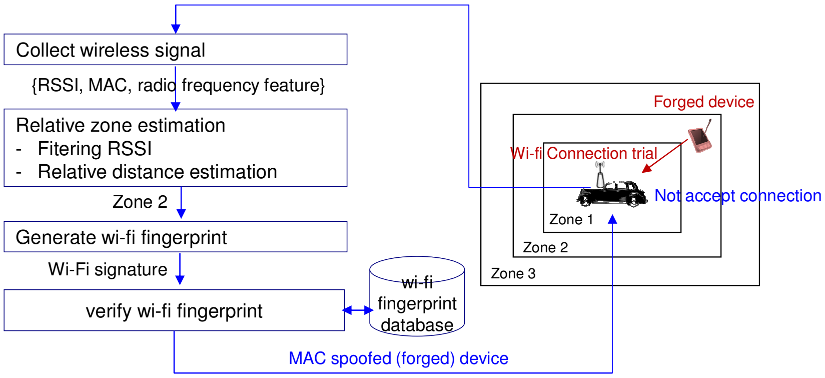


Figure 4. Detection process of rogue device

2.3. Experiments result

We developed hardware platform to collect wi-fi radio frequency signal and extract wireless features. Figure 5 shows the developed HW platform which can be installed to the head unit in the connected vehicle. This hardware platform includes Atheros 9380 WLAN chipset for monitoring wi-fi signals. We developed test platform with related user interface and dashboard, and we developed also wireless attack tool for evaluating developed system. Figure 6 shows the screenshots of our attacking tool to create cloned MAC. Figure 7 shows the UI of test platform for MAC spoofing device through wi-fi signature analysis and verification. Table 1 shows the experiment result.

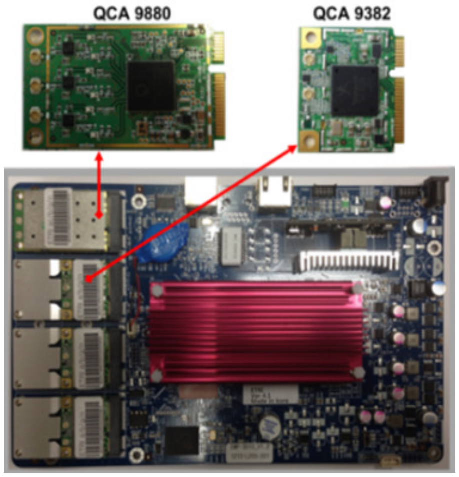
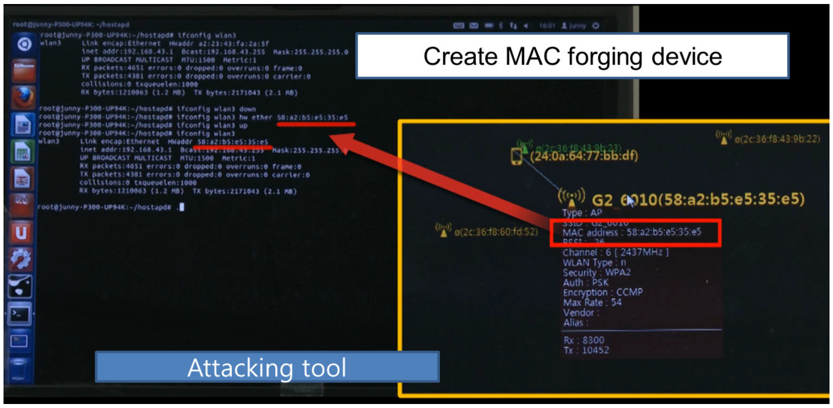


Figure 5. Prototype hardware platform which supporting wireless radio frequency feature extraction


Create MAC forging device

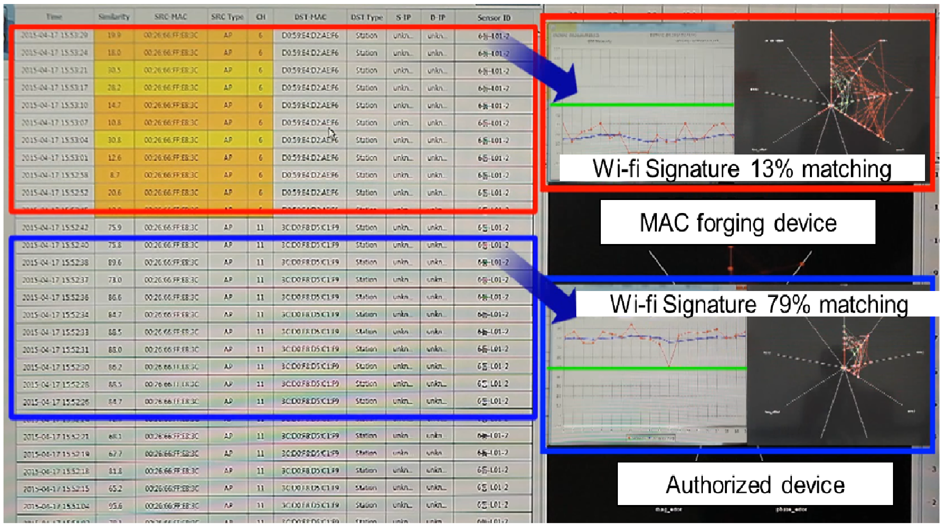
```
root@junny-P300-UP94K: ~/hostapd
root@junny-P300-UP94K:~/hostapd# ifconfig wlan3
wlan3     Link encap:Ethernet  HWaddr a2:23:43:fa:2a:5f
          inet addr:192.168.43.1  Bcast:192.168.43.255  Mask:255.255.255.0
          UP BROADCAST MULTICAST  MTU:1500  Metric:1
          RX packets:4651 errors:0 dropped:0 overruns:0 frame:0
          TX packets:4381 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:1210063 (1.2 MB)  TX bytes:2171043 (2.1 MB)

root@junny-P300-UP94K:~/hostapd# ifconfig wlan3 down
root@junny-P300-UP94K:~/hostapd# ifconfig wlan3 hw ether 58:a2:b5:e5:35:e5
root@junny-P300-UP94K:~/hostapd# ifconfig wlan3 up
root@junny-P300-UP94K:~/hostapd# ifconfig wlan3
wlan3     Link encap:Ethernet  HWaddr 58:a2:b5:e5:35:e5
          inet addr:192.168.43.1  Bcast:192.168.43.255  Mask:255.255.255.0
          UP BROADCAST MULTICAST  MTU:1500  Metric:1
          RX packets:4651 errors:0 dropped:0 overruns:0 frame:0
          TX packets:4381 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:1210063 (1.2 MB)  TX bytes:2171043 (2.1 MB)

root@junny-P300-UP94K:~/hostapd# .
```

Type : AP
SSID : G2_6010
MAC address : 58:a2:b5:e5:35:e5
RSSI : -26
Channel : 6 [2437MHz]
WLAN Type : n
Security : WPA2
Auth : PSK
Encryption : CCMP
Max Rate : 54
Vendor :
Alias :

Rx : 8300
Tx : 10452

Attacking tool

Figure 6. Snapshot of the attack tool, it shows the creation of MAC forging device



Time	Similarity	SRC-MAC	SRC Type	CH	DST-MAC	DST Type	S-IP	D-IP	Sensor ID
2015-04-17 15:53:29	19.9	00:26:66:FF:E8:3C	AP	6	D0:59:E4:D2:AE:F6	Station	unkn...	unkn...	6[illegible]-L01-2
2015-04-17 15:53:24	18.0	00:26:66:FF:E8:3C	AP	6	D0:59:E4:D2:AE:F6	Station	unkn...	unkn...	6[illegible]-L01-2
2015-04-17 15:53:21	30.5	00:26:66:FF:E8:3C	AP	6	D0:59:E4:D2:AE:F6	Station	unkn...	unkn...	6[illegible]-L01-2
2015-04-17 15:53:17	28.2	00:26:66:FF:E8:3C	AP	6	D0:59:E4:D2:AE:F6	Station	unkn...	unkn...	6[illegible]-L01-2
2015-04-17 15:53:10	14.7	00:26:66:FF:E8:3C	AP	6	D0:59:E4:D2:AE:F6	Station	unkn...	unkn...	6[illegible]-L01-2
2015-04-17 15:53:07	10.8	00:26:66:FF:E8:3C	AP	6	D0:59:E4:D2:AE:F6	Station	unkn...	unkn...	6[illegible]-L01-2
2015-04-17 15:53:04	30.8	00:26:66:FF:E8:3C	AP	6	D0:59:E4:D2:AE:F6	Station	unkn...	unkn...	6[illegible]-L01-2
2015-04-17 15:53:01	12.6	00:26:66:FF:E8:3C	AP	6	D0:59:E4:D2:AE:F6	Station	unkn...	unkn...	6[illegible]-L01-2
2015-04-17 15:52:58	8.7	00:26:66:FF:E8:3C	AP	6	D0:59:E4:D2:AE:F6	Station	unkn...	unkn...	6[illegible]-L01-2
2015-04-17 15:52:52	20.6	00:26:66:FF:E8:3C	AP	6	D0:59:E4:D2:AE:F6	Station	unkn...	unkn...	6[illegible]-L01-2
[illegible]	[illegible]	[illegible]	AP	6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2015-04-17 15:52:42	75.9	00:26:66:FF:E8:3C	AP	11	3C:D0:F8:D5:C1:F9	Station	unkn...	unkn...	6[illegible]-L01-2
2015-04-17 15:52:40	75.8	00:26:66:FF:E8:3C	AP	11	3C:D0:F8:D5:C1:F9	Station	unkn...	unkn...	6[illegible]-L01-2
2015-04-17 15:52:38	89.6	00:26:66:FF:E8:3C	AP	11	3C:D0:F8:D5:C1:F9	Station	unkn...	unkn...	6[illegible]-L01-2
2015-04-17 15:52:37	73.0	00:26:66:FF:E8:3C	AP	11	3C:D0:F8:D5:C1:F9	Station	unkn...	unkn...	6[illegible]-L01-2
2015-04-17 15:52:36	86.6	00:26:66:FF:E8:3C	AP	11	3C:D0:F8:D5:C1:F9	Station	unkn...	unkn...	6[illegible]-L01-2
2015-04-17 15:52:34	84.7	00:26:66:FF:E8:3C	AP	11	3C:D0:F8:D5:C1:F9	Station	unkn...	unkn...	6[illegible]-L01-2
2015-04-17 15:52:33	88.5	00:26:66:FF:E8:3C	AP	11	3C:D0:F8:D5:C1:F9	Station	unkn...	unkn...	6[illegible]-L01-2
2015-04-17 15:52:31	88.0	00:26:66:FF:E8:3C	AP	11	3C:D0:F8:D5:C1:F9	Station	unkn...	unkn...	6[illegible]-L01-2
2015-04-17 15:52:30	86.2	00:26:66:FF:E8:3C	AP	11	3C:D0:F8:D5:C1:F9	Station	unkn...	unkn...	6[illegible]-L01-2
2015-04-17 15:52:28	88.5	00:26:66:FF:E8:3C	AP	11	3C:D0:F8:D5:C1:F9	Station	unkn...	unkn...	6[illegible]-L01-2
2015-04-17 15:52:26	84.7	00:26:66:FF:E8:3C	AP	11	3C:D0:F8:D5:C1:F9	Station	unkn...	unkn...	6[illegible]-L01-2
[illegible]	[illegible]	[illegible]	AP	11	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2015-04-17 15:52:21	68.1	00:26:66:FF:E8:3C	AP	11	3C:D0:F8:D5:C1:F9	Station	unkn	unkn	6[illegible]-L01-2
2015-04-17 15:52:19	67.7	00:26:66:FF:E8:3C	AP	11	3C:D0:F8:D5:C1:F9	Station	unkn	unkn	6[illegible]-L01-2
2015-04-17 15:52:18	81.8	00:26:66:FF:E8:3C	AP	11	3C:D0:F8:D5:C1:F9	Station	unkn...	unkn...	6[illegible]-L01-2
2015-04-17 15:52:15	65.2	00:26:66:FF:E8:3C	AP	11	3C:D0:F8:D5:C1:F9	Station	unkn...	unkn...	6[illegible]-L01-2
2015-04-17 15:51:04	95.6	00:26:66:FF:E8:3C	AP	11	3C:D0:F8:D5:C1:F9	Station	unkn...	unkn...	6[illegible]-L01-2
[illegible]	[illegible]	[illegible]	AP	11	3C:D0:F8:D5:C1:F9	Station	unkn	unkn	[illegible]

Wi-fi Signature 13% matching

MAC forging device

Wi-fi Signature 79% matching

Authorized device

Figure 7. Snapshot of the test platform, it shows the MAC verification process in GUI

Table 1. Experiment result (FAR: False Accept Rate, FRR: False Reject Rate, EER: Equal Error Rate)

Test device	**threshold**	**FAR**	**FRR**	**EER**
Mobile device with different chipset (smart phone w/ Broadcom chipset, laptop computer w/ intel chipset, laptop computer w/ atheros chipset)	0.91	2.7%	0%	0.8%
Mobile device with same wi-fi chipset (iphone4s smart phone w/ broadcom's BCM 4330, iphone4 smart phone w/ broadcom's BCM 4329)	0.86	10.04%	0%	5.34%

3. CONCLUSIONS

In this paper, we present wifi fingerprint-based approach to securing the connected vehicle against wireless attack.

This paper provide the security mechanism to detect rogue device that has a spoofed (i.e, forged) MAC by analysing wi-fi fingerprint. We generate wi-fi fingerprint by analysing radio frequency features such as error vector magnitude (EVM). And we also utilizing distance information for improving detection ratio. The distance information is derived by RSSI value which in included in wi-fi signal. The prototype of the proposed mechanism with considering EVM as a radio frequency feature is implemented and we provide experimental results. The proposed mechanism analyse a characteristics of the wi-fi radio frequency signal of the device for detecting MAC spoofing device. We also developed wireless attacking tool and test platform with GUI.

In our experiments, the FAR is 2.7% in case that test mobile device has different chipsets and 10.4% in case that test mobile device has same chipsets. The detection rate should be improved when rogue device with a same manufacturers and wi-fi chipset with authorized one. Currently, we are designing the algorithm consider additional wi-fi radio frequency features such as IQ offset, sync correlation and so on.

ACKNOWLEDGEMENTS

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No.R-20160226-002842, Development of V2X Service Integrated Security Technology for Autonomous Driving Vehicle)

REFERENCES

[1] Hacking the Mitsubishi Outlander PHEV hybrid, https://www.pentestpartners.com/blog/hacking-the-mitsubishi-outlander-phev-hybrid-suv/, Pen test partners, 2016

[2] Hackers take Remote Control of Tesla's Brakes and Door locks from 12 Miles Away, http://thehackernews.com/2016/09/hack-tesla-autopilot.html, The Hacker News, 2016

[3] Hao, Peng, "Wireless Device Authentication Techniques Using Physical-Layer Device Fingerprint" (2015). Electronic Thesis and Dissertation Repository. Paper 3440. Western university, http://ir.lib.uwo.ca/etd/3440

[4] H. Kwon, G.An, S.H.Kim and B.H.Chung, "Detecting cloned devices in wireless network using RSSI and RF Features", ICONI, Dec., 2014

[5] J. Son and S. Kim, "kNNDD-based One-Class Classification by Nonparametric Density Estimation," Journal of the Korean Institute of Industrial Engineers, Vol. 38, No. 3, pp. 191-197, Sep. 2012.

[6] JP Hubaux, S Capkun, J Luo, The security and privacy of smart vehicles, IEEE Security & Privacy Magazine, 2004

[7] AirTight Patent, Method and system for monitoring a selected region of an airspace associated with local area networks of computing devices, Patent# US 7,002,943 Feb, 2006

[8] Y. Shi and Michael A. Jensen, Improved Radiometric Identification of Wireless Devices Using MIMO Transmission, IEEE Transactions on Information Forensics and Security, Dec. 2011

[9] R. Beyah and A. Venkataramen, Rogue-Access-Point Detection - Challenges, Solutions, and Future Directions, IEEE Security & Privacy, vol.9, issue 5, pp.56-61 (2011)

[10] Agilent 8 Hints for Making and Interpreting EVM Measurements, Agilent Technologies, 2005

AUTHORS

Hyeokchan Kwon

Received PhD degree in computer science from Chungnam National University in 2001. Since 2001, he is currently a principal researcher in electronics and telecommunications research institute (ETRI) in Korea. His research interests include automotive security, wireless intrusion prevention system and IoT security, etc.

Sokjoon Lee

Received MS degree in computer engineering from Seoul National University in 2000. Since 2000, he is currently a principal researcher in electronics and telecommunications research institute (ETRI) in Korea. His research interests include cryptographic protocol and ICT-Physical convergence security such as medical security, automotive security, etc.

Byung-ho Chung

Received PhD degree in computer science from Chungnam National University in 2004. He joined Agency for Defense Development (ADD) in 1988 where he was a senior researcher for 12 years. Since 2000, he is currently a principal researcher in electronics and telecommunications research institute (ETRI) in Korea. His research interests include automotive security, medical security, wireless security, multimedia security, etc.

AUTHOR INDEX