David C. Wyld
Jan Zizka (Eds)

Computer Science & Information Technology

4th International Conference on Computer Networks & Data Communications (CNDC-2017)
December 23~24, 2017, Sydney, Australia

AIRCC Publishing Corporation

Volume Editors

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

Jan Zizka,
Mendel University in Brno, Czech Republic
E-mail: zizka.jan@gmail.com

ISSN: 2231 - 5403
ISBN: 978-1-921987-76-2
DOI : 10.5121/csit.2017.71601 - 10.5121/csit.2017.71610

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

Preface

The 4th International Conference on Computer Networks & Data Communications (CNDC-2017) was held in Sydney, Australia, during December 23~24, 2017. The 7th International Conference on Digital Image Processing and Pattern Recognition (DPPR-2017), The 7th International Conference on Artificial Intelligence, Soft Computing and Application (AIAA-2017), The 4th International Conference on Wireless and Mobile Network (WiMNET-2017) and The 7th International Conference on Advances in Computing and Information Technology (ACITY-2017) was collocated with The 4th International Conference on Computer Networks & Data Communications (CNDC-2017). 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 CNDC-2017, DPPR-2017, AIAA-2017, WiMNET-2017, ACITY-2017 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, CNDC-2017, DPPR-2017, AIAA-2017, WiMNET-2017, ACITY-2017 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 CNDC-2017, DPPR-2017, AIAA-2017, WiMNET-2017, ACITY-2017.

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.

David C. Wyld
Jan Zizka

Organization

General Chair

David C. Wyld	Southeastern Louisisna University, USA
Jan Zizka	Mendel University in Brno, Czech Republic

Program Committee Members

Aalya Alajaji	Prince Sultan University, Saudi Arabia
Abdalah Rababah	Jordan University of Science and Technology, Jordan
Abdulhamit Subasi	Effat University, Saudi Arabia
Adnan A. Rawashdeh	Yarmouk University, Jordan.
Aihua Mao	South China University of Technology, China
Ajao Lukman Adewale	Federal University of Technology, Nigeria
Amine Laghrib	Faculté des Sciences Beni-Mellal, Marocco
Amir Rastegarnia	University of Malayer, Iran
Amizah Malip	University of Malaya, Malaysia
Antonia Plerou	Ionian University, Greece
Arifuzzman A K M	University of Alabama at Birmingham, USA
Asma Ayed Al Drees	King Khalid University, Saudi Arabia
Baver Okutmustur	Middfle East Technical University, Turkey
Boukenadil Bahidja	University of Tlemcen, Algeria
Carlo Sau	Università di Cagliari, Italy
Chaker LARABI	Université de Poitiers , France
Chuanzong Zhang	Aalborg University, Denmark
Dac-Nhuong Le	Haiphong University, Vietnam
Dalel BOUSLIMI	Institut Mines- Telecom, France
Dimitris Kontoudis	University of Macedonia, Greece
Dongpo Xu	Northeast Normal University, China
Duan Keqing	Wuhan Early Warning Academy, China
Efthimios Alepis	University of Piraeus, Greece
Erman Cakit	Aksaray University, Turkey
Fabio Gasparetti	Roma Tre University, Italy
Fatma Outay	Zayed University DXB, UAE
Fernando Bobillo	University of Zaragoza, Spain
Hacer Yalim Keles	Ankara University, Turkey
Hamid Alasadi	Basra University, Iraq
Hamzeh Khalili	Universitat Politecnica de Catalunya (UPC), Spain
Hani Bani-Salameh	The Hashemite University, Jordan
Hanming Fang	Logistical Engineering University, China
Hao-En Chueh	Yuanpei University Of Medical Technology, Taiwan
Hari Krishna Garg	National University of Singapore, Singapore
Hasnaoui Salem	Al-Manar University, Tunisia
Hassan Ugail	University of Bradford, UK
Islam Atef	Faculty of Engineering-ALexandria, Egypt

Ismail Shahin	University of Sharjah, UAE
Issac Niwas Swamidoss	Nanyang Technological University, Singapore
Jamel Hattay	University of Tunis El Manar,Tunisia
Joao Gama	University of Porto, Portugal
Jose Gonzalez	University of A Coruña, Spain
Jose Vicente Berna Martinez	University of Alicante, Spain
Kamlesh Dutta	NIT Hamirpur, India
Khider Shoaib	Dalian University of Technology China, Dalian
Kishan Rao Kalitkar	Vaagdevi Group of Technical Institutions, India
Lakshmi Rajamani	Osmania University, India
Lei ZHANG	University of Surrey, UK
Lim Seng Poh	Wawasan Open University, Malaysia
M.Sadanandam	Kakatiya University, India
Madya Dr. Yuhanis binti Yusof	Universiti Utara Malaysia, Malaysia
Mike Turi	California State University-Fullerton, USA
Mimoun Hamdi	Ecole Nationale d'Ingenieurs de Tunis, Tunisia
Mohamad Badra	Zayed University, Dubai, UAE
Mohamad heidari	Islamic Azad University, Iran
Mohamed Arezki MELLAL	M'Hamed Bougara University, Boumerdès
Mohammad Amin Shayegan	Islamic Azad University, Iran
Mohammad Siraj	King Saud University, Saudi Arabia
Mohammadreza Balouchestani	Indiana Purdue Fort Wayne University,USA
Mohammed Fatehy Soliman	Suez Canal University, Egypt
Nayeem Ahmad Khan	University Malaysia Sarawak, Malaysia
Ognjen Kuljaca	Brodarski Institute, Croatia
Pasi Luukka	Lappeenranta University of Technology Finland
Philippe Thomas	Université de Lorraine, France
Philomina Simon	University of Kerala, India
Pratyay Kuila	National Institute of Technology, India
Riccardo Pecori	eCampus University, Italy
Rocío Pérez de Prado	University of Jaén, Spain
Ronghuo Zheng	University of Texas, United States
Sanjay K. Dwivedi	Babasaheb Bhimrao Ambedkar University, India
Sanjay Sharma	University of london, UK
Sayeed Tavakoli	University of Sistan and Baluchestan, Iran
Sherif Rashad	Florida Polytechnic University, USA
Shifei Ding	China university of mining and technology, China
TAN, TSE GUAN	Universiti Malaysia Kelantan, Malaysia
Tzung-Pei Hong	National University of Kaohsiung, Taiwan
Upasna Vishnoi	Sr. Digital Design Engineer,USA
Veton Kepuska	Florida Institute of Technology, Australia
Vicki Allan Utah	State university USA
Victor Banos	Technical University of Catalonia, Spain
Yao-Nan Lien	Asia University, Taiwan
Yuan Zhuang	Bluvision Inc, USA
Yue Cao	Northumbria University, UK
Yuhanis binti Yusof	Universiti Utara Malaysia, Malaysia
Yuriy Mishchenko	Izmir University of Economics,Turkey

Technically Sponsored by

Computer Science & Information Technology Community (CSITC)

Networks & Communications Community (NCC)

Digital Signal & Image Processing Community (DSIPC)

Organized By

Academy & Industry Research Collaboration Center (AIRCC)

TABLE OF CONTENTS

4th International Conference on Computer Networks & Data Communications (CNDC-2017)

7th International Conference on Digital Image Processing and Pattern Recognition (DPPR-2017)

7th International Conference on Artificial Intelligence, Soft Computing and Application (AIAA-2017)

4th International Conference on Wireless and Mobile Network (WiMNET-2017)

7th International Conference on Advances in Computing and Information Technology (ACITY-2017)

AN UNMANNED AERIAL VEHICLE BASED FOREST PATROL SYSTEM

Songsheng LI

Department of Computer Engineering,
Guangdong College of Business and Technology, Zhaoqing, China

ABSTRACT

Although technology has advanced at breathless pace, wildfire still bursts out every year all over the world. There are various wildfire monitoring system employed in diferrent countries, most of them are depended on photos or videos to identify features of wildfire, after wildfire happened. Delay of confirmation is diverse based on used techonologies. An autonomous forest patrol system by Unmanned Aerial Vehicle(UAV) is presented in this paper, which try to employ the fashinable UAV to patrol in forest and collect environmental data in order to monitor and predict wildfire before it really erupts. From limited practical data collected, the monitoring data such as temperature and humidity are effective to reflect the real situation, prediction requires more data and time to prove.

KEYWORDS

UAV, Bluetooth, Forest, Wildfire, Patrol

1. INTRODUCTION

Autonomous forest patrol is a classical topic as forest occupy almost 31% surface of the earth in 2012 according to EARTH POLICY INSTITUTE from Rutgers University [1]. In general, forests are not under surveillance, it is too wide for human to cover even parts of it before. Modern invention such as airplane and robot arouse possibility of forest exploration. Satellites [2] are employed to monitor wildfire, but they only help after burning. PETER KOURTZ [3] wrote the paper named "A VISUAL AIRBORNE FOREST FIRE DETECTION PATROL ROUTE PLANNING SYSTEM" when working in Canadian Forest Service in 1973. The scene of the paper is that a pilot flies an airplane in a planned route, observes the forest on the ground for sign of wildfire. Route is significant because of budget restriction and time limitation if wildfire happens; it is planed according to fire occurrence pattern from historical data. As there is no real computer at that time, all the test was done almost by hand. Comparably, UAV of today is tiny, flexible and intelligent, should be used effectively. Luis Merino et al. proposed an Unmanned Aircraft System for wildfire monitoring [4], which employ several aerial vehicles and a central station, vehicles collaborate and automatically obtain wildfire information by means of on-board infrared or visual camera. Every vehicle equips with local perception and software component; they make decision on the real-time progress of the fire front shape, applying techniques of prediction model, fire contours extraction, image vibrations eliminating, feature matching etc. Their experiment showed that fire front shape is adequately extracted from the images, and estimation of the fire front shape was successful, the system is applicable. At least two issues should be stressed, one is scale, their experiment employed 3 vehicles, but the coverage area is limited to visual range of people. As more than one vehicle is used, collaboration is another problem, the system need visual photos from different vehicles, so how to organize their

David C. Wyld et al. (Eds) : CNDC, DPPR, AIAA, WiMNET, ACITY - 2017
pp. 01– 10, 2017. © CS & IT-CSCP 2017 DOI : 10.5121/csit.2017.71601

synchronization and path to get useful vision is key factor of their success. It will turn out to be very complex if more than one wildfire bursts.

M. Hefeeda and M. Bagheri present the design and evaluation of a wireless sensor network (WSN) for early detection of forest fires [5]. They focus on the Fire Weather Index (FWI) from the Canadian Forest Service, treat the forest fire detection problem as a k-coverage problem in WSN. In addition, they present a simple data aggregation scheme to prolong the network lifetime by only delivering the data of interest to the application. Their simulation way is far from practice, first is deployment, with increase of k, the massive nodes are need based on the algorithm, and nodes die quickly.

Paper [6] describes a hierarchical WSN for early fire detection in risky areas. Advantage it elaborates includes 1) no communication network in advance, 2) time from the GPS module, 3) not just fire early detection, environment monitoring too, 4) no cameras/images, 5) middleware layer to share data. There are some questions left of the paper, first is the time synchronization, which is not described how to synchronize time from CNs to SNs, and relevant energy consumed. Second is battery of SNs, it was said that battery with a capacity of 600mAh can last at least 2 years without prove.

Habiboglu et al. proposed a real-time video smoke detection system that uses correlation descriptors with an SVM classifier [7]. They use temporally extended correlation matrices to combine color, spatial and temporal information together in the decision process. They proved that the proposed method is computationally efficient. Duration for the algorithm to identify the wildfire is the most crucial index for wildfire detection. The algorithm needs three stages, 1) slow moving object detection in video, 2) smoke-colored region detection, 3) correlation based classification. So time is related to performance of camera and computer.

This paper [8] presents the system architecture, hardware and software framework of WSN based wildfire monitoring system. They claimed main contribution of the paper is the design can meet the goal of reactivity and reliability, robustness and network lifetime. Specifically, reactivity depends on threshold, reliability needs diagnostic phase and link quality based routing protocol, robustness achieves with water-and-fire-proof boxes, network lifetime is extended by a MAC protocol named SMAC and dynamic power management. Left issues are listed as 1）balance of coverage and cost, 2) location information of data is not solved, 3) how will the mobile station work.

Comparison in [9] summarizes four main ways, human based observation, satellite system, optical cameras and WSNs, to forest fire detection in seven aspects including cost, efficiency and practicality, faulty alarms repetition, fire localizing accuracy, detection delay, small fire behavior information, can be used for other purposes. They all have advantage and disvantage, but optical cameras and WSN are relatively preferred.

Titled “Use of Remote Sensing in Wildfire Management” [10], the chapter elaborates wildfire management in a total different way for pre-fire and post-fire conditions monitoring, which is space-borne remotely sensed imagery, consist of optical remote sensing, thermal infrared remote sensing and radar remote sensing. Authors admit that further studies are required to establish models to estimate fuel moisture, and further to assess optical and thermal infrared images, radar images for monitoring pre-fire conditions too. Other required input for monitoring fire danger, like wind parameters, cannot be derived from remote sensing data. And geo-stationary satellites provide a lower revisit time (1 to 2 daily passes) so it is not ideal for real-time monitoring of wildfire.

Lina Tang and Guofan Shao reviewed applications of drone in forestry research and practice [11]. They first distinguished drone remote sensing from crewed aircraft remote sensing and satellite remote sensing then some projects employed drone successfully, such as surveying, mapping, measuring, tracking of forest situation, and supporting intensive forest management etc. They agree that drones are very susceptible to weather and human-related accident, and multidisciplinary collaborations to promote the standardization of drone is iminent.

We present a localization framework of WSN [12] [13] using dynamic path of mobile beacon (DPMB). The imaged scenario of the application is that an unmanned vehicle (UV) such as UAV or a driverless car, moves on the field where massive sensor nodes are deployed, its moving path will be decided dynamically by information from nodes. UV is the master of algorithm behind, it starts with first triple positions to ensure all nodes inside the triangle be converted to reference nodes, then it walks on vertex of equilateral triangle, in every position, “flood 3” employs reference to localize more nodes, then “Try all 2” checks nodes only received 2 messages based on their localized neighbors, and those two steps are repeated until no new nodes are localized, then UV moves to next position. we adapted online Reinforcement Learning (RL) as brain of UV to drive UV on the field to perform localization. According to the principle of exploration or exploitation, UV could regret the first choice and have a second chance to localize more nodes [14]. An alternative metric, direction, specifically, Vector Cosine Similarity (VCS) instead of distance is employed when decision on neighbours [15]. Nodes are grouped and grouped weight is calculated, which UV is driven by. Simulation proves that algorithms are lightweight and effective with help of UAV.

In this paper, the forest patrol system is implemented by UAV, observation points (OP) are distributed in forest, their locations are marked when installation, UAV fly to them according to the locations and receive data by Bluetooth. Before Bluetooth 4.1 [16], a Bluetooth Low Energy (BLE) device is either master or slave, cannot be both. This is how BLE used in the system, a UAV with a BLE data agent (BLEDA) acts as a mobile central, collects data from peripheral, which is BLE data collector (BLEDC) in OPs.

2. FOREST PATROL SYSTEM

The proposed forest patrol system is based on the currently popular UAV and combined with BLE to implement automatic forest patrol and predict the possible eruption of wildfire.

2.1. System Introduction

The system is displayed in figure 1, there are some OPs are selected in the surveilled forest; UAV stops at station usually and flies along a fixed and scheduled flight route which covers all the OPs and starts from station and ends at station. All the detail about OP, UAV and station are described in the following.

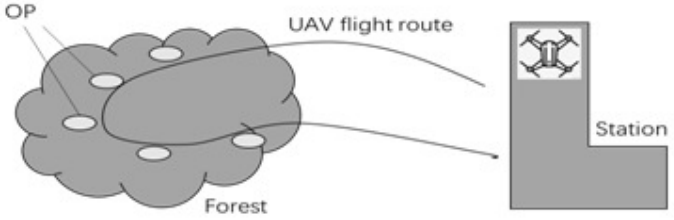


Figure 1. Introduction

2.1.1. Selection of OP

OP should be chosen in the middle of the surveilled forest, where that part of forest is important or vulnerable, then a BLE data collector (BLEDC) attached with a solar panel will be installed, which is a box with sensors embedded in surface. It will collect various sensor data continuously and dynamically, and report to UAV regularly. GPS information of the OP will be recorded when installation.

2.1.2. UAV Collecting Data

UAV equipped with a BLEDA will fly to OPs according to the GPS information in storage, which are recorded at installation stage. BLEDA will connect to BLEDCs, receive collected sensor data, return to station, and report data to cloud database. This is the routine task of UAV in the system. Two features of Intelligent Flight Modes [17] of UAV are very helpful for our system, one is waypoints, they are multiple GPS points which are set in control part of UAV. If they are set, the UAV will automatically fly to them, for our system, those points are OPs. The other is point of interest (POI), if a point is set in the control of UAV as POI, UAV will continuously circle around it for information.

2.1.3. Station

Station is the base for UAV, it provides resources the system requires, which include database server, Wi-Fi router, recharge for UAV and BLEDA, etc. There will be shelter for UAV and staffs.

Two key components in the system are BLEDC and BLEDA, they are explained in the following.

2.2. BLEDC

BLEDC is responsible for collecting sensor data and reporting to UAV, whose structure is shown in Figure 2. As it is installed in severe wild environment, its structure is strictly specified.

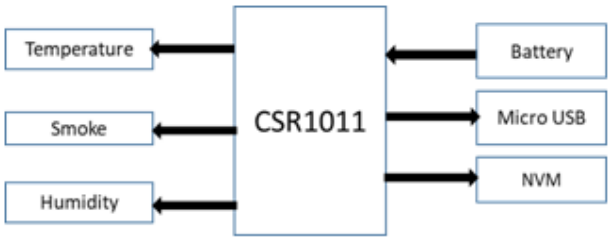


Figure 2. BLEDC

Three main sensors, temperature, humidity and smoke detector are integrated to IO of CSR1011 which is a decent single-mode BLE product from CSR. It collects data from those sensors dynamically, attaches timestamp, saves to non-volatile memory (NVM) and reports to BLEDA when it comes. Some specific details:

BLEDC plays role of GATT server in BLE architecture. According to BLE standard [18], as a GATT server, BLEDC must setup at least one Service with Universally Unique Identifier (UUID) and inside this Service there is at least one Characteristic with UUID too. Different BLE devices can be identified by this specific Service UUID. BLEDA use this UUID to find BLEDC.

Cyclic storage is designed for the application, for example, 512 bytes are applied from NVM, a head pointer for read and a tail point to write are defined, when head reach tail, no new data is available, when tail reach head, no more space available. This is the flexible and general way in practice.

Dynamical data means recording data base on threshold which is a value between normal and abnormal. Data beyond threshold will be recorded more frequently than normal. In such way significant data are recorded and normal data could be ignored, as a result, saving power and memory of BLEDC.

CSR1011 is designed by consideration of power exhaust, only a small 3v button battery CS2032 can provide energy of months, but for the system solar panel is a better choice if it is feasible, in other words, if the OP can provide enough sunshine for the panel.

In general, BLEDC only executes slow advertising based on schedule, but in special situation such as value from sensor change dramatically, it will keep slow adverting to open possibility for connection from BLEDA in order to report emergence on time.

Taking the wild bad climate and environment into consideration, the whole equipment adopts engineering plastics and sensors are embedded on the surface to promise them expose to the true environment, and all the related materials are chosen strictly to meet the industry requirement. No matter solar panel is used or not, save energy is the first principle as overcast could last long enough to exhaust solar power.

2.3. BLEDA

It is carried by UAV to approach OP to receive data from BLEDC. Its structure is shown in Figure 3. The core of BLEDA is ESP32, which is a Wi-Fi and Bluetooth combo chip, all the data reported from BLEDCs will be uploaded to database on Station by Wi-Fi. A GPS chip is integrated to the equipment to provide accurate time and location information. Storage employs the similar circular management way as BLEDC. Power supply in BLEDA will be very flexible as UAV can return to Station to recharge. LED applies different colors and flash to represent various working situation. Long press of button will power on or off the equipment and short press will change mode of BLEDA. There are two working modes designed for BLEDA, one is setup mode, the other is client mode.

Figure 3. BLEDA

The purpose of setup mode is to check the GPS information of OPs. UAV carries BLEDA and flies on the route of patrol to Ops after all the GPS information of Ops are set to UAV as waypoints. When flying, BLEDA starts BLE and search for BLE device which owns Service with UUID of this system, if BLEDA is connected and reports information, Bluetooth address and GPS information of every BLEDA are saved into NVM.

The client is the general mode of BLEDA. It is called client because BLEDC is defined as GATT Server in Bluetooth 4 and BLEDA is GATT client. Its main task is receiving data from BLEDC. In flight, BLEDA keeps tracking GPS information, compare it with GPS information of BLEDCs from NVM, if they are similar or close, then open BLE and connect to the BLEDC and receive data and save them to NVM.

2.4. System Setup Procedure

The procedure is shown in Figure 4. The flow chart explains itself and the detail of Bluetooth communication is hidden. Slow advertising is a way for BLEDC to notice BLEDA that it is waiting for connection. Compared with its counterpart, fast advertising, it saves energy, but it is a waste of energy if it keeps advertising. This could be a one-time effort if anything goes well. This procedure makes sure that the information of BLEDCs is consistent with the original selection of Ops. Application in Station server will compare the information BLEDA collected with records of Ops, if any difference emerges, all human error and performance of UAV should be ruled out first, then coverage range of BLE should be taken into consideration as it can be affected by environment remarkably.

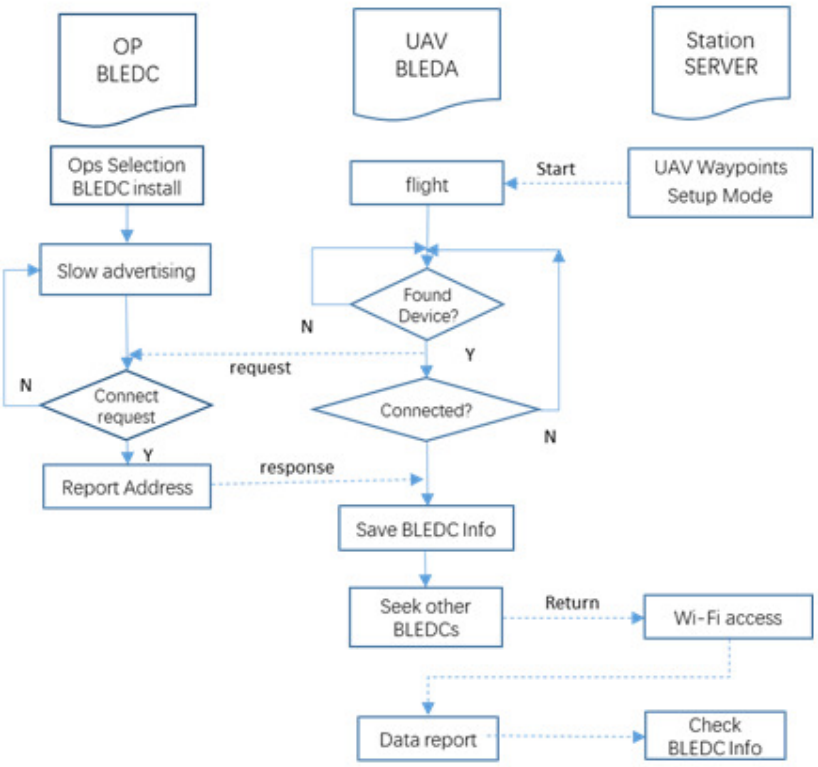


Figure 4. Setup procedure

2.5. Routine Procedure

Compared to setup procedure, the routine deserves more elaboration, which is displayed in Figure 5. The significance of time synchronization, scheduled slow advertising and exception handle are explained below.

2.5.1. Time Synchronization

System clock of CSR1011 is based on crystal oscillator which is not absolutely accurate, its error will be larger with time last longer, so it should be corrected with accurate time from GPS. So a significant step of routine procedure is time synchronization, which is happened right after BLEDA connected to BLEDC, BLEDC receives time by the way of concerted protocol, updates its own time which will be used from then on, and uploads sensor data to BLEDA. So the timestamp of data from first routine could be deflected.

2.5.2. Scheduled Slow Advertising

In routine procedure, scheduled slow advertising is employed instead of slow adverting which is employed in setup procedure. The former is scheduled, so it only slow advertising in certain time with certain interval, and the latter will always advertise in a slow way. The necessity of the former is power efficiency which is key topic in WSN, it is possible that cloudy and rainy weather could last long then power from solar panel exhausts.

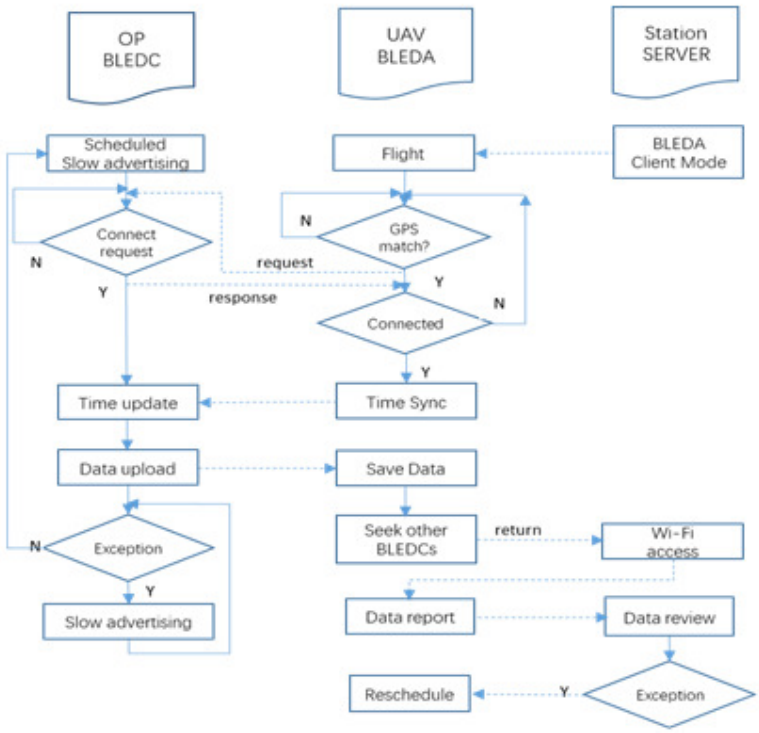


Figure 5. Routine procedure

As waypoints are applied in UAV, the path and distance is relatively fixed, and the flying speed is stable too, so it is possible that total duration of one routine procedure is countable. For instance, if UAV is scheduled to fly once an hour, duration for UAV flies from Station to the targeted forest is 5 minutes, and duration for UAV to collect all the data from BLEDCs is 10 minutes, so if BLEDCs are scheduled to start slow advertising once an hour at 5 minutes past the hour, and last 10 minutes, and once connected, not advertising in the same hour. If the schedules fulfill the requirement of the procedure, it is satisfactory and saves energy. The accurate advertising time synchronization will be developed in the future, for now, coarse schedule is enough for the system.

2.5.3. Exception Handle

Station exists as a support center, it provides Wi-Fi router for UAV to access to Internet, so all the collected data from BLEDCs of every routine procedure will be uploaded to data server. As new data arrive, application in server reviews the data in a fast way first, the application checks if the degree of deviation of various sensor data using their threshold as reference, then increase the frequency of flight or start an emergence flight immediately depending on the degree of exception, in order to catch all the possible issue of surveilled forest. Similarly, BLEDC has the same mechanism to check sensor data, if emergence appears, it will keep slow advertising and waiting for connection to report data until the data return to normal and its advertising mode goes back to scheduled slow advertising. By this way, possible problem of the surveilled forest is not missed, the effectiveness of the system is kept.

3. EXPERIMENTS

Prototype of BLEDC is ready, which is showed in figure 6. CSR1011 is the core of BLEDC, temperature sensor is TMP112, SHT20 is alternative option as it is a combination of temperature and humidity sensors. In figure 6, a CSR1011 is connected with two TMP112s and one SHT20, and powered by a tiny CR2032 button battery. The choice of smoke dectctor is still under consideration, the current prototype is used to check feasibility of the system. BLEDA is under development, but it can be replaced by an ordinary mobile phone which is equipped with Bluetooth 4 and GPS. An Android mobile is used in experiments as it fullfills the hardware requirements and an APP is developed to implement all the procedures of BLEDA, then it is tied with camera of UAV to work as BLEDA. Three prototypes of BLEDC are hung on trees for several days, distance between each two of them is more than 100m to avoid confusion of GPS information. Then their GPSs are set to DJI Phantom3 as waypoints, and each waypoint is set as point of interest, so DJI can fly to those positions and circle around and receive data. Data of different points are saved in mobile storage as seperate files, so they can ba accessed and evaluated later.

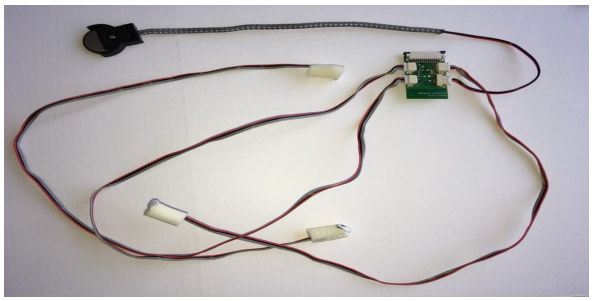


Figure 6. BLEDC prototype

Two sample data are displayed in Figure 7 and Figure 8. The data in Figure 7 is from about 20:00 of 09/21 to 18:00 of 09/22, almost a whole day of typical summer. Temperature is expressed as Celsius and humidity as percentage. In about 7am, humidity reaches its highest and temperature in its lowest, in contrast, in midday, about 13:00, temperature in its highest, 55℃ and humidity in its lowest. At other time, changes are accordant with natural law, temperature and humidity develop up and down alternatively. Figure 8 is the data of three days of rainy in a row. From afternoon of the first day, it starts raining, humidity rises and temperature drops drastically. Humidity keeps in high situation, even in the midday, the lowest percentage value of humidity is still higher than the highest ℃ value of temperature. The results are highly accordant with the real weather, which confirm that feasibility of the system.

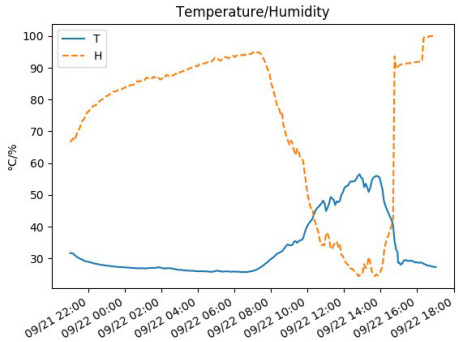


Figure 7. Normal summer day

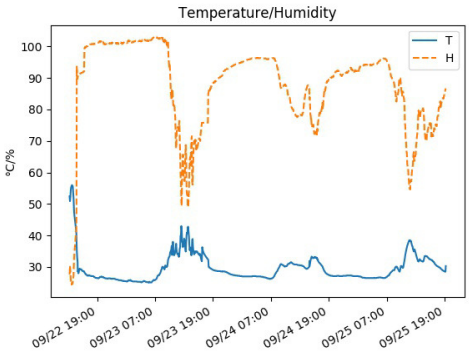


Figure 8. Three rainy days

4. CONCLUSIONS

UAV acts as a forest patrol worker, repeats the process again and again without fatigue to collect sensor data. Not only data from sensors, but also photos and video from action camera which is usually equipped in UAV, the combination will provide a better profile of the surveilled forest. Such a system with features of automatic, economical, one-time investment and long-term benefit, can find hidden danger of wildfire in advance, if measures can be taken before explosion of wildfire, the system is successful.

The system heavily depends on UAV which is vulnerable of severe weather, such as storm, but if the purpose of patrol is mornitoring wildfire, it is not a big issue, as wildfire usually happens in sunny hot summer not in wet days. The other weakness of UAV is its duration of flight and recharge, so the area of surveiled forest should be accordant to the ability of UAV and the schedule of flight should be planed in consideration of its battery volume.

In the future, all the hardwares should be designed resonably. First, smoke detector could be either photoelectric, which detect smoke optically, or ionization, smoke detected by physical process, no matter which one is selected, it has to be integrated to IO port of CSR1011. The current using mobile phone is much higher than the designed plan of BLEDA in cost. The third choice of BLEDA is no new hardware. As all the UAV manufactures are opening their API for developers, firmware of UAV could be adapted to implement the control and procedure of BLEDA. It will be a big save of cost. The ultimate goal of the system is prediction of wildfire which depends deeply on big data. If cooperation with local forest administaration department achieves, more and more data accumulate, a prediction pattern will emerge by data analysis and/or machine learning.

REFERENCES

[1] http://www.earth-policy.org/indicators/C56/forests_2012

[2] http://www.esri.com/services/disaster-response/wildfire

[3] A visual airborne forest fire detection patrol route planning system. (1973). Kourtz, P.H. Canadian Forestry Service, Forest Fire Research Institute, Ottawa, Ontario. Information Report FF-X-45. 34 p.

[4] An Unmanned Aircraft System for Automatic Forest Fire Monitoring and Measurement, (2012), Journal of Intelligent & Robotic Systems, Volume 65, Number 1-4, Page 533, Luis Merino, Fernando Caballero, J. Ramiro Martínez-de-Dios, Iván Maza, Aníbal Ollero

[5] M. Hefeeda and M. Bagheri, (2017), "Wireless Sensor Networks for Early Detection of Forest Fires," 2007 IEEE International Conference on Mobile Adhoc and Sensor Systems, Pisa, pp. 1-6. doi: 10.1109/MOBHOC.2007.4428702

[6] Antonio Molina-Pico, David Cuesta-Frau, Alvaro Araujo, Javier Alejandre, and Alba Rozas, (2016), "Forest Monitoring and Wildland Early Fire Detection by a Hierarchical Wireless Sensor Network," Journal of Sensors, vol. 2016, Article ID 8325845, 8 pages, doi: 10.1155/ 2016/ 8325845

[7] Hakan Habiboglu, Y & Gunay, Osman & Cetin, A. (2011). Real-time wildfire detection using correlation descriptors

[8] Yanjun Li, Zhi Wang and Yeqiong Song, (2006),"Wireless Sensor Network Design for Wildfire Monitoring, " 2006 6th World Congress on Intelligent Control and Automation, Dalian, pp. 109-113.

[9] Ahmad A. A. Alkhatib, (2014), A Review on Forest Fire Detection Techniques, International Journal of Distributed Sensor Networks Vol 10, Issue 3, 2014

[10] Brigitte Leblon, Laura Bourgeau-Chavez and Jesús San-Miguel-Ayanz (2012). Use of Remote Sensing in Wildfire Management, Sustainable Development - Authoritative and Leading Edge Content for Environmental Management, Dr. Sime Curkovic (Ed.), InTech, DOI: 10.5772/45829.

[11] Lina Tang, Guofan Shao, (2016), Drone remote sensing for forestry research and practices, Journal of Forestry Research, 2015, Volume 26, Number 4, Page 791

[12] Li, S., Lowe D., Kong X., Braun R. (2011), Wireless Sensor Network Localization Algorithm Using Dynamic Path of Mobile Beacon, APCC 2011: 17th Asia-Pacific Conference on Communications, October 2011, Sabah, Malaysia.

[13] Li S., Kong X. & Sandrasegaran, K. (2013) "Dynamic Path of Mobile Beacon in Localization of Wireless Sensor Network" International Journal of Sensor Networks

[14] Li S., Kong X., Lowe D. (2012) "Dynamic Path of Mobile Beacon Employing Reinforcement Learning in WSN Localization". International Workshop on Data Management for Wireless and Pervasive Communications, March 2012 Japan

[15] Li S., Kong X., Lowe D. (2012) "Wireless sensor network localization with autonomous mobile beacon by path finding", International Conference on Information Science and Applications (ICISA 2012), May 2012 Suwon, S. Korea

[16] Darroudi SM, Gomez C., (2017), Bluetooth Low Energy Mesh Networks: A Survey, Sensors., 17(7): 1467.

[17] http://www.dji.com/intelligent-flight-modes/v1-doc

[18] https://www.bluetooth.com/

AUTHORS

Songsheng Li received his BS degree from Beijing University of Post and Communication in 1995 and MS degree from Xi'an Jiaotong University, P. R. China in 2002. He earned his Ph.D. of Engineering from University of Technology, Sydney in 2013. His research interests include Wireless Sensor Network, IOT and Data Mining.

APPLICATION BASED ON FUZZY LOGIC TO DETECT AND PREVENT CYBERBULLYING THROUGH SMARTPHONES

José Á. Concepción-Sánchez, Pino Caballero-Gil and Jezabel Molina-Gil

Department of Computer Engineering and Systems,
University of La Laguna, Tenerife, Spain

ABSTRACT

Derived from bullying, cyberbullying is a new problem that is spreading because of the many advances in technology, like the Internet and smartphones, affecting especially to the world's youth population. Currently, there are some studies to investigate their effects on victims or suggest different solutions to detect it. However, a definitive tool that can detect and prevent this harassment is no yet available. For this reason, this paper proposes a novel mobile application for smartphones that will allow detecting whether a person is being a victim of cyberbullying. For this, the application is based on a set of data processing techniques and fuzzy logic that make up a system for decision making, capable of detecting harassment effectively. In addition, this paper also includes some experimental results obtained by performing cyberbullying detection tests with the application.

KEYWORDS

Cyberbullying, Fuzzy Logic, Data Processing, Mobile Application, Security

1. INTRODUCTION

In recent years, when the Internet has expanded around the world and has become an indispensable tool in our everyday lives, cybersecurity is playing an increasingly fundamental role. That is why, in a society that lives constantly connected, the Internet is now an ideal space for cybercrimes, such as cyberbullying [1]. Unlike traditional bullying, where harassment of a person is usually verbal or physical, cyberbullying is characterised by harassment or intimidation produced through social networks or instant messaging systems. The consequences of suffering this problem are many and varied. Some of these are low self-esteem, emotional disorders, depression or anxiety among others [2].

Cyberbullying is a global problem (see Figure 1) [3]. It has increased by 88% in the last five years in countries like the UK [4]. Others countries, such as Singapore [5] or Argentina [6], have seen how more and more young people between the ages of 7-18 are being affected by this serious problem. There are countries that are starting to take action, as is the case of Germany, which has made a law proposal to penalise social networks that do not eliminate offensive and humiliating messages [7]. However, this is not enough since many of the messages are sent through instant messaging systems, which are much more difficult to control. In this way, all these data indicate that there is an urgent need to look for possible solutions that can detect, eradicate and prevent this problem.

David C. Wyld et al. (Eds) : CNDC, DPPR, AIAA, WiMNET, ACITY - 2017
pp. 11– 23, 2017. © CS & IT-CSCP 2017 DOI : 10.5121/csit.2017.71602

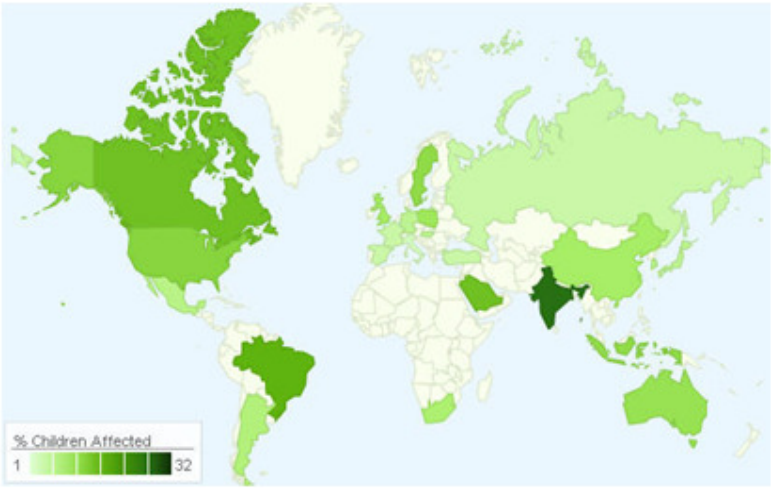


Figure 1. Children affected by global cyberbullying.

This paper presents a novel mobile application that will allow the detection of possible cases of cyberbullying at an early stage in order to act accordingly, presenting itself as a real alternative to eliminate this problem. To do this, the proposed mobile application will analyse all the messages that are received on the victim's smartphone and will decide, using data processing and decision-making techniques, if cyberbullying can exist in the analysed content.

One of the techniques used in this paper is fuzzy logic [8], on which the system for decision making is based. Fuzzy logic is a computational intelligence technique that allows working with information with a high degree of imprecision. It is a multivalued logic that allows intermediate values to be able to define evaluations between yes and no. This is the main difference with the conventional logic that works with well-defined and precise information. This work is based on the use of this mathematical tool since it is not possible to use conventional logic to determine if a person is being a victim of cyberbullying. Throughout the paper, the operation of the system will be further detailed.

This paper is structured as follows. In section 2, some works related to this proposal are mentioned. In section 3, some mobile application features and operation are presented. Section 4 describes the proposed architecture for cyberbullying detection. In section 5 some experimental results of the cyberbullying detection are presented. Finally, the paper is closed with some brief conclusions and future works in section 6.

2. RELATED WORKS

There are numerous papers that study different ways of trying to combat against cyberbullying [9] [10]. For instance, in [11] police are described as an actor in addition to parents, students, schools and service providers on the Internet to combat this problem. According to that study, the police is one more means that can help in preventing cyberbullying by carrying out information tasks for students, parents and schools, creating online information systems (in addition to face-to-face channels), identifying perpetrators and helping the victims.

Research has been also carried out based on paradigms of text mining for topics related to the detection of cyberbullying such as the detection of online sexual harassers [12], the detection of vandalism [13] or the detection of cyberterrorism [14] [15]. However, few studies have been developed to find technical solutions that allow the detection of cyberbullying. Among these

studies, there are some focused on the detection of cyberbullying through patterns in the analysed texts. For example, the paper [16] proposes the use of machine learning to detect cyberbullying. According to it, through automatic learning, the proposed tool can detect language patterns used by bullies and victims, and develop rules to automatically detect bullying content. The data that were used to carry out the tests were extracted from the web Formspring.me, with a result of 78.5% accuracy in the detection of messages with harassment.

On the other hand, the papers [17] [18] [19] propose that for the detection it is necessary to take into account the context as well as the profile and characteristics of the users being studied. The obtained results reflect an improvement in the accuracy for the detection of cyberbullying. In our proposed system based on fuzzy logic, we have taken into account these studies, incorporating a series of input variables that will allow deducing if there are patterns where messages can be discarded depending on the context and the studied user.

Another interesting study [20] proposes a mechanism that allows recognising in social networks both content and highly potential users in terms of cyberbullying with a high degree of effectiveness. However, this proposal has been designed for its use in social networks, so it would not contemplate other forms such as instant messaging systems, which is one of the most used means for cyberbullying [21].

Generally, studies related to cyberbullying are focused on the search of patterns to detect it as well as on the effects it produces on its victims. In addition, a few proposed applications have focused on this specific area, being a clear disadvantage since today there are numerous different ways and tools that could be used to harass victims.

The system proposed in this paper does not focus on a specific area. The application installed on the victim's smartphone will detect messages and notifications from different media. Once detected, the app will proceed to process the content and, in case there are any word or expression that can be classified as a possible case of cyberbullying, fuzzy logic will be used as a mechanism to make the decision as to whether the victim may be actually being harassed.

3. MOBILE APPLICATION

According to statistical data, by the year 2020, there will be 6.1 trillion active smartphones around the world [22], which means that 70% of the world's population will use them [23]. In addition, the increase in the use of smartphones is growing faster among young people aged 16 to 24 [24]. All these data show that smartphones have become an essential part of our society, so it can be used in the worst case as a tool for cybercrime. Therefore, this work is focused on the development of an application for these mobile devices that serves as a tool to detect and prevent a global problem such as cyberbullying. Its main features are:

- Privacy: As in most cases the victims of harassment try to hide their situation [25], this application is designed so that the parents of the youth can install it on their mobile phone without he/she having to know it. To do this, the application will be installed on the smartphone in a hidden way, without an icon. Also, the bullies will also not know that the victim has the application installed on the mobile device. Thus, in case it is detected that the youth is being a victim of cyberbullying, only the parents will be notified so that they can take the appropriate measures (see Figure 2).

- Easy to use: At the usability level, the application is very intuitive and easy to use, since most of its operation is on the background. Parents will only have to configure the contact parameters and the code to access the application.

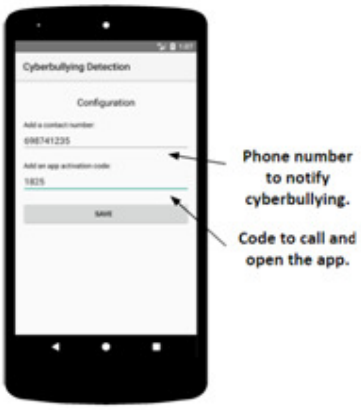


Figure 2. Settings screen in the mobile application.

Once the application is installed on the smartphone of the possible victim, the number of the contact to be notified via SMS should be indicated in the event that a case of cyberbullying is detected. In addition, it is also necessary to enter a numeric code that will be used later to return the application to the foreground by making a call to that code. After setting these parameters, it will only be necessary to press a button that will leave the application running in the background until the call to the code provided in the initial configuration is made.

Note that even if the mobile reboots or goes off, the proposed application will start automatically and continue to run in the background when the smartphone turns on.

3.1. Notification Detection

The mobile application will be listening in the background so that every time a notification to the smartphone arrives, it can collect the information from it (see Figure 3). For this, the application parses the information that is in the notification to extract the texts and if possible, also the application and its issuer. Once obtained, if the content is sufficient, the app proceeds to the analysis to check if there is cyberbullying. Otherwise, it will be saved encrypted in a local database until there is more content to be able to analyse it.

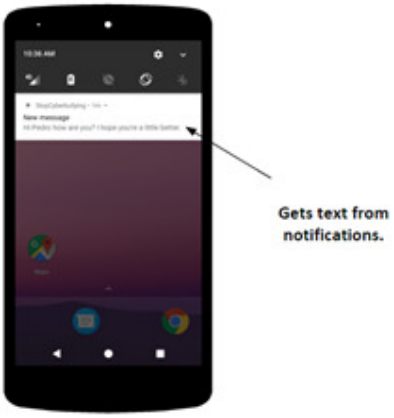


Figure 3. Sample notification that the application can detect.

Thus, any notification of instant messaging applications like WhatsApp or Telegram, email applications like Gmail or even social networks like Facebook as well as any other application

that sends some type of text can be analysed to be able to detect if the user of the smartphone is a victim of cyberbullying.

3.2. Privacy and Security

The information that this application works with is very sensitive because most of this information is formed by conversations and personal notifications of the user of the application. Therefore, it is necessary that all this information is treated as carefully as possible because it could get to make a misuse of it and invade the privacy of the user.

In this way, no one will be able to obtain complete texts of notifications or conversations, neither the parents nor legal custodians who installed the application on the child. They will only get, in case cyberbullying is detected, the words and expressions that have triggered the alarm. No other content will be accessible to anyone.

In addition, the mobile application will only work locally with the data and will only save the collected information of the notifications in case it is insufficient to be able to analyse it, avoiding unnecessary analyses that could affect the battery of the mobile phone. In this case, the information will be encrypted with the cryptographic algorithm AES 256 CBC mode [26], in order to avoid that the information can be subtracted. Once there is enough content to analyse, the stored information will be removed from the database after analysis.

Finally, the local database of the application containing all the words and expressions related to cyberbullying is updated automatically by a call to a remote database that is continually updated with new words and expressions. In this case, no additional security measures are necessary since the call to update the database is done using HTTPS.

4. PROPOSED ARCHITECTURE

The proposed system architecture to detect possible cases of cyberbullying (see Figure 4) is composed of three main steps:

- Get data from notifications: Messages that the possible victim of cyberbullying receives in his smartphone.
- Data processing: Analyse the messages to eliminate unnecessary words (prepositions, articles, ...) and look for matches with the local database fed by words and expressions that can refer to bullying.
- Fuzzy logic system: Verify, from a series of entries, if there really is a possible case of cyberbullying.

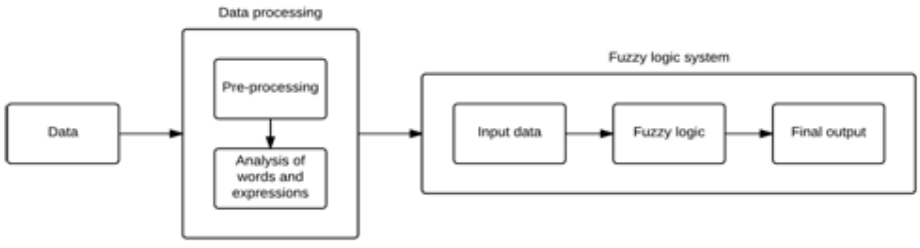


Figure 4. Structure of the decision-making system.

In the following subsections, the operations of each of the steps will be described in more depth.

4.1. Get Data From Notifications

The objective of this first step is to obtain the text that is going to be analysed in search of possible cyberbullying. To do this, once the application detects a push notification [27], it will check if it has enough content to analyse it at the moment, or if there are messages saved from previous notifications that, concatenated with the content of the current notification, give a set of reasonable information for analysis. If it is sufficient, the next step will be executed. Otherwise, the contents of the notification will be stored in the database for future analysis.

In addition, the mobile application will attempt to obtain the name of the contact or phone number that generated the push notification by parsing the content of the notification. In this way, if there is cyberbullying, the stalker will be detected too.

It is worth mentioning that it has been decided to store and concatenate several messages when they do not contain a minimum amount of text because the battery in the smartphones is limited and running the analysis process for each message can consume a lot of battery. So, we guarantee that it will only run when there is a reasonable amount of content for analysis.

4.2. Data Processing

Once the messages that will be analysed are obtained, in this step they are processed to simplify the chains and to check if there are possible words or expressions identified with cyberbullying.

4.2.1. Pre-Processing

Data pre-processing [28] is a very important step because most of the content that is sent in instant messaging systems or social networks is usually negligible. In this way, a data cleaning method has been implemented to allow rejecting words that lack information such as articles, prepositions or other predefined words that are not to be evaluated later. Using this mechanism, the processing times in the subsequent steps are reduced.

Once the negligible content is removed, the resulting string is divided into words and word sets that will be used in the next step. The result would be classified as shown in Table 1.

Table 1. Results after text pre-processing.

Id	Value
1	$word_m$
m	…
m+1	$expression_{m+n}$
m+n	…

4.2.2. Analysis of Words and Expressions

As mentioned in Subsection 3.2 about privacy and security, there is a local database in the mobile application that contains a set of predefined words and expressions that may be related to cyberbullying. This database will be used to check if there is any match between its content and the set of words and expressions obtained in the previous step. Its operation is detailed in Figure 5.

If no match is detected, the possibility of cyberbullying will be discarded in the analysed messages and the use of the fuzzy logic mechanism will not be necessary. On the other hand, if any coincidence is detected with the database, the inputs will be prepared for entry into the fuzzy logic system.

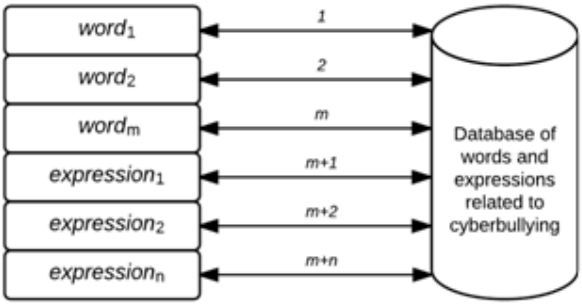


Figure 5. Checking matches.

4.3. Fuzzy Logic System

The fuzzy logic system will be responsible for making the final decision about whether the potential victim is receiving cyberbullying. This mechanism is used because it allows obtaining a greater range of decision making, where a deterministic system would not be able to solve the problem. Therefore, the outputs of this system are not only based on yes/no, but also a third option is considered where an incidence is generated. When the system considers that there are not sufficient data to indicate that the user is receiving cyberbullying but the possibility cannot be ruled out, a new incidence will be saved in the database of the application that will be taken into account for future decision making.

In this way, if the process of analysis of words and expressions finds some match between the database and the analysed chains, the input values composed of the following linguistic variables will be initialized:

- Different Detected (DD): Number of different words and expressions that have been detected in the analysed chain.
- Total Detected (TD): The total number of words and expressions detected in the analysed chain, including repetitions of the same.
- Last Incident (LI): Days since the last time an incident was generated.
- Total Incidents (TI): Total number of generated incidents in the last thirty days.

In turn, each of the linguistic variables is composed of the linguistic terms HIGH, MEDIUM and LOW, which indicate the possibility that the possible victim is suffering cyberbullying. Figure 6 shows the graphs where they are represented. The X coordinate corresponds to the values that can be taken by the linguistic variables and the Y coordinate with the probabilities corresponding to the linguistic terms.

In the case of DD, it is considered that up to two different expressions detected in the strings have LOW-MEDIUM probability that it is cyberbullying since often young people use a colloquial vocabulary to talk to each other. However, from the four expressions, it is considered that there could be a high degree of probability that the individual is a victim of cyberbullying.

On the other hand, in TD the value ranges of the linguistic terms increase because the repetitions of the expressions or words are taken into account. Thus, the linguistic term HIGH takes its maximum probability from the eight total coincidences.

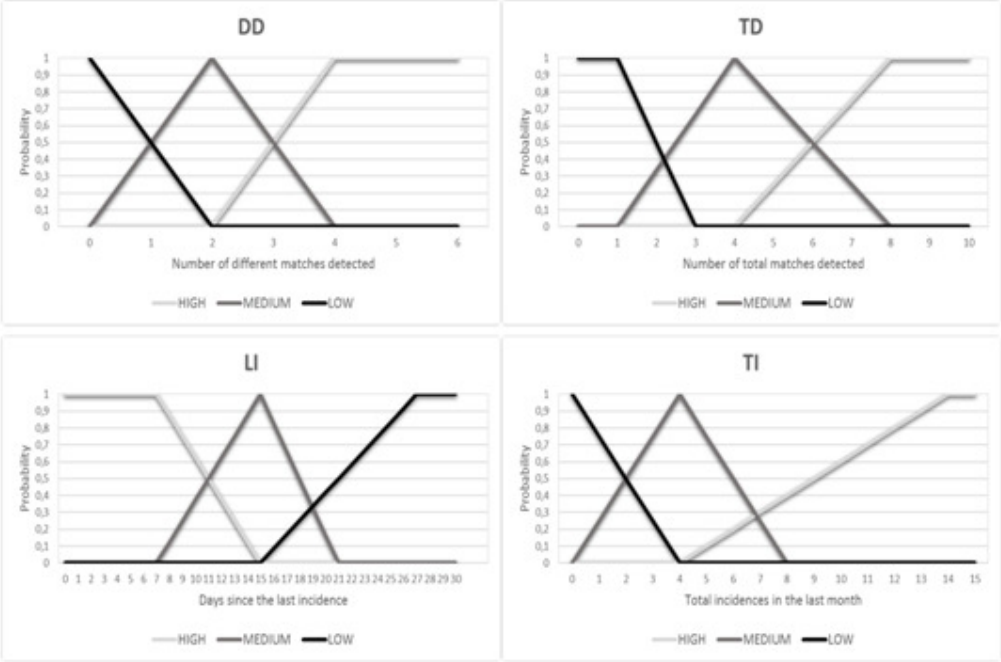


Figure 6. Graphs where the linguistic terms for each of the linguistic variables are represented.

For its part, the linguistic variable LI takes the linguistic term HIGH with its maximum probability until seven days since the last incidence occurred. A set of incidents in a short period of time can be a clear indicator that the possible victim is receiving cyberbullying.

Finally, the higher the TI value, the greater the likelihood that the individual is being harassed, with the linguistic term HIGH being most likely from the fourteen incidences.

Each of these linguistic terms is represented by a membership function [29] that defines them. For our case, the triangular type membership functions have been used. In Figure 7, the membership functions corresponding to the linguistic terms of the variable LI are shown as an example.

$$f_{HIGH}(x) = \begin{cases} 1, & x \leq 7 \\ \dfrac{15-x}{15-7}, & 7 < x < 15 \\ 0, & x \geq 15 \end{cases}$$

$$f_{MEDIUM}(x) = \begin{cases} 0, & x \leq 7 \\ \dfrac{x-7}{15-7}, & 7 < x \leq 15 \\ \dfrac{21-x}{21-15}, & 15 < x < 21 \\ 0, & x \geq 21 \end{cases}$$

$$f_{LOW}(x) = \begin{cases} 0, & x \leq 15 \\ \dfrac{X-15}{27-15}, & 15 < x < 27 \\ 1, & x \geq 27 \end{cases}$$

Figure 7. Membership functions of LI's linguistic terms.

The next step after fusing is the formulation of specific rules for expressing the combination of influences. As an example, Figure 8 shows a simple rule structure, where DECISION is an output linguistic variable defined by three linguistic terms: YES, INCIDENCE and NO. These linguistic terms are associated with their corresponding membership functions, where the output will depend on which linguistic term has the maximum probability, using the max-membership defuzzification method. Besides, there may be more than one value assignment rule for the DECISION. In this case, the assignments to the DECISION are combined by an implicit AND, so the probability corresponding to the DECISION corresponds to the minimum value between all the input linguistic variable probabilities.

```
Input: The fused values of DD, TD, LI and TI.
Output: The fused values of YES, INCIDENCE and NO.
if (TI == HIGH);
 then
    DECISION = YES;
end
if (DD == HIGH) and (LI == LOW);
 then
    DECISION = INCIDENCE;
end
if (DD == MEDIUM) and ((TD == LOW) or
  (TD == MEDIUM));
 then
    DECISION = NO;
end
```

Figure 8. Sample rule structure.

As an illustrative example using the rules shown in Figure 8, if DD has one match as value, its linguistic terms will be fuzzified with 0.5 as LOW and 0.5 as MEDIUM. On the other hand, if TD has a value of four matches, its linguistic term MEDIUM will be fuzzified with a probability of 1. Once fuzzified the linguistic terms, of the three rules established in the example, the third is the one that would be fulfilled, reason why DECISION would have NO as a result. Defuzzifiying this linguistic term would indicate that the text analysed does not contain evidence of cyberbullying. If more rules were fulfilled, they would be combined by AND, and the linguistic term with the highest probability will be chosen, as mentioned above.

Finally, once the system returns the output, the mobile application will proceed to perform the corresponding actions depending on the value obtained.

5. EXPERIMENTAL RESULTS

Along with the application and proposed fuzzy-logic-based architecture, a series of experimental tests have been carried out to verify the degree of efficiency in the detection of possible cases of cyberbullying through smartphones. For this, tests have been performed simulating different environments that could be given in a real case. These are described in Table 2

These environments have been selected for the study because they are the most likely to detect cyberbullying. For example, a case has not been added in which texts have no content related to cyberbullying because the application will directly discard it using data processing. However, in the first case, the application could detect cyberbullying because young people often use, to communicate with each other, words that could be found within the application database.

. Table 2. Description of the cases studied.

Case	Description
1	Use of slang or informal language in conversations between friends. In this case, the application should not detect cyberbullying.
2	Isolated cases where there could be numerous coincidences with words and expressions related to cyberbullying. An incidence should be generated for future analysis.
3	Intermediate harassment. It is the most difficult case to classify because there is evidence of possible cyberbullying but it is difficult to locate where the boundary is between harassment or not.
4	The person is a victim of cyberbullying. It should be detected.

For each of these environments, notifications have been generated with texts that simulate each one to check the degree of effectiveness of the application when it comes to decision making and detection of cyberbullying. The obtained results are classified in Table 3, where:

- No-Cyberbullying represents the percentage of samples that were analysed without finding evidence of cyberbullying. In these cases, the application will not take any action.

- Incidences are the messages that generated some incidence. The application will save the incident in the database so that it can then be taken into account for the next analysis.

- Cyberbullying are the messages that generated alarms. An SMS will be sent with the words and expressions that generated the alarm to the number phone entered in the application settings. If possible, the name of the application will also be sent from where the cyberbullying was done as well as the name of the stalker.

Table 3. Final results.

Case	No-Cyberbullying (%)	Incidences (%)	Cyberbullying (%)
1	94%	6%	0%
2	79%	18%	3%
3	39%	48%	13%
4	9%	69%	22%

As can be seen in the first case, most of the texts were classified without cyberbullying content. This is due to the data processing mechanism detailed in the previous section and to the combination of parameters used in the fuzzy logic system, thus avoiding that slang or informal conversations among youth, are identified as cyberbullying. On the other hand, in the second and third cases, a greater number of incidents were generated, something normal due to the detection of a greater number of words and expressions identified with cyberbullying. In these situations, if the data are not very certain, the system will assimilate it by default as cyberbullying and notify the parents of the possible victim, preventing a future problem. Finally, in the latter case, the percentage of alarms increased because there were a big number of incidents, which is normal because the messages that were used in this case always had expressions or words related to cyberbullying.

The obtained results have been satisfactory, since the detection of cyberbullying through the notifications has been in accordance with the environment studied. However, there is still room for improvement. For example, in the first case, although no alarms were generated, 0% of incidents should have been obtained because the analysed texts simulated a conversation between friends. In this way, it may be necessary to add new mechanisms and systems that allow to refine

and improve the decision-making process for the detection of this problem and to prevent false positives.

6. CONCLUSIONS

Cyberbullying is a problem that affects many young people around the world, and every year the numbers of affected continue to grow very quickly. In the present, there are some studies and small contributions that propose methods to solve this problem but only in specific situations.
This paper proposes the use of a mobile application that will be installed on the smartphone of the possible victim to analyse the received messages and notifications, allowing the detection at an early stage of a possible case of harassment or cyberbullying. For this, the application has been enriched with a set of data processing methods and a system based on fuzzy logic that will be in charge of determining if the user is being a victim of cyberbullying. The obtained results showed that the proposal promises to be an effective tool to combat against this problem.

Among the main advantages of this system are the use of smartphones as a means to detect cyberbullying due to its widespread use, as well as the analysis of texts of notifications that any application can generate, being different from other proposals that focus on a specific application or system. In addition, it is very easy to use and totally invisible to save the privacy of the possible victim.

Finally, as future works, it is necessary to study the inclusion of new complementary models that allow improving the accuracy of the proposed system to decide if an individual is being harassed, as mentioned in the previous section. For instance, to include the use of artificial neural networks could help the application to acquire its own knowledge and that it then serves in decision making. In addition, the application should also be improved with voice recognition systems so that it can analyse the notifications that contain this type of messages. Finally, it is also intended to develop the application for the rest of mobile operating systems, so that the application can be accessible to all users regardless of the smartphone.

ACKNOWLEDGEMENTS

Research supported by the Spanish Ministry of Economy and Competitiveness, the European FEDER Fund, and the CajaCanarias Foundation, under Projects TEC2014-54110-R, RTC-2014-1648-8, MTM2015-69138-REDT and DIG02-INSITU.

REFERENCES

[1] “What is Cyberbullying?” raising awareness of cyberbullying, for students, parents and educators - what is cyberbullying? [Online]. Available: http://www.cyberbullying.info/whatis/whatis.php. [Accessed: 03-Nov-2017].

[2] K. Rigby, (2001) Health Consequences of Bullying and Its Prevention. Peer harassment in school: The plight of the vulnerable and victimized, 310.

[3] “Cyberbullying World Map – Australia Rated as 5th Worst,” iCyberSafe.com - Living in a Connected World, 23-Apr-2012. [Online]. Available: https://icybersafe.com/2012/04/22/cyberbullying-world-map-australia-rated-as-5th-worst/. [Accessed: 03-Nov-2017].

[4] “Online bullying counselling on increase, says Childline,” BBC News, 14-Nov-2016. [Online]. Available: http://www.bbc.com/news/uk-37970725. [Accessed: 03-Nov-2017].

[5] "Bullying in Singapore: Between the Classroom and the Office," NoBullying - Bullying & CyberBullying Resources, 06-Dec-2016. [Online]. Available: https://nobullying.com/bullying-in-singapore/. [Accessed: 03-Nov-2017].

[6] "Bullying in Argentina," NoBullying - Bullying & CyberBullying Resources, 26-Mar-2017. [Online]. Available: https://nobullying.com/bullying-in-argentina/. [Accessed: 03-Nov-2017].

[7] "Germany to fine social media giants up to €50 million for hate speech," The Local, 05-Apr-2017. [Online]. Available: https://www.thelocal.de/20170405/germany-to-fine-social-media-giants-up-to-50-million-for-hate-speech. [Accessed: 03-Nov-2017].

[8] L.A. Zadeh, (1965) Fuzzy sets. Information and control 8, 3 (1965), 338–353.

[9] M. Marczak and I. Coyne, (2010) Cyberbullying at School: Good Practice and Legal Aspects in the United Kingdom. Australian Journal of Guidance and Counselling 20, 2, 182–193. https://doi.org/10.1375/ajgc.20.2.182

[10] M.L. Genta, A. Brighi, and A. Guarini, (2009) European project on bullying and cyberbullying granted by Daphne II programme. Zeitschrift für Psychologie/Journal of Psychology 217, 4, 233.

[11] H. Vandebosch, L. Beirens, W. D'Haese, D. Wegge, and S. Pabian, (2012) Police actions with regard to cyberbullying: The Belgian case. Psicothema 24, 4, 646–652.

[12] A. Kontostathis, (2009) ChatCoder: Toward the tracking and categorization of internet predators. In Proc. Text Mining Workshop 2009 Held in Conjunction with the Ninth Siam International Conference on Data Mining. Sparks, NV.

[13] K. Smets, B. Goethals, and B. Verdonk, (2008) Automatic vandalism detection in Wikipedia: Towards a machine learning approach. In AAAI workshop on Wikipedia and artificial intelligence: An Evolving Synergy. 43–48.

[14] D.A. Simanjuntak, H.P. Ipung, A.S. Nugroho, et al, (2010) Text classification techniques used to faciliate cyber terrorism investigation. In Second International Conference on, Advances in Computing, Control and Telecommunication Technologies (ACT). 198–200.

[15] Y. Elovici, A. Kandel, M. Last, B. Shapira, and O. Zaafrany, (2004) Using data mining techniques for detecting terror-related activities on the web. Journal of Information Warfare 3, 1, 17–29.

[16] K. Reynolds, A. Kontostathis, and L. Edwards, (2011) Using machine learning to detect cyberbullying. In 10th International Conference on, Machine learning and applications and workshops (ICMLA), Vol. 2. 241–244.

[17] M. Dadvar, D. Trieschnigg, R. Ordelman, and F. de Jong, (2013) Improving cyberbullying detection with user context. In European Conference on Information Retrieval. Springer, 693–696.

[18] M. Dadvar, F.M.G. de Jong, R.J.F. Ordelman, and R.B. Trieschnigg, (2012) Improved cyberbullying detection using gender information. (2012), 23–25.

[19] A. Squicciarini, S. Rajtmajer, Y. Liu, and C. Griffin, (2015) Identification and Characterization of Cyberbullying Dynamics in an Online Social Network. In Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '15). ACM, New York, NY, USA, 280–285. https://doi.org/10.1145/2808797.2809398

[20] Y. Chen, Y. Zhou, S. Zhu, and H. Xu, (2012) Detecting offensive language in social media to protect adolescent online safety. In International Conference on and 2012 International Confernece on Social Computing (SocialCom), Privacy, Security, Risk and Trust (PASSAT). 71–80.

[21] R.M. Kowalski, S.P. Limber, S. Limber, and P.W. Agatston, (2012) Cyberbullying: Bullying in the digital age. J. Wiley & Sons.

[22] Lunden, Ingrid. "6.1B Smartphone Users Globally By 2020, Overtaking Basic Fixed Phone Subscriptions." TechCrunch, 2 June 2015, techcrunch.com/2015/06/02/6-1b-smartphone-users-globally-by-2020-overtaking-basic-fixed-phone-subscriptions/. [Accessed: 03-Nov-2017].

[23] 70 Percent of Population Will Have Smartphones by 2020." PCMAG, 3 June 2015, www.pcmag.com/article2/0,2817,2485277,00.asp. [Accessed: 03-Nov-2017].

[24] "UK: smartphone ownership by age 2017." Statista, www.statista.com/statistics/271851/smartphone-owners-in-the-united-kingdom-uk-by-age/. [Accessed: 03-Nov-2017].

[25] M. Fekkes, F.I.M. Pijpers, and S.P. Verloove-Vanhorick, (2005) Bullying: who does what, when and where? Involvement of children, teachers and parents in bullying behaviour. Health Education Research 20, 1, 81. https://doi.org/10.1093/her/cyg100arXiv:/oup/backfile/content_public/journal/her/20/1/10.1093/her/cyg100/2/cyg

[26] J. Daemen and V. Rijmen, (2013) The design of Rijndael: AES-the advanced encryption standard. Springer Science & Business Media.

[27] K.M. Bell, D.N. Bleau, and J.T. Davey, (2011) Push notification service. (Nov. 222011). US Patent 8,064,896.

[28] S.S. Baskar, L. Arockiam, and S. Charles, (2013) A systematic approach on data pre-processing in data mining. Compusoft 2, 11, 335.

[29] L.A. Zadeh, (1975) The concept of a linguistic variable and its application to approximate reasoning—I. Information sciences 8, 3, 199–249.

AUTHORS

José Á. Concepción-Sánchez is a student of a Master in Mobile Application Development. He graduated as Computer Science Engineer in 2016 from the University of La Laguna and belongs to the CryptULL Research group. During his research period, he has participated in various national and international conferences and also participates in the research projects of the group to which he belongs.

Pino Caballero-Gil is a Full Professor of Computer Science and Artificial Intelligence at the University of La Laguna, Spain, where she leads the CryptULL research group on Cryptology. Her major research interests are in secure mobile applications, stream ciphers, strong identification, cryptographic protocols, vehicular networks and security in wireless networks.

Jezabel Molina-Gil received her Computer Science Engineering Degree from the University of Las Palmas de Gran Canaria (España) in 2007 and her Ph.D. from the University of La Laguna in 2011. Her research is on VANET security, specially in cooperation and data aggregation. She belongs to the CryptULL research group devoted to the development of projects on cryptology since 2007, and is involved in several projects and publications related to this area. She has authored several conference and journal papers.

INTENTIONAL BLANK

ACCESS CONTROL SYSTEM BASED ON RASPBERRY PI AND ANDROID SMARTPHONES

Jonay Suárez-Armas and Pino Caballero-Gil

Department of Computer Engineering and Systems,
University of La Laguna, Tenerife, Canary Islands, Spain

ABSTRACT

In venues where there are restricted access areas, access control systems that work by entering a code or with some extra element such as a card are used. This paper proposes an access control system based on controlling each restricted area using a Raspberry Pi with an NFC reader, in addition to using Android smartphones emulating an NFC tag as an identification method thanks to an application for these devices. The management of the system is performed through a web panel that also allows to view all the data. Moreover, the system is based on the role access control model. In order to protect the information exchanged between the different elements of the system, security mechanisms are used.

KEYWORDS

Security, Access control, NFC, Host-based Card Emulation

1. INTRODUCTION

For many years, physical security in different environments has been a matter of concern for people, and different alternatives have been sought to prevent the entry of unwanted people into certain places, such as homes, businesses or restricted areas within a building. With this aim, alarms and video surveillance systems have been installed in the most exposed places. If apart from detecting the intrusion of someone in an area, we want to prevent someone from entering, or that even some people can enter and not according to which zone, we talk about access control systems.

In large venues where there is a large number of people, like airports, it is not feasible to use simple keys to access the different areas, either because of the great number of copies of keys that would be circulating or because of the time it would take to request them. Therefore, access control systems are used in these places, which may have different methods of identification, such as the introduction of a number code, or even the introduction of a card in a reader. Another advantage of the access control system is to have a log with all access attempts, whether they have been allowed or not to pass.

In order to improve the current access control systems, this paper proposes a system based on the use of a Raspberry Pi for the control of each area and connected with a server that provides them with the necessary information about the access permissions that the users have in real time, each time a user tries to enter an area. In addition, the system is complemented by the use of Android

David C. Wyld et al. (Eds) : CNDC, DPPR, AIAA, WiMNET, ACITY - 2017
pp. 25– 32, 2017. © CS & IT-CSCP 2017 DOI : 10.5121/csit.2017.71603

smartphones as an identification device in each zone thanks to the NFC Host-based Card Emulation (HCE) mode [1], available for Android devices with version 4.4 or higher, and also thanks to an NFC reader that incorporates each Raspberry Pi.

The next section shows some related works. The details of the system architecture are explained in Section 3. Section 4 describes the designed applications while Section 5 exposes the security mechanisms used to protect the information exchanged between the elements of the system. Finally, a short conclusion closes the paper.

2. RELATED WORKS

Apart from access control systems that can be found in stores, there are different works that address this issue in order to try to design a good access control system with low-cost and usable. One of the trends that currently exists is to use low-cost minicomputers, such as Raspberry Pi or Arduino, to create small access control systems that can be installed in homes. In [2], the authors propose an access control system in which a Raspberry Pi is used to control the access to a house via the Internet, with the ability to view who wants to enter thanks to a camera and also to send a message through a small screen. In [3], a system security with Raspberry Pi has been designed, which is based on image recognition with extra functionalities, such as sending intrusion e-mails to the nearby police department or sending notifications via SMS. There are also works about access control that use an Arduino board with an NFC reader that allows reading NFC tags with the access credentials needed to access a zone [4] or even replace these tags with Android smartphones that are able to simulate an NFC card [5]. The NFC cards emulation on Android mobile phones [6] [7] is also a topic studied by different authors because of the advantages of being able to leave the cards at home and do different operations simply swiping the mobile phone to the reader, or even with an NFC-enabled smartwatch.

In access control systems, it is possible to use different identification methods. These methods can be classified into 3 groups: identification using something that is known, identification through something that is, identification using something that you have. The identification using something that is known is the most used method, since the passwords are in this group. In the second group, identification through something that is, fingerprint identification [8], identification using facial recognition [9] and identification based on iris analysis [10] are included. In the identification using something that you have is where the identification using NFC [11] cards fits in. To make a system stronger, it is possible to combine two identification methods of two different groups, for example the use of an NFC card together with the fingerprint. Authentication with NFC is also studied in some works [12] [13].

3. DESCRIPTION OF THE SYSTEM

The designed system is composed of some elements. The interconnection between them is in Fig. 1. The connection between the user's smartphones and the server is through the Internet, and the connection between the server and each Raspberry Pi is through Ethernet, or even through a Wireless LAN connection in areas without LAN connection.

The server is the centre element and is the brain of the system. It contains all the data needed to manage the access to the restricted areas, and the web panel of management is lodged in it. This element is connected with the other elements of the system in order to give them the necessary information.

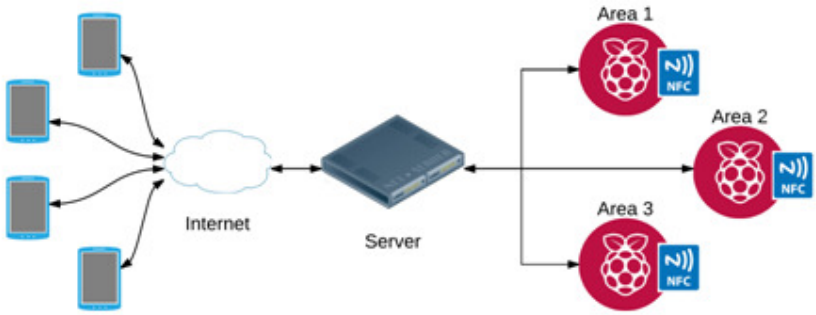


Figure 1. System architecture scheme

Another essential element of the designed access control system is what we will call area control subsystem, because it is composed of different elements managed by a Raspberry Pi. In each restricted area, there will be an area control subsystem. This subsystem consists of a Raspberry Pi together with an NFC reader (MFRC522), a 16x2 LCD display to indicate some information to the users, a green LED to indicate that access has been allowed and a red LED to indicate otherwise, a camera to capture images of people who have accessed the different areas, and a relay that allows the connection of the Raspberry Pi with the door opening system of the controlled area. In Fig. 2, the connection scheme of the Raspberry Pi with the different elements that make up the area control subsystem is shown.

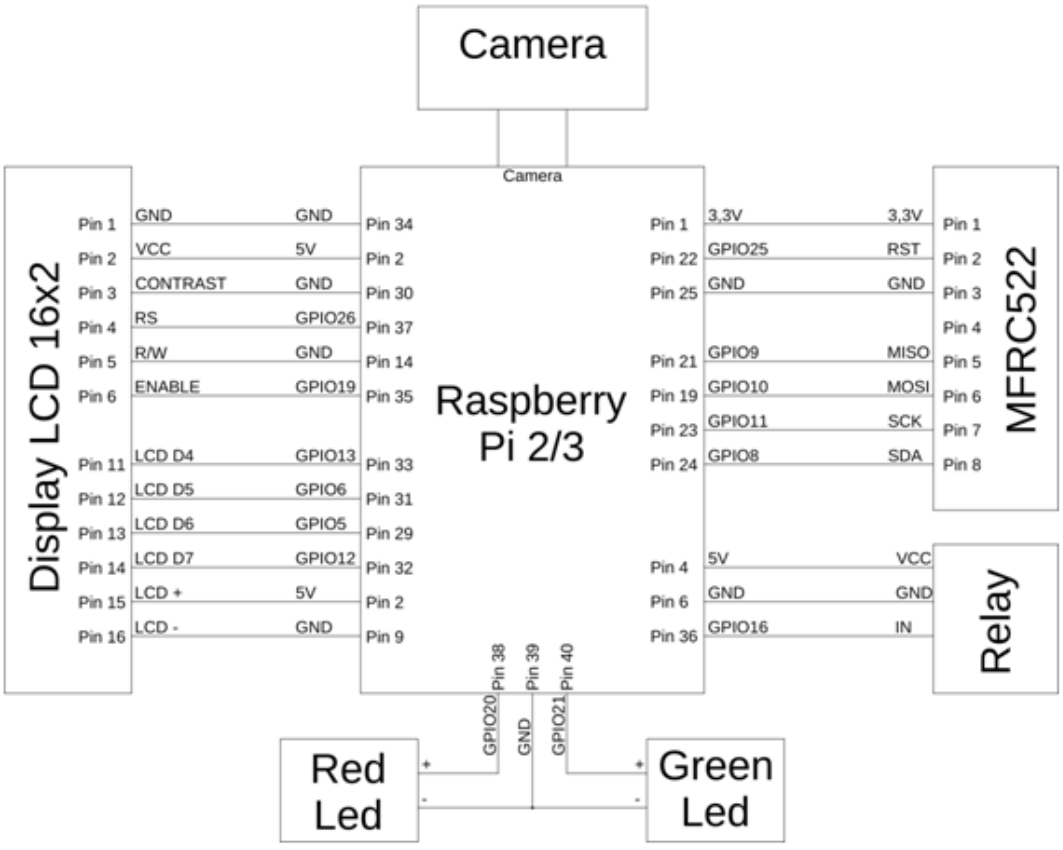


Figure 2. Raspberry Pi connection scheme

The last point of the system is composed of the users' own mobile phones. These Android smartphones are NFC-enabled and they have a version of Android 4.4 or higher, from which the NFC Host-based Card Emulation (HCE) tag emulation mode is available, which allows to emulate an NFC tag on the mobile device, so by bringing the phone closer to an NFC reader, the user requests to access an area. To do this, each of these phones will have an application installed that allows you to connect to the system server to gather the necessary information and then to be able to be identified in each restricted area.

3.1. Role access control model

To improve the granting of permissions, the system is based on the roles access control model [14], in which permissions are not granted to each user, but assigned to a role, and then roles are assigned to system users.

This is an advantage, since it is not necessary to manually assign the permissions to each user, so that if a series of permissions are repeated for several users, they are only assigned once to a role, and then the created role is assigned to the corresponding users. In addition, it is possible to assign several roles to a user, so it is not necessary to modify a role that affects several users, but you can create a new one that will be complemented with those already assigned.

To avoid permissions conflicts, in the designed system permissions are not denied, because by default users do not have any permissions. In this way, when assigning several roles to a user the system does not have to decide whether to allow or deny an access, since all of them would be positive permissions.

4. APPLICATIONS

Three applications have been designed to make up the system: server application, Raspberry Pi application, and Android application.

The web application hosted on the server is in charge of the control of the system, and could be divided into two parts. On the one hand, there is the application in charge of communications with other elements of the system to provide them with the information they need at any time, especially the information of permissions of users each time an access attempt occurs. On the other hand, there is the web panel that allows administrators to perform different system configurations (user creation, role creation, role and permissions assignment) and monitor the access attempts of each area.

The application of Raspberry Pi is responsible for making the communication with the server every time a user approaches his mobile phone to try to access a controlled area. In that communication, it queries if said user has permission to access said zone. It is also responsible for operating the NFC reader, opening the door if necessary and controlling the other elements that are connected.

The application installed in the mobile phones of the users is responsible for establishing NFC communication with the Raspberry Pi of each controlled area. In addition, it connects to the server through an Internet connection to request the access credentials that are sent through NFC when a user tries to access a restricted zone. For security reasons, NFC communication only occurs if the device is unlocked, preventing another person from accessing an unauthorized area using a third-party phone. Through this application it is also possible to view your own access logs.

4.1. Operation

The typical operation mode is divided into several steps (see Fig. 3):

1. From the management web panel, users must be created in the system so that they can get identification data from their Android mobile phones.

2. To assign corresponding permissions to the users created previously. On the one hand, permissions to enter the web panel for reading data or even to modify it will be assigned, and on the other hand restricted areas access permissions will be assigned. The permissions assignment is based on the role access control model previously mentioned.

3. In the Android application, every user must introduce their username and password to get the identification data and to be able to use the system.

4. The mobile phone is set in emulation tag mode and is ready to establish a communication with the NFC readers of each area.

5. Swiping the smartphone by an NFC reader, the system checks if the user has permission to access the corresponding area.

6. The door is opened if the user is authorized.

7. The access attempt is registered and saved in database whether the user could access or not to the restricted area. The data recorded on each attempt are: date and time, user, user's photo, area, and if access has been allowed or denied.

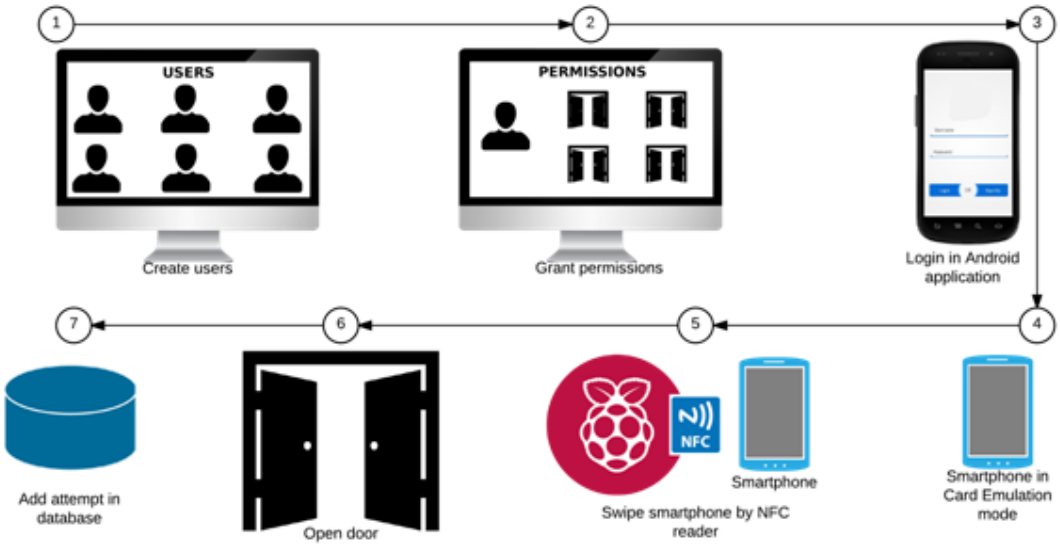


Figure 3. Operation mode

5. SECURITY

An access control system is a platform dedicated to the physical security of a venue, so the logical security is an essential part of the system. For this reason, the corresponding security mechanisms are used to protect each part of the designed system. In Fig. 4 the mechanisms chosen in each part are shown.

First, the communications between the server and the web application that manages and monitors the system, are protected by the use of HTTPS. This protocol is also used to protect the exchange of information between the server and the Raspberry Pi of each restricted area.

On the other hand, is the information that is sent between the server and the mobile application, generally to request to the server the identification data for the access control system. That information is encrypted with 256-bit Advanced Encryption Standard (AES) in Cipher Block Chaining (CBC) mode. In addition, the encryption key used in AES is previously agreed between the server and the smartphone using the Elliptic Curve Diffie-Hellman (ECDH) key agreement algorithm. This agreement is carried out by exchanging the public keys between these two parties through an insecure channel, such as the Internet, and applying in each part the operations corresponding to the ECDH algorithm we obtain on both sides the key that will be used to encrypt with AES.

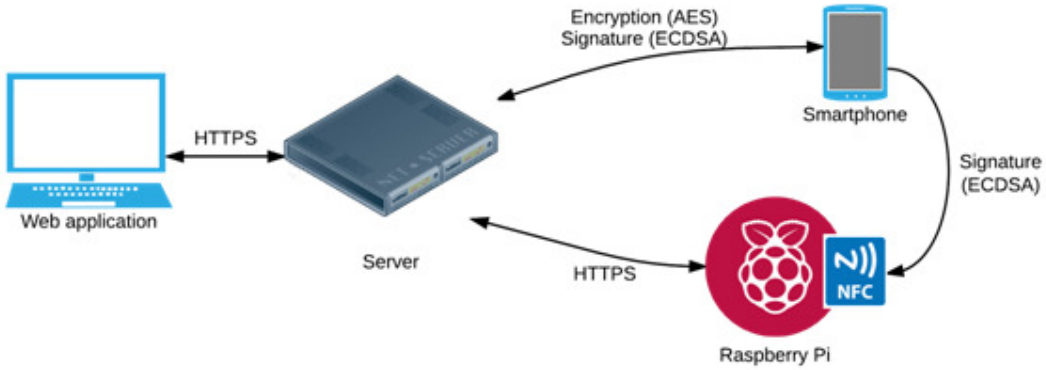


Figure 4. Security scheme

Moreover, the information exchanged between the server and smartphone is digitally signed in order to verify that the information received is correct and also comes from the entity that is expected. For this purpose, Elliptic Curve Digital Signature Algorithm (ECDSA) is used. The use of these two algorithms, provides the system with confidentiality, integrity and authenticity, since the information travels signed and encrypted. The latter involves maintaining a logical order in applying these security algorithms. First, the information is signed and then encrypted, which means that on the other side you first have to decrypt the information and finally verify it before using it or performing any operation with it.

Finally, in order to avoid possible attacks that NFC technology is exposed [15], NFC communications between smartphones and Raspberry Pi also apply security methods. In this case, the information sent from the smartphone to the Raspberry Pi is signed, and then it is verified before using it. For this purpose, the ECDSA digital signature algorithm is also used.

6. CONCLUSIONS

The proposed access control system improves current systems in cost and functionalities thanks to the use of Raspberry Pi as a controller in each restricted area. One of the main advantages of this system compared to those that can be found in stores is the flexibility to add new functionalities and different forms of identification, as well as different sensors that allow to obtain some kind of relevant information. Besides, the use of smartphones in each area as a method of identification is here proposed to replace keys or cards.

Moreover, it is possible to control remote zones in which an Ethernet connection is not available using a wireless network connection because Raspberry Pi has a wireless network interface.

Apart from using Android phones, it is also intended to be able to use iPhone as an identification method, but unfortunately NFC technology is not open to iOS developers.

In the future, cameras connected to the Raspberry Pi will be used to perform video surveillance and to do motion detection tasks using fuzzy logic.

In addition to identification by using smartphones emulating an NFC card, fingerprint identification will be added to provide a higher level of security for the most critical areas. In this way, these two methods will be combined, having to pass both to access the controlled area.

To check the effectiveness of the system, it will be implemented and tested in a real environment. When it is in operation, tests will be carried out to obtain data on its performance. Attacks will also be launched to identify weak points in the security of the system and then correct them.

ACKNOWLEDGEMENTS

Research supported by the Spanish Ministry of Economy and Competitiveness, the European FEDER Fund, and the CajaCanarias Foundation, under Projects TEC2014-54110-R, RTC-2014-1648-8, MTM2015-69138-REDT and DIG02-INSITU.

REFERENCES

[1] "Host-based Card Emulation," Android Developers, 03-May-2017. [Online]. Available: https://developer.android.com/guide/topics/connectivity/nfc/hce.html. [Accessed: 04-Nov-2017].

[2] M. N. Chowdhury, M. S. Nooman, and S. Sarker, "Access Control of Door and Home Security by Raspberry Pi Through Internet," International Journal of Scientific and Engineering Research, vol. 4, pp. 550–558, 2013.

[3] R. Manjunatha and R. Nagaraja, "Home Security System and Door Access Control Based on Face Recognition," International Research Journal of Engineering and Technology (IRJET), 2017.

[4] M. W. D. Saravia, "Access control system using NFC and Arduino," 2015 IEEE Thirty Fifth Central American and Panama Convention (CONCAPAN XXXV), pp. 1–6, 2015.

[5] R. S. Basyari, S. M. Nasution, and B. Dirgantara, "Implementation of host card emulation mode over Android smartphone as alternative ISO 14443A for Arduino NFC shield," 2015 International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC), pp. 160–165, 2015.

[6] N. Saparkhojayev, A. Nurtayev, and G. Seydaliyeva, "NFC-based Access Control and Management System Using Smartphones as Keys," International Journal of Applied Engineering Research, vol. 11, no. 8, pp. 5519–5522, 2016.

[7] N. Saparkhojayev, A. Dauitbayeva, A. Nurtayev, and G. Baimenshina, "NFC-enabled access control and management system," 2014 International Conference on Web and Open Access to Learning (ICWOAL), pp. 1–4, 2014.

[8] S. Shigematsu, H. Morimura, Y. Tanabe, T. Adachi, and K. Machida, "A single-chip fingerprint sensor and identifier," IEEE Journal of Solid-State Circuits, vol. 34, no. 12, pp. 1852–1859, 1999.

[9] J. S. Coffin and D. Ingram, "Facial recognition system for security access and identification," Nov. 23 1999, US Patent 5,991,429.

[10] J. G. Daugman, "Biometric personal identification system based on iris analysis," Mar. 1 1994, US Patent 5,291,560.

[11] V. Sharma, P. Gusain and P. Kumar, "Near field communication," Conference on Advances in Communication and Control Systems 2013 (CA2S 2013), vol. 248001, 2013.

[12] M. Q. Saeed and C. D. Walter, "Off-line NFC Tag Authentication," Internet Technology and Secured Transactions, 2012 International Conference for IEEE (ICITST-2012), pp. 730–735, 2012.

[13] H. Lee, W.-C. Hong, C.-H. Kao, and C.-M. Cheng, "A User-Friendly Authentication Solution Using NFC Card Emulation on Android," 2014 IEEE 7th International Conference on Service-Oriented Computing and Applications, pp. 271–278, 2014.

[14] D. Ferraiolo, J. Cugini, and D. R. Kuhn, "Role Based Access Control: Features and Motivation," In Proceedings of 11th annual computer security application conference, 1995.

[15] E. Haselsteiner and K. Breitfuß, "Security in near field communication (NFC)," In Workshop on RFID security, pp. 12–14, 2006.

AUTHORS

Jonay Suárez-Armas is a Computer Engineer graduated at the University of La Laguna in 2016, and currently is a Master Student in Mobile Application Development and member of CryptULL, the cryptology group of the University of La Laguna. He has participated in different conferences and is the author of several papers.

Pino Caballero-Gil is a Full Professor of Computer Science and Artificial Intelligence at the University of La Laguna, Spain, where she leads the CryptULL research group on Cryptology. Her major research interests are in secure mobile applications, stream ciphers, strong identification, cryptographic protocols, vehicular networks and security in wireless networks.

SEAMLESS MOSAIC OF UAV IMAGES FOR DENSE URBAN AREA

Ming Li[1,2], Ruizhi Chen[1,2], Xuan Liao[1] and Weilong Zhang[1]

[1]State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing, Wuhan University, Wuhan City, China
[2]Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan City, China

ABSTRACT

This paper aims to put forward a seamless mosaic method of UAV image for dense urban area, which can effectively avoid seam-line pass through the edge of the building, so as to eliminate the ghosting, dislocation and seam in the image mosaic process. Firstly, the radiation error of UAV image are corrected by Wallis algorithm, and extract the corresponding points from the adjacent images by SIFT algorithm, to correct the left and right pending matching images to the virtual unified reference image, to ensure the images are in the same coordinate system. Then, in view of the shortcomings of the classical Duplaquet method, we proposed a new more robust UAV image mosaic algorithm by changing the energy accumulation criterion of energy function for dynamic programming. Finally, the comparative experiments show that our method can find the optimal seam-line to avoid it through the edge of houses, especially in dense urban area.

KEYWORDS

UAV Images, Seamless Mosaic, Seam-line, Dynamic Programming, Dense Urban Area

1. INTRODUCTION

Unmanned aerial vehicles (UAV) remote sensing system has been widely used in environmental protection, ecological agriculture, disaster emergency and 3D urban reconstruction with its predominance of low cost, fast and easy to operation. However, UAV due to flight height, camera perspective constrains, the coverage of single UAV image is small. In order to expand the field of view, obtain more remote sensing information about the target area, we need to mosaic multiple UAV images into one image. Because of UAV is lighter in quality, vulnerable to high-altitude winds, flight attitude is unstable, the overlapping area of adjacent images is often irregular, and the image exposure is uneven, they always lead to stitching images prone to ghosting, blurred, dislocation, colour inconsistencies and so on[1-5].

Internationally, researchers have proposed a lot of methods to solve the problem of seamless mosaic of UAV images. Among them, seam-line-based algorithms are an important research branch in this area. This kind of algorithms are intended to find an optimal path with less grayscale and geometric differences, energy function-based algorithm is one of the important methods to resolve the problem. It focus on considering the energy difference between the images, and its effect is more superior to others. In the energy function-based algorithms, references [6-8] adopt the Dijkstra's shortest path algorithm to search for the optimal seam-line, which improves the ghosting and dislocation problems due to objects movement, registration

David C. Wyld et al. (Eds) : CNDC, DPPR, AIAA, WiMNET, ACITY - 2017
pp. 33– 39, 2017. © CS & IT-CSCP 2017 DOI : 10.5121/csit.2017.71604

errors, but its search efficiency is low, and the method of weight determination is complex too. The seam-line searching by the ant colony algorithm can avoid the area where the colour contrast is larger on the image, but the algorithm is sensitive to the number of ants, this will causes the search process of seam-line easy to fall into the local optimum [9]. However, the dynamic programming algorithm is relatively mature, has a relatively complete theoretical system, not easy to innovation of it, and after modified the energy function of dynamic programming algorithm, it still has a strong image direction correlation, that leads to reduce its robustness, especially there is a significant difference in brightness between the adjacent images. Most of the current algorithms still cannot achieve satisfactory results when dealing with image mosaics in dense urban areas. Therefore, this paper attempts to propose a new UAV image mosaic method based on dynamic programming for dense urban area.

2. IMPROVED IMAGE MOSAIC METHOD

2.1. Classic method of dynamic programming

The Duplaquet method is a classic dynamic programming method for seam-line searching. Formula (1) is the energy criterion defined for the algorithm [10]. The algorithm can ensure that the length of the alternative seam-lines are equal, and the seam-line with the smallest accumulated energy values is the optimal.

$$C(x,y)=C_{dif}(x,y)-\lambda C_{edge}(x,y) \tag{1}$$

Thereinto, C_{dif} (x, y) is the mean value of the grey level difference of the pixel in the neighbourhood V which belong to the overlapping areas between the two adjacent images, C_{edge} (x, y) is the minimum gradient value of the pixel in the overlapping areas of image pair, λ is a weighing factor, which be used for adjusting the proportion of grey difference and structure difference in energy function. There are many improved methods. Reference [11] is a representative improvement method based on Duplaquet algorithm, the method put the ratio value of accumulate energy value and the length of seam-line as a measurement criteria, and took the smallest value as the best one, but it still has some problems. It can be seen from the Figure 1, compared with the Duplaquet method, its best seam-line also still pass through a large number of houses, this is not conducive to UAV image mosaic in dense urban areas.

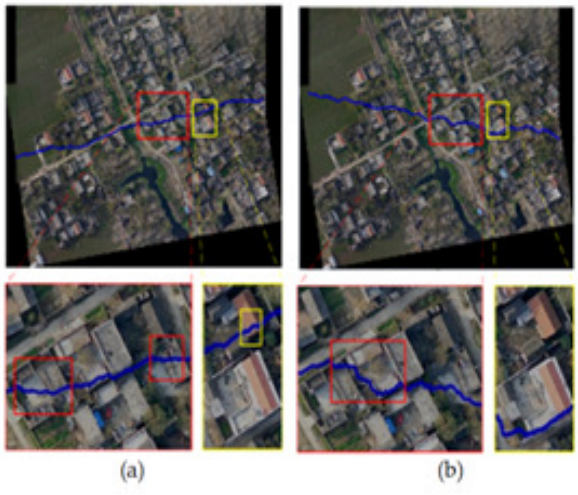


Figure 1. Seam-lines by different methods. (a) Duplaquet method; (b) Reference [11] Method

2.2. Our improved method

The methods mentioned above have the following two problem: 1) Gradient guidance direction of energy function does not support omnidirectional searching. 2) The accumulation of energy function has directionality. The optimal energy function aggregation takes only three directions (energy aggregation directions) into consideration, and the direction of energy aggregation is also limited from left to right, from top to bottom direction (energy traversal direction). The energy criterion proposed by the classic Duplaquet algorithm only considers the horizontal and vertical gradients, and only compares the pixels in the three adjacent directions near the current pixel. When the overlapping area has a large number of dense distribution of buildings, and these houses with different height, due to the deformation are not consistent from image point to the ground point in roof, the seam-lines searched by the Duplaquet algorithm are likely to pass through the edge of the buildings, and when the image matching errors are large, the stitched image is easy to appear obvious dislocation phenomenon.

In order to solve these problems. Firstly, on the basis of considering the 8 directional neighbourhood information of current pixel and its similarity of surrounding structure, we use a new operator to calculate the pixel gradients according to reference [12]. Then, we introduce the fourth direction on the basis of the original three direction in the process of seam-line searching, to correct the problem of seam-line serious deviation from the ideal seam-line by changing the strategy of energy aggregation direction. At last, on the foundation of the Duplaquet method, we redefined the energy criterion and proposed a new dynamic programming method based on two-channel energy accumulation to improve the optimal seam-line searching. As shown in Figure 2, there is a schematic diagram of two-channel-based optimal seam-line searching, which optimizes the seam-line searching criteria by detecting the eight pixels (contain the horizontal direction) that surrounding the current pixel neighbourhood.

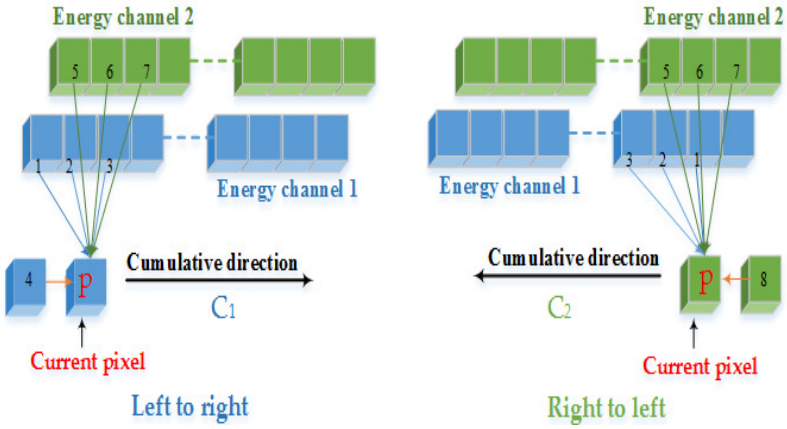


Figure 2. The idea of our improved method

In Figure 2, set P as the current pixel, and redefine the eight comparison directions of P, which respectively is 1(upper left of the current pixel for energy aggregation channel 1), 2(upper of the current pixel for energy aggregation channel 1), 3(upper right of the current pixel for energy aggregation channel 1), 4(left of the current pixel for energy aggregation channel 1), 5(upper left of the current pixel for energy aggregation channel 2), 6(upper of the current pixel for energy aggregation channel 2), 7(upper right of the current pixel for energy aggregation channel 2), 8(right of the current pixel for energy aggregation channel 2), the initial invalid direction is the current pixel itself.

Based on the theoretical analysis, this paper constructed the mathematical abstract expression of the theoretical model that proposed in this paper. Assuming that image 1 and image 2 are original image pair be used for stitching, the energy function defined in this paper is shown in formula (2):

$$E = \sum_{(x,y)\in O} B(x,y)\sigma(|I_1(x,y) - I_2(x,y)|) + \sum_{(x,y)\in O} \sigma\left(\max\left(|\underset{0\le k\le 7}{d_k}(I_1(x,y)) - \underset{0\le k\le 7}{d_k}(I_2(x,y))|\right)\right) + \sum_{(x,y)\notin O} N(x,y) \tag{2}$$

In formulation (2), $B(x, y)$ to determine whether the current pixel (x, y) is in the boundary of overlapping area of the adjacent images, when $B(x, y) =1$, it means that it is not in the boundary region, when $B(x, y) =10$, it means that it is in the boundary region. $\sigma(*)$ is the Gaussian smoothing term, which uses the information in the local window to enhance the local influence of the current pixel, $I_1(\cdot)$, $I_2(\cdot)$ respectively is the pending stitching image, O is the overlapping area, $d(*)$ represents the gradient function of one of the eight directions, $N(x, y)$ is the energy value of the invalid area, which is the constant term, and the value is 100 times than the maximum value of O. If the overlapping area is irregular, it can be extended to a regular area using the smallest circumscribed rectangle of the overlapping area.

3. EXPERIMENT RESULTS ANALYSIS

3.1. Experimental data and environment

In order to verify the effectiveness of our proposed method, this paper selects the UAV image from dense urban area to testing. And compares the method proposed in this paper with the classic Duplaquet method. Before seam-line searching, the radiation errors of UAV images are corrected by Wallis [13], and the pending mosaic images extracted the corresponding points by SIFT [5] to correct them to a same virtual unified coordinate system. In this paper, it use Visual C++ based on OpenCV open source library to program the proposed improvement algorithm. Figure 3 is two group of experimental images. The experimental computer environment is Windows 7 operating system, with 32G computer memory, Intel core 7 CPU.

(a) (b)

Figure 3. The experimental data. (a) Data 1; (b) Data 2

3.2. Results analysis

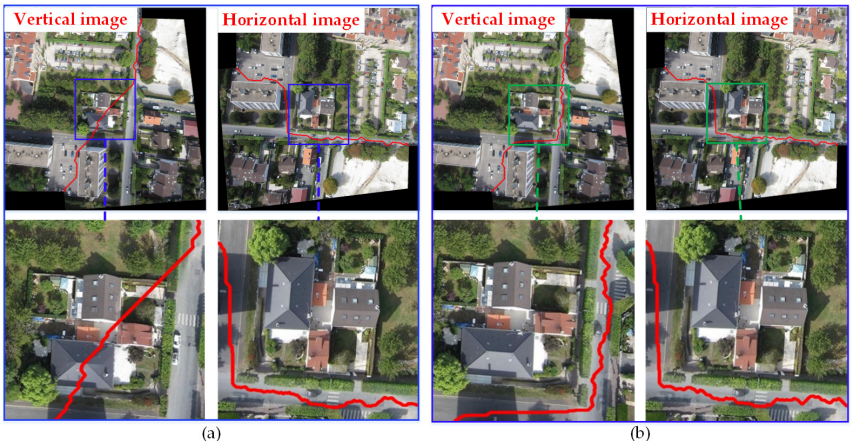


Figure 4. The seam-lines of two methods under rotation. (a) Duplaquet method; (b) our method

Firstly, the vertical image in Figure 3(a) was rotated to be horizontal image, and then, we used the Duplaquet method and the method proposed in this paper to get the seam-lines. Figure 4 shows the results, it can be seen from the pictures of partial enlargement that the best seam-lines searched by the Duplaquet method has changed significantly before and after rotation, the seam-line pass through the edge of buildings before rotation, but the seam line avoid the buildings after rotation. The seam-lines are searched by our algorithm are basically no change before and after rotation, they still along the direction of the road forward, and are very good to avoid the ground buildings. This shows that the traditional methods are sensitive to the direction of images, that is to say the minimum value of energy function is related to the direction of energy aggregation and traversal. Therefore, due to this paper has made specific improvements to the above issues, our algorithm has the advantage of adaptability, and it is more robust. So, it is more suitable to the UAV images mosaicking.

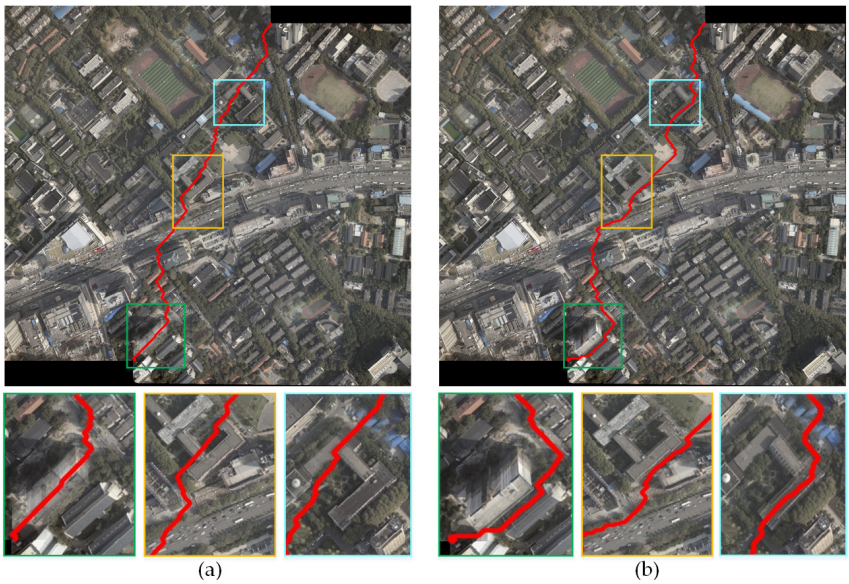


Figure 5. Compared with seam-lines. (a) Duplaquet method; (b) our method

In order to further verify the superiority of our method, the Duplaquet algorithm and our algorithm are used to search the best seam-lines of image pairs in Figure 3(b) with irregular

overlapping areas. Figure 5(a) and 5(b) are the results respectively. It can be seen from Figure 5 that the seam-lines of two data sets have obviously difference with different methods. From the local zoom view of Figure 5, we can find that the Duplaquet method not only appear the seam-lines across the edge of houses, but also there is a ghosting phenomenon appears in Figure 5(a). In this paper, the optimal seam-lines searched by ours are basically forward along the road direction, which avoid across the ground buildings, it will greatly reduce the probability of dislocation and seams as the reason like image matching errors.

4. CONCLUSIONS

This paper reviewed many mainstream image mosaic methods at first, then selected the classical Duplaquet method as the research object, and pointed out the defects of the dynamic programming method based on the Duplaquet algorithm theory. At last, this paper proposed a new dynamic programming algorithm to search the best seam-lines through improved several essential and key problems of the classic Duplaquet method. And furthermore, the superiority and effectiveness of the method proposed in this paper are verified by the comparative experiments of two image pairs with irregular overlapping area. The results are better than the Duplaquet method, and it is proved to be directional independent and better robustness.

ACKNOWLEDGEMENTS

This Study is supported by the National Key Research and development Program of China (2016YFB0502201), the NSFC (91638203), the State Key Laboratory Research Expenses of LIESMARS.

REFERENCES

[1] LI, D. & Li, M., (2014) “Research advance and application prospect of unmanned aerial vehicle remote sensing system”, Geomatics and Information Science of Wuhan University. Vol. 39, No. 5, pp505-513.

[2] Chen, R.; Chu, T.; Landivar, J.; Yang, C. & Maeda, M., (2017) “Monitoring cotton (Gossypium hirsutum L.) germination using ultrahigh-resolution UAS images”, Precision Agric. Vol.18, pp 1-17.

[3] Chen, S.; Lafer, D.; Mangina, E., (2016) “State of Technology Review of Civilian UAVs”, Recent Patents on Engineering, 2016, 10(3): 160-174.

[4] Zhang, W.; Li, M.; Guo, B.; Li, D. & Guo, G., (2017) “Rapid texture optimization of three-dimensional urban model based on oblique images”, Sensors. Vol.17, No. 4, pp911-916.

[5] Li, M.; Li, D. & Fan D., (2012) “A study on automatic UAV image mosaic method for paroxysmal disaster”, International archives of the photogrammetry, remote sensing and spatial information science. Vol. B6, pp123-128.

[6] Dijkstra, E., (1995) “A note on two problems in connexion with graphs”, Numerische Mathematic. Vol. 1, No. 1, pp269-271.

[7] Davis, J., (1998) “ Mosaics of scenes with moving objects”, IEEE Computer Society Conference on Computer Vision & Pattern Recognition, Santa Barbara, USA.

[8] Chon, J.; Kim, H. & Lin, C.,(2010) “Seam-line determination for image mosaicking: A mismatch and the global cost”, ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 65, No. 1, pp86-92.

[9] Zhang, J.; Sun, M. & Zhang, Z., (2010) "Automated seamline detection for orthophoto mosaicking based on ant colony algorithm", Geomatics and Information Science of Wuhan University. Vol. 34, No. 6, pp675-678.

[10] Duplaquet, L., (1998) "Building large images mosaics with invisible seam-lines", Proceedings of SPIE, Vol. 3387, pp369-377.

[11] Xu, Y.; Xing, C. & Chen, X., (2011) "A mosaicking method for UAV sequence images based on seam line", Geomatics and information science of Wuhan University, Vol. 36, No. 11, pp1265-1269.

[12] Cheng, X., (2011) "Research on fast produce of orthophoto with UAV sequence images", Wuhan University.

[13] Luo, S., (2015) "Improved dodging algorithm based on Wallis principle". Geomatics Science and Technology. Vol. 3, pp51-58.

AUTHORS

Ming Li received the Ph.D. degree from Wuhan University, Hubei, China in 2016. He is currently a postdoc researcher at LIESMARS, Wuhan University and an assistant researcher at Collaborative Innovation Center of Geospatial Technology. His research interest includes computer vision, indoor positioning and geographic information application.

Ruizhi Chen received the Ph.D. degree from University of Helsinki, Finland in 1991. He is currently a professor at LIESMARS, Wuhan University and a researcher at Collaborative Innovation Center of Geospatial Technology. His research interest includes computer vision, indoor positioning and so on.

Xuan Liao received the B.S. degree in GIS form China University of Petroleum, Qingdao, Shandong province, China, in 2017. Currently she is pursing the Master degree in Wuhan University, Wuhan, Hubei province, China. Her primary research interest is the application of Remote Sensing and Computer Vision.

Weilong Zhang received the M.S. degree from Xidian University, Shanxi, China in 2015. He is pursing Ph.D. degree in Wuhan University, Wuhan, Hubei province, China. His research interest includes computer vision, photogrammetric survey.

INTENTIONAL BLANK

CLUSTERING BASED LOCAL TONE MAPPING ALGORITHM FOR DISPLAYING HDR IMAGES ON LDR DEVICES

Taeuk Kang, Jeonghyun Lee[1] and JechangJeong[2]

[1]Department of Electronic and Computer Engineering,
Hanyang University, Seoul, Korea
[2]Department of Electronic Engineering, Hanyang University, Seoul, Korea

ABSTRACT

In order to display HDR (High Dynamic Range) images with increased dynamic range on LDR (Low Dynamic Range) monitors, it is necessary to perform a tone mapping technique, which is a process of compressing a dynamic range of an image. Representative techniques are global tone mapping and local tone mapping. Though global tone mapping is simple to compute, it has low local contrast and loses details. Local tone mapping has high local contrast but it demands high computational complexity. In order to take advantage and to compensate of two techniques, we propose local tone mapping based on clustering. Clustering reduces the complexity of local tone mapping implementation by dividing images. In the local tone mapping process, the local adaptation is obtained by combining the cluster-level log mean and global log mean. Using local characteristics, that is local adaptation, based on clustering, the results has high local contrasts and local detail is improved. Experiment result shows that proposed algorithm has better performance than conventional algorithm.

KEYWORDS

High Dynamic Range(HDR) Imaging, Tone Mapping, k-means Clustering,

1. INTRODUCTION

The dynamic range means the ratio of the brightest and darkest values in an image. The human visual system(HVS) has wide dynamic range. A human can perceive luminance ranging from 10^-6 to 10^8 (cd/m2). Consumer display device normally presents luminance up to 500 (cd/m2) [1]. Since it has smaller dynamic range than HVS, it doesn't have the ability to represent bright and dark region simultaneously. To solve this problem, high dynamic range (HDR) imaging technique emerged.

The purpose of HDR imaging technique is to capture all natural luminance range. HDR images are acquired using modern high-quality imaging sensor or exposure fusion technique. HDR image contains luminance up to 10,000 (cd/m2). Because an HDR image can represent a wide dynamic range, they can express dark and bright region similar to the human eye [1]. Becausethe dynamic range of the low dynamic range(LDR) device is smaller than the dynamic range of HDR image, the dynamic range must be compressed in order to display the HDR image on the LDR device. This process is called tone mapping.

David C. Wyld et al. (Eds) : CNDC, DPPR, AIAA, WiMNET, ACITY - 2017
pp. 41– 51, 2017. © CS & IT-CSCP 2017 DOI : 10.5121/csit.2017.71605

Tone mapping is divided into a global tone mapping and a local tone mapping. Global tone mapping is a way to treat all pixels equally using a simple function when processing one image. Since this method handles all pixels in the same way, it has low computational complexity and is easy to design, but it does not take local characteristics into account and loses details corresponding to high frequencies. In addition, the local tone mapping is a method of calculating based on local characteristics when processing an image. This technique has high computational complexity because it considers regional characteristics, but has the advantage of implementing tone mapping while preserving the details of high frequencies. In this case, however, there is a high probability that defects such as a halo artifacts or gradient reversal[2].

In other classification, the tone mapping is divided into two parts, one based on HVS and the other constructed through experiments. As an example, tone mapping algorithm of Reinhard et al. [3] applies a zone system that divides brightness based on human visual characteristics used in photographic techniques. The tone mapping technique of Drago et al. [4] uses adaptive logarithmic function similar to human visual perception to the image. On the other hand, a typical example of experimentally constructed algorithm is the algorithm of Schlick et al [5], and it is a technique that applies quantization to tone mapping.

In this paper, we propose an algorithm that performs tone mapping locally by dividing images into several clusters. The proposed algorithm is implemented based on HVS because it uses adaptive logarithmic mapping [4]. And because it takes into account the regional characteristics, the local contrast and detail can be improved.

The composition of this paper is as follow. Section 2 describes the conventional algorithm. The proposed algorithm is described in Section 3. Section 4 presents and analyzes the experimental results, Section 5 concludes this paper.

2. CONVENTIONAL ALGORITHM

2.1. Adaptive Logarithmic Mapping

The response of HVS about luminance is non-linear. It is approximated with equation(1) that is Weber-Fechner law.

$$B = k_1 ln(\frac{L}{L_0}) \tag{1}$$

L_0 means the luminance of background, k_1 is constant

The response is transformed to use it in image processing as equation (2)

$$L_d = \frac{log\ (L_w + 1)}{log\ (L_{wmax} + 1)} \tag{2}$$

$$L_w = \frac{L}{L_{wa}} \tag{3}$$

$$L_{wa} = exp\ (\frac{1}{N}\sum log\ [\delta + L(x, y)]) \tag{4}$$

L_d means displayed luminance. L_w represents the value that input luminance is divided with L_{wa}. L_{wa} is world adaptation that is global log average. Because L_{wa} is different value corresponding to image, tone mapping function is transformed depends on image. Delta is very small value that prevent the log function to be divergence.

Equation (2) yields the same result for all the bases, since the base of the logarithm is the same. If the base is same, it is deleted in calculation. In order to obtain different results according to the base of the log, it is transformed as the following equation(5).

$$L_d = \frac{log_{bs}(L_w + 1)}{log_{10}(L_{wmax} + 1)} \tag{5}$$

(a)$bs = 2$ (b)$bs = 10$

Figure1. experiment result according to differentbs

If mapping is calculated based on equation (5), we use the fixed base value of bs. You can see the result depends on different base value in the figure 1. In $bs = 2$, the detail was well preserved and the brightness of the dark area was appropriate, and the brightness of the bright area was appropriately compressed when $bs = 10$. In order to maintain two advantages, equation (6) is proposed by modifying equation (5). The output value can be calculated using Equation (6), which is a tone mapping function.

$$L_d = \frac{L_{dmax} \cdot 0.01}{log_{10}(L_{wmax} + 1)} \cdot \frac{log\,(L_w + 1)}{log\left(2 + \left(\frac{L_w}{L_{wmax}}\right)^{\frac{log\,(b)}{log\,(0.5)}} \cdot 8\right)} \tag{6}$$

L_d, L_w, L_{wmax} is the same parameter as used in Eq. (2), and L_{dmax} is the maximum displayable luminance. b means user-defined constant and adjusts the degree of under-transformation. Drago et al. In [4], set the optimal b to 0.85 through a survey of preference for b.

2.2. Rendering HDR Image Using Integrated Global and Local Processing

Shin et al. [6] proposed a tone mapping algorithm that integrates global and local characteristics to overcome the drawbacks of global tone mapping. This algorithm is implemented using a different tone mapping function for each pixel, and the function is obtained by considering both

the global and the regional characteristics around each pixel. It is divided into basic global tone mapping and local tone mapping using block level averaging.

The local tone mapping is performed by referring to the characteristic of the block after forming a block of a certain size based on the current pixel. The block size should be large enough to capture local characteristics, but if it is set too large, it is difficult to determine the correct characteristics and computational complexity increases. Considering these properties, the authors determined the block size to be 3 and implemented the algorithm. To maintain overall contrast, use an average of all images mixed. The mapping function is calculated by equation (7) using mixed mean.

$$L_d = \frac{L(x,y) \cdot \left(1 + \frac{\alpha L_{mb}(x,y)}{L_{max}}\right)}{L(x,y) + C \cdot \left[(1-\alpha)L_{avg} + \alpha L_{mb}(x,y)\right]} \tag{7}$$

$$L_{mb} = exp\,(\frac{1}{N} \sum_{p \in block} log\,[\delta + L_p]) \tag{8}$$

L_{mb} is the block level log mean and is defined by equation (8).L_{max} is the maximum value of the input image luminance, and α is a user parameter that is determined in the range of $0 < \alpha < 1$ and controls the ratio of the global average and the block level average. Figure 2 shows the experimental results for various α. The authors can obtain reasonable results in the range of $0.4 < \alpha < 0.6$ through experiments, and assume that $\alpha = 0.4$ is optimal and implement the algorithm. In the process of calculating block level averages, edge pixels have large impact on the average, resulting in halo artifacts.

Figure 2. experiment result according to alpha

3. PROPOSED ALGORITHM

In this paper, we propose a local tone mapping algorithm based on clustering. In Section 3.1, we introduce the clustering technique that divides the image into several clusters. In Section 3.2, local tone mapping based on clustering is interpreted to perform tone mapping according to the characteristics of each cluster. In Section 3.3, we describe a technique for eliminating boundary artifactresulting from Section 3.2.

3.1. Clustering Algorithm

Clustering is an algorithm that divides the image into several regions. When constructing a cluster, it is called a superpixel to bind a human visual cognition to something similar. In the proposed algorithm, the tone mapping is performed considering the regional characteristics after dividing into several regions using the SLIC (Simple Linear Iterative Clustering) technique.

The SLIC technique generates a superpixel through a process similar to k-means clustering. The SLIC technique is easy to use and can flexibly change the number of clusters to be created. The process of the SLIC begins with defining k corresponding to the size of the superpixel, which is performed in the CIELAB color space. The rest of the process is divided into two phases, initialization and assignment. In the initialization process, initialize information about the center of each cluster and the number of clusters of the corresponding pixel. Then, the assignment step is repeatedly performed to obtain an appropriately divided image.

In the initialization step, C_k and $label(i)$ are initialized. Each cluster is composed of squares with the same number of pixels, andC_k stores information about the center of each cluster. At this time, the length of one side of the square is set to S. Next, the value of $label(i)$ is set to k for all pixels existing in the k-th square.

In the assignmentstep, each information is updated to fit the image, and after a number of iterations, the information of each cluster is determined. The assignment steps follow the flowchart in figure 3. below. This step is repeated until the E corresponding to the previous result and the current result change is smaller than the specified threshold thr while repeating. Each iteration sets the surrounding pixels for the center of all the clusters. The search area of $2S \times 2S$ is set at the center of each cluster and the distance from center is calculated. If the calculated distance is smaller than the minimum distance $d(i)$ stored in the corresponding position, it is determined that the calculated distance is closer to the current cluster than the stored cluster, and the minimum distance $d(i)$ is newly set and the $label(i)$ of the corresponding position is changed.If we performed on all the clusters, we calculate the center for the new clusters, obtain the difference E from the center of the previous clusters, and compare with thr. If it is less than thr, end the iteration and complete the assignment step.

At the end of the assignment phase, the clustered images can be obtained using label information for all locations.And performs local tone mapping that improves local contrast by using clustering information.

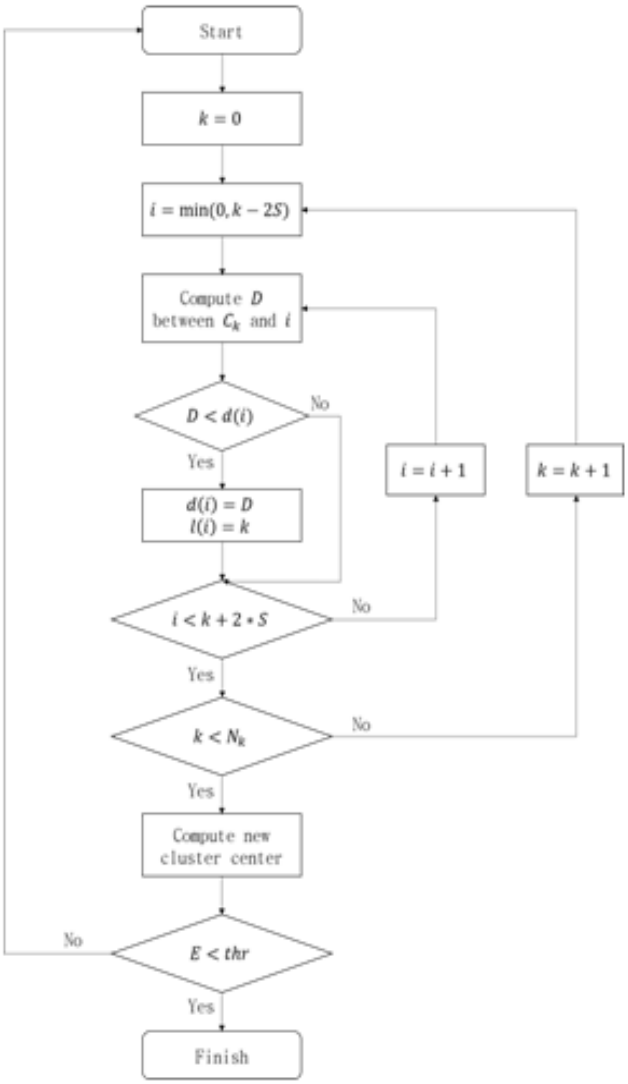


Figure 3. flow char of assignment process

3.2. Clustering Based Local Tone Mapping

In this paper, we propose local log mapping considering regional characteristics for each cluster in the image divided into several clusters. The proposed tone mapping scheme consists of the following flowchart. In the first step, the luminance component is extracted from the color image and the tone mapping process is performed in black and white. The tone mapping process is divided into a global process and a local process. In the global processing, the global log mean of the whole image is calculated. In the process of local processing, the image is divided into several clusters and the local log mean is obtained for each region. Prior to obtaining the local log mean, we used the SLIC technique to divide the image into several clusters that are visually cognitively similar. Find the local log mean for each subdivision.

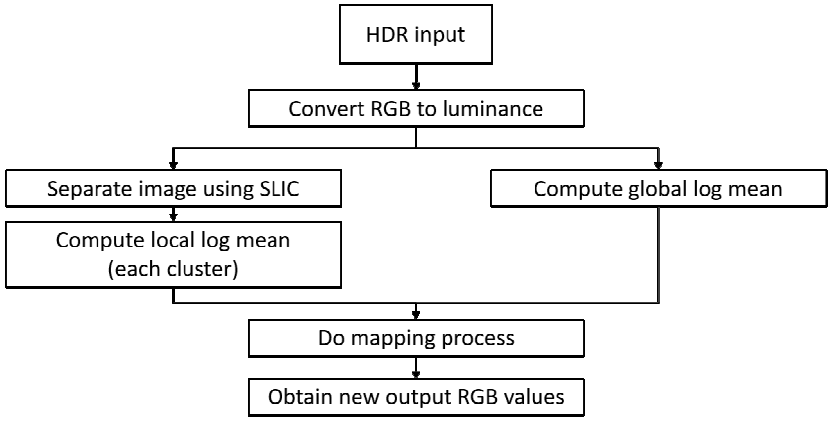


Figure 4. flow chart of proposed algorithm

The mapping process is performed after the global log average and the local log average are calculated. The mapping process is performed using different functions for each region and uses local adaptation, which is a mixture of global log average and local log average. Because local log averages differ from region to region, local adaptation also varies from region to region. When calculating the local adaptation, the global log average plays a role in maintaining overall brightness and the local log average helps to make high local contrast.

$$L_{la} = (1-\alpha)L_{avg} + \alpha L_{Cavg} \tag{9}$$

$$L_{Cavg} = exp\,(\frac{1}{N_C}\sum_{x,y\in cluster} log\,[\delta + L(x,y)]) \tag{10}$$

At each cluster, tone mapping process is computed using equation (11).

$$L_d = \frac{L_{dmax}\cdot 0.01}{log_{10}(L_{wmax}+1)}\cdot\frac{log\,(L_w+1)}{log\left(2+8\cdot\left(\frac{L_w}{L_{wmax}}\right)^c\right)} \tag{11}$$

$$L_w = \frac{L}{L_{la}} \tag{12}$$

c stands for user-defined constant and controls the degree to which the base is deformed. In this case, c has the same role as b used in the algorithm of Drago et al. [4] and has a relation of $c = -\log_2 b$. L_w is divided by L_{wa} in the conventional method, whereas in the proposed tone mapping technique, it is divided by the previously calculated L_{la}. This allows you to implement tone mapping that reflects local characteristics. The result of applying the proposed tone mapping technique is shown in Figure 5.

Figure5. experiment result of proposed algorithm

Experimental results show that details and local contrast are improved, but boundary artifact occur at the boundaries of the cluster. Because they are transformed using different tone mapping functions in different regions, there can be large differences in the boundaries of the clusters and clusters. In Section 3.3, we introduce techniques for removing boundary artifact.

3.3. Dealing with Boundary Artifact

Analysis of the previous experimental results shows that boundary artifact appear at the edge of the cluster. When performing the tone mapping process to remove boundary artifact, the tone mapping function of surrounding clusters is considered. We use the boundary artifact removal technique introduced in the algorithm of Duan et al [7].

This technique is performed as shown in Fig. 6. Using the tone mapping functions of surrounding clusters, the current pixel is input to each function and the weighted sum of the output result is mapped. The weight for each result is measured in inverse proportion to the distance from the current pixel, and the distance is calculated by Euclidean distance between the center of the cluster and the current pixel, and the weight is calculated by equation (13). If we calculate each weight, we get the tone-mapped result through equation (14). All pixels undergo such a transformation that results in the elimination of borderline defects.

$$w_d = exp\left(-\frac{d_n}{\sigma_n}\right) \tag{13}$$

$$d_n = \frac{\sum TMO\big(D(x,y)\big) \cdot w_d(n)}{\sum w_d(n)} \tag{14}$$

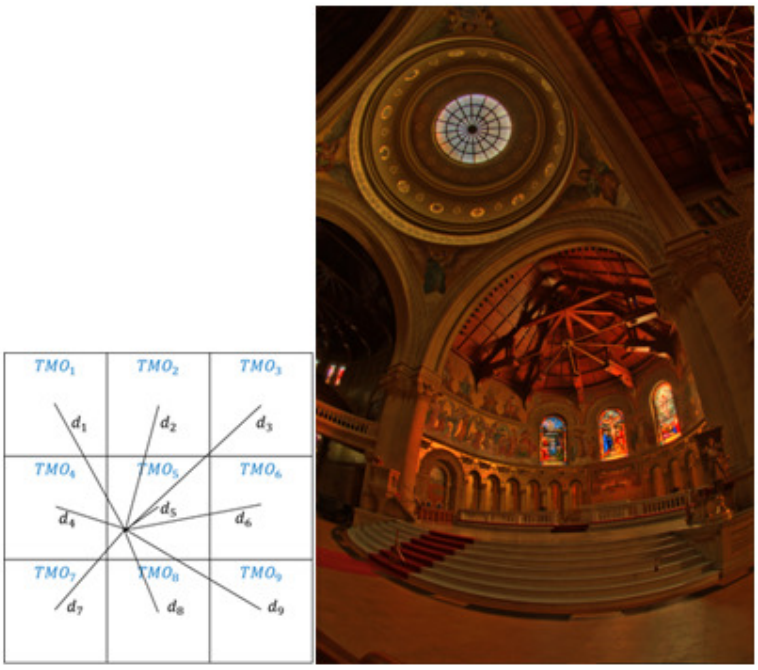


Figure 6. removing boundary artifact. (a) example of calculating weight, (b) experiment result removed boundary artifact.

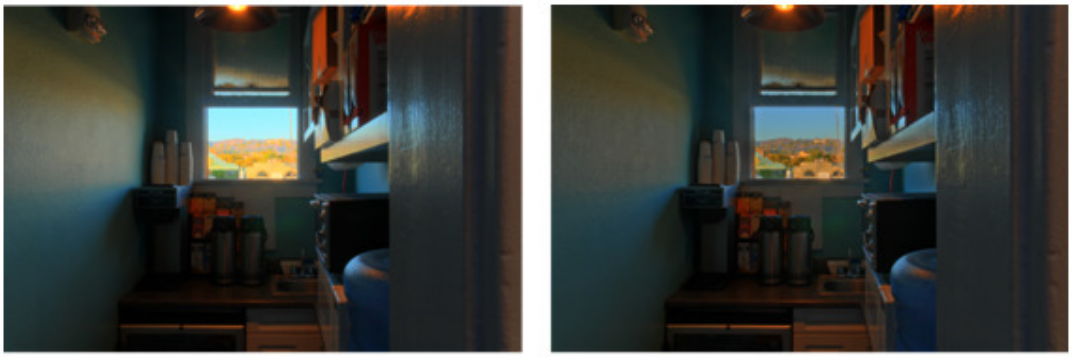
4. EXPERIMENT RESULT

In this section, we compare and analyze the experiment results to compare the proposed tone mapping algorithm with the conventional tone mapping algorithm.All image processing is performed via MATLAB 2015a in a PC environment of Windows10 64-bit operating system Intel Core i5-3470 CPU @ 3.20GHz. In HDR images, there is no objective image quality index commonly used such as PSNR (Peak Signal to Noise Ratio), which is mainly used for image quality comparison of general images. Therefore, it is necessary to rely on subjective image quality comparison. We evaluate the performance of the proposed tone mapping technique through subjective image quality comparison.

Figure 7 compares the experimental results of Drago et al. [4] and the proposed tone mapping algorithm for KitchenWindow.hdr image. The proposed tone mapping preserves more of the local contrast or details. At (a), it can be seen that the overall brightness of the image is uniformly distributed, and the dynamic range is appropriately compressed. However, you can see the shadows in the dark areas. And when you look at the part of the window, it is too bright to know exactly what kind of landscape it is.The result of the proposed tone mapping of (b) shows the result of complementing the disadvantage of (a) mentioned above. First, we got the result of utilizing the detail of the dark part. The letter in the box on the right shelf can be seen more clearly compared to (a). You can also recognize the coffee machine or the stuffs next to it. Finally, when you look at the image on the window side, you can see in detail what kind of scenery is out of the window and color information can be expressed more abundantly.

Figure 8 compares the experimental results for memorial.hdr images. This image has a very bright area of stained glass and a window in the ceiling, and a very dark area in the corner of the ceiling. The results of (a) show that the overall brightness is distributed evenly, but the details are not preserved in very bright or dark areas. Especially, the stained glass on rightmost side can not recognize any picture. Looking at (b), the details were preserved in the corner areas of the ceiling, which is a dark area, and many details were preserved in the stained glass area. The picture of

rightmost stained glass which was not seen in (a) is recognizable. Comparing to (a), it can be seen that more information is expressed.



(a) (b)

Figure 7. subjective quality comparison, (a): the result of Drago et al. [4], (b): the result of proposed algorithm

(a) (b)

Figure 8. subjective quality comparison

(a) (b) (c)

Figure 9. subjective quality comparison

5. CONCLUSION

In this paper, we propose a tone mapping algorithm considering local characteristics after divides image with several regions based on clustering. In order to identify local characteristics, cluster level log mean was calculated for each cluster, and local adaptation was performed by mixing global log mean and local log mean of each cluster, and tone mapping was performed using local adaptation. In order to maintain the visual quality in the tone mapping process, the result image is configured to have a brightness distribution similar to that of human visual system using adaptive logarithmic mapping based on the visual recognition. Experiments have shown that visual quality is maintained similar to adaptive log mapping based on visual cognition, that local contrast is high and local detail is improved. And we show that the proposed tone mapping has superior performance compared with conventional tone mapping algorithm.

ACKNOWLEDGEMENTS

The authors would like to thank everyone, just everyone!

REFERENCES

[1] R. Boitard, M. Pourazad, P. Nasiopoulos, and J. Slevinsky, (2015) “Demystifying High-Dynamic-Range Technology: A new Evolution in Digital Media,” IEEE Consum. Electron. Mag., Vol. 4, No. 4, pp. 72-86.

[2] C. Ha, J. Lee, and J. Jeong, (2016) “Tone Reproduction for High-Dynamic Range Imaging based on Adaptive Filtering,” Opt. Eng., Vol. 53, No.3, p.33103.

[3] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, (2002) "Photographic Tone Reproduction for Digital Images," ACM Trans. Graph., Vol. 21, No. 3, pp. 267-276.

[4] F. Drago, K.Myszkowski, T. Annen, and N. Chiba, (2003) "Adaptive Logarithmic Mapping for Displaying High Contrast Scenes," Comput. Graph. Forum, Vol. 22, No. 3, pp. 419-426.

[5] C. Schlick, (1994) “Quantization Techniques for Visualization of High Dynamic Range Pictures,” Proceeding of the Fifth Eurogrphics Workshop on Rendering, pp. 7-18.

[6] H. Shin, T. Yu, Y. Ismail, and B. Saeed, (2011) “Rendering High Dynamic Range Images by using Integrated Global and Local Processing,” Opt. Eng., Vol. 50, No.11, p.117002.

[7] J. Duan, M. Bressan, C. Dance, and G. Qiu, (2010) “Tone-mapping High Dynamic Range Images by Novel Histogram Adjustment,” Pattern Recognition, Vol. 43, No. 5, pp. 1847-1962.

[8] K. Lee, W. Choe, J. Kwon, and S. Lee, (2009) “Locally Adaptive High Dynamic Range Image Reproduction Inspired by Human Visual System,” Proc. SPIE 7241, pp. 72410T

INTENTIONAL BLANK

NAIVE BAYESIAN FUSION FOR ACTION RECOGNITION FROM KINECT

Amel Ben Mahjoub[1], Mohamed Ibn Khedher[2], Mohamed Atri[1] and Mounim A. El Yacoubi[2]

[1]Electronics and Micro-Electronics Laboratory, Faculty of Sciences of Monastir, Monastir University, Tunisia
[2]SAMOVAR, Telecom SudParis, CNRS, University of Paris Saclay, France

ABSTRACT

The recognition of human actions based on three-dimensional depth data has become a very active research field in computer vision. In this paper, we study the fusion at the feature and decision levels for depth data captured by a Kinect camera to improve action recognition. More precisely, from each depth video sequence, we compute Depth Motion Maps (DMM) from three projection views: front, side and top. Then shape and texture features are extracted from the obtained DMMs. These features are based essentially on Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP) descriptors. We propose to use two fusion levels. The first is a feature fusion level and is based on the concatenation of HOG and LBP descriptors. The second, a score fusion level, based on the naive-Bayes combination approach, aggregates the scores of three classifiers: a collaborative representation classifier, a sparse representation classifier and a kernel based extreme learning machine classifier. The experimental results conducted on two public datasets, Kinect v2 and UTD-MHAD, show that our approach achieves a high recognition accuracy and outperforms several existing methods.

KEYWORDS

Action recognition, Depth motion maps, Features fusion, Score fusion, Naive Bayesian fusion, RGB-D.

1. INTRODUCTION

The field of action recognition has been considered as an active challenging domain in computer vision research for more than two decades. It is necessary for several applications such as intelligent video surveillance, robot control, video understanding, healthcare, etc. In the past few years, further investigations [1–4] have been initially focused on recognizing actions from RGB video sequences recorded by traditional 2D cameras. Recently, the emergence of low-cost RGB-D cameras, such as Microsoft Kinect v2, has gained much attention in computer vision thanks to its excellent accuracy in action recognition. Kinect v2 provides RGB and depth data modalities. It has been used to improve the performance of human action recognition systems. The rapid development of such cameras has opened the door to a rich representative work [5–10] in learning and recognizing actions based on depth video sequences. Depth maps have various

David C. Wyld et al. (Eds) : CNDC, DPPR, AIAA, WiMNET, ACITY - 2017
pp. 53– 69, 2017. © CS & IT-CSCP 2017 DOI : 10.5121/csit.2017.71606

advantages compared to traditional color videos. First, they are insensitive to change in lighting conditions. Second, they provide a three-Dimensional (3D) structure and shape information that improves the distinguish ability of different poses. These innovations have been behind producing a lot of multimodal datasets dedicated to human action recognition systems. [11] described most RGB-D datasets currently exploited in recognizing actions. Three levels of information fusion have shown an improvement in accuracy: (i) data level, where data from several sensors can be integrated to supply new data; (ii) feature level, where the different feature sets extracted from a data source are fused to create a new fused feature vector; and (iii) decision level, where the fusion of multiple classifiers is used to make the final classification decision.

This paper addresses how to enhance recognition accuracy using feature and score fusion levels. First, three Depth Motion Maps (DMMs) [7] are computed in order to represent each action video sequence. Next, the description of the obtained DMMs is performed on the basis of Histogram of Oriented Gradients (HOG) [12] and Local Binary Patterns (LBP) [13] descriptors that encode contour and texture depth features. The HOG-LBP feature fusion approach is applied to carry out a compact DMM representation. To get action prediction outputs from the feature variables, we train three classifiers: Collaborative Representation Classifier (CRC) [10, 14], Sparse Representation Classifier (SRC) [15, 16] and Kernel based Extreme Learning Machine (KELM) [17]. These techniques are among the most widely used methods in the literature [9,10,14,15,18,19], as they have shown good performances for activity recognition systems, but as far we know, this is the first time that these three classifiers are fused together to classify action. Finally, we consider a Naive-Bayes approach to combine the classification scores, that shows an improvement in the accuracy of human action recognition when tested on publicly available datasets [14] [20]. The naive Bayesian approach is a commonly known methodology for classifier output fusion, is proposed in various works as [21–23]. Our experimental results substantiated that our proposed human action recognition approach performs better than various state-of-the-art methods.

The rest of the paper is organized as follows. In section 2, a state of the art of human action recognition methods is presented. Section 3 describes the DMM as well as our proposed fusion and classification approaches. The experimental results are presented in section 4. Section 5 includes a conclusion and perspectives.

2. STATE OF THE ART

Several recent action recognition approaches have been presented recently [24–26]. Earlier, action recognition data was provided from an RGB camera. However, human activity recognition from color video sequences has many difficulties such as illumination changes and variations in human appearance.

Recently, by the appearance of depth cameras like Microsoft Kinect, several RGB-D-based human action recognition methods have been developed, as reviewed in [27–29]. The Kinect sensor captures data as color and depth information. In the literature, these provided RGB-D data in addition to skeleton joints have been well explored to improve human action recognition.

In [5], the authors define Space-Time Occupancy Patterns (STOP) to represent 3D depth maps. Both space and time axes are divided into several segments to present a 4D grid. An occupancy feature, calculated in each grid cell, represented the number of occupied space-time points. The

occupancy values of all cells formed STOP feature vectors. A nearest neighbor classifier was used to recognize human actions.

Wang and Lie [6] extracted the Random Occupancy Pattern (ROP) from depth sequences by considering a 3D depth sequence as a 4D shape. ROP features were calculated by applying a weighted sampling scheme founded on rejection sampling. An elastic-net regularized model was utilized to choose the most discriminative features to train red a Support Vector Machine (SVM) classifier for action recognition.

In [7], the authors proposed shape and motion-action representations. Each 3D depth frame was projected into three 2D maps and then the difference between two consecutive maps yielded a motion energy. The concatenation of all these motion energies over the video give out the DMM. The HOG was computed from front, side and top DMM maps as a distribution of local intensity gradients. The DMM-HOG features were matched by an SVM classifier for action recognition.

Oreifej and Liu [8] introduced the histogram of oriented 4D normals which was the extension of histograms of oriented 3D normals [30] by appending the time derivative. The depth, time and spatial coordinates of a 4D space were quantified by a regular polychoron, and then each human action was modeled as a distribution of the normal surface orientations.

Moreover, Chen [9] presented a DMM based on an LBP descriptor for human action. The DMM was presented from three projection views to characterize the 3D local motion. The LBP features were extracted from the depth maps to measure the local image texture by encoding each pixel with decimal numbers. All the extracted LBP features from each projected DMM were concatenated to give a single feature vector, used to train a KELM classifier.

Farhad and Jiang in [10] developed a new descriptor that computed HOG features from DMMs based on contourlet sub-bands. A Contourlet Transform (CT) combined the Laplacien pyramid and the directional filter bank technique to decompose the DMMs into low-frequency and high-frequency sub-bands. This method was used to decrease the noise and clearly present the depth shape information at several scales and directions. Afterwards, HOG features were extracted from these DMM contourlet sub-bands. The combination of the histograms obtained from the three depth views provided the final DMM-CT-HOG feature vector, trained by the CRC to classify human action.

The work of [31, 32] was inspired by the success of deep learning in human action recognition. A new deep learning based action recognition framework using depth and skeleton data was defined in [32]. The deep convolutional neural network was used to extract the spatio-temporal features from depth sequences. A jointVector feature was obtained by computing the angle and position between skeleton joint information. A SVM classifier matched high-level and jointVector features separately to get class probability vectors. The fusion of these two kinds of weighted vectors gives final action recognition.

Ivan in [33] introduced an approach of body-pose estimation based on RGB-D video sequences to recognize complex human activities. Two geometric and motion descriptors were applied to each RGB-D datum to encode respectively the spatial configuration and dynamic features of every body-pose. A hierarchy of three levels was modeled to produce the global activities prediction. At the first level, each activity was decomposed into atomic actions. The intermediate level

represented the atomic human action with sparse composition. At a higher level, the complex human activities were described based on spatio-temporal compositions of atomic actions.

Various publications have appeared in the recent years demonstrating the importance of fusion methods in improving the accuracy of action recognition systems. One of the examples for decision-level fusion was presented in [14]. The authors extracted feature vectors from depth, skeleton and inertial data. Next, they matched three CRC classifiers separately to get the video label. Finally, a Logarithmic Opinion Pool (LOGP) was carried out for decision-level fusion. Imran and Kumar in [34] suggested a new human action recognition method based on classification fusion. Four deep convolutional neural networks were utilized to classify the Motion History Image (MHI) vector descriptors from RGB data and three DMM vectors from depth sequences. The fusion of the output scores of these four networks was done using average and product rule approaches. The Dempster-Shafer (DS) method was put forward in [35] for decision level fusion. In the latter work, both depth and inertial data were exploited to extract feature vectors. The authors used the DS technique to fuse the decision outputs of two CRC classifiers. The authors in [18] described three DMMs by CT-HOG, LBP and Edge Oriented Histograms (EOH) feature descriptors. Then, they used three KELM classifiers to make a decision for each feature vector. The LOGP and majority voting methods were proposed to combine the outputs of classifiers in order to improve activity recognition accuracy. A probabilistic classification fusion approach utilizing a Bayes formalism was performed in [23]. Multiple Hidden Markov Models (HMM) classifiers were matched given the features from accelerometer sensors placed on the body. The naive Bayesian fusion technique was executed to concatenate the classification output vectors from all HMMs.

3. PROPOSED APPROACH

In this section, we present our approach which is broadly described in Figure 1. In order to identify the action of a person in the depth video sequence, we firstly extract the shape and texture features using HOG and LBP descriptors. The key idea of our method is to fuse these two kinds of feature vectors before classification. A dimensionality reduction is secondly performed based on the Principal Component Analysis (PCA) technique. Thirdly, the training is done by applying CRC, SRC and KELM classifiers to get different classification score outputs. Finally, we fuse the probabilistic information sources via a Naive-Bayes technique to output the label of the sequence.

4. FEATURE EXTRACTION

Two features are extracted to describe the DMM images : 1) HOG and 2) LBP.

4.1 Depth Motion Map

The DMM [7] is used to characterize the 3D structure and shape information from depth maps. A depth video sequence contains M frames, where each frame is projected over three orthogonal Cartesian planes to build DMM_f , DMM_s and DMM_t images from front, side and top views, respectively. The computation of motion energy is performed by subtracting the consecutive maps from each projected DMM. The sum of motion energy over the video sequence yields the $DMM_{f,s,t}$ as follows:

$$DMM_{f,s,t} = \sum_{i=o}^{M-1} (DMM_{f,s,t}^{i+1} - DMM_{f,s,t}^{i}| > \epsilon) \quad (1)$$

where f,s,t are the front, side and top views and $\in$ 2 is a threshold. For each projected map, the entire sequence frames are not used but just their extracted regions of interest. These extracted foreground DMMs are normalized to generate the final DMM features. Figure2 shows an example of three projected maps of a right hand high wave depth sequence.

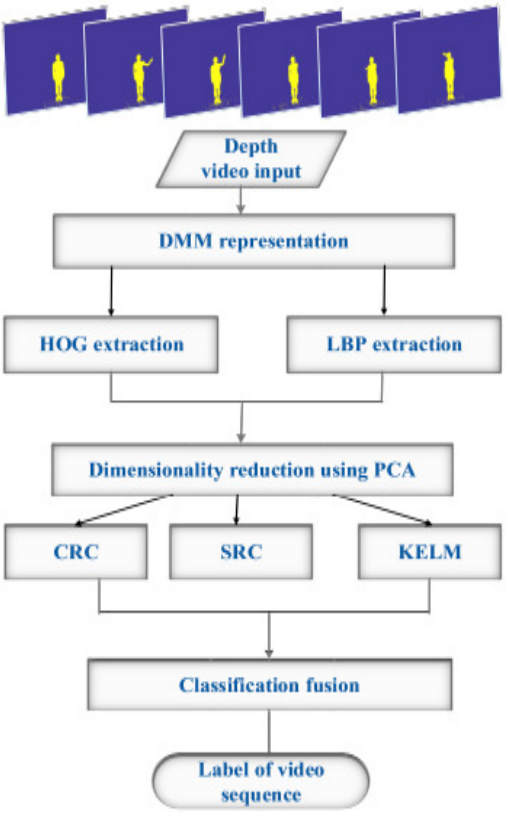


Figure 1. Flowchart of our approach

Figure 2. DMM views: DMM_f, DMM_s and DMM_t

4.2 Histogram of Oriented Gradients

The HOG was introduced in [12] and was used to encode the local appearance and shape on DMM maps [7] with the distribution of local intensity gradients or edge directions. After finding the object using depth information, the idea is to calculate the occurrences of discretized gradient orientations in the depth local region to represent the body shape and motion information. We divide every projected depth map into 8x8 non overlapping cells, where each cell has nine orientation bins. The pixels of these cells throw a weighted vote for an orientation histogram based on the value of the gradient magnitude to yield a histogram of nine gradient directions. It results in three HOG vectors that describe map features from front, side and top DMMs. These vectors are concatenated to produce a 6,588 dimensional final DMM-HOG descriptor for the entire action video sequence.

4.3 Local Binary Patterns

The first LBP encoding schemes were proposed for quantifying the local image contrast. Ojala in [13] extended the LBP to an arbitrary circular derivation to describe the local texture pattern. The Computation of the texture can be carried out by thresholding a neighborhood by the gray value of its center and by assigning decimal numbers to the pixels of the image. Let g_c be a scalar value of the center pixel and g_P(**P=0, ... P-1**) be the gray values of its neighborhood of P which are pixels equally spaced on a circle with a radius R. This circle constitutes a circular symmetric whole of its neighbors. The LBP feature is established by subtracting the g_P neighbors from the center value g_c to produce a P digit binary number converted to a decimal form as follows:

$$LBP_{P,R} = \sum_{p=o}^{P-1} S(g_P - g_c)2^P \tag{2}$$

where P is the number of neighborhood pixels and S(x) is

$$S(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

We then obtain 2^P uniform patterns. The evaluation of a histogram over an image, which represents the frequency of each occurring number, is generated to encode the texture information. In this paper, we use the LBP operator to extract features from the DMM maps as in [9].

4.4 Feature fusion

Recently, the use of information fusion has attracted the attention of researchers in the action recognition domain as a consequence of its greater accuracy. The concatenation of feature vectors to improve the system performance is a simple and traditional method frequently used in the literature. The feature vectors of different kinds have been concatenated together to find a single long feature vector to train the classifier. In [36], Wang and Han employed this fusion approach to combine HOG and LBP feature vectors for human detection. Dimitrovski and Kocev [37] performed the low level feature fusion approach to describe medical images. Several extracted features have been concatenated in a single feature vector before the classification step. Local

feature and boundary based shape features were concatenated in [38] to improve the recognition performance of the object class. In [39], the authors started by extracting the HOG and LBP features. After that, they applied PCA to each of them to reduce the dimension. Finally, the concatenation of the two obtained feature vectors was performed to give the mixed HOG-LBP descriptor.

In our work, we extract HOG and LBP features from the DMM to represent the depth sequence from diverse prospects. We then apply the fusion algorithm based on the PCA to concatenate DMM-HOG and DMM-LBP feature vectors.

5. CLASSIFICATION ALGORITHM

5.1 Collaborative Representation Classifier

The CRC has been employed in various work [10, 14] owing to its performance and efficiency in classification. Given c as the number of classes, we have n training samples belonging to c classes, denoted $\mathbf{X= [X_1, X_2, ..., X_c]} = [x_1, x_2, ...,x_n]$, where $\mathbf{X} \in \mathbf{R}^{d \times n}$ and the total number of training samples is $n= n_1 + n_2 + ... + n_c$, n_k being the number of samples pertaining to class k. The test sample y is described in the CRC as a linear association: $\mathbf{y = \alpha X}$, where $\alpha = [\alpha_1, \ldots, \alpha_c]$ represents the coefficient vector of the corresponding training sample. We then apply the l_2 norm to optimize α by a minimizing formulation as follows:

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\alpha} \| \mathbf{y} - \mathbf{X}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\mathbf{L}\|_2^2 \tag{3}$$

Where λ is a parameter of regularization and L is the Tikhonov regularization matrix that represents the distance weighted matrix by giving less weight to the training samples dissimilar from the test samples as follows:

$$\mathbf{L} = \begin{bmatrix} \|\mathbf{y} - \mathbf{x}_1\|_2 & 0 & \ldots & 0 \\ 0 & \|\mathbf{y} - \mathbf{x}_2\|_2 & \ldots & 0 \\ 0 & 0 & \ldots & \|\mathbf{y} - \mathbf{x}_n\|_2 \end{bmatrix} \tag{4}$$

The Tikhonov regularization is used to solve this minimization problem:

$$\hat{\boldsymbol{\alpha}} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{L}\mathbf{L}^T)^{-1}\mathbf{X}^T\mathbf{y} \tag{5}$$

A minimization of the reconstruction error is made to take a classification decision.

$$label(\mathbf{y}) = \arg\min_{k} e_k(\mathbf{y})) = \| \mathbf{y} - \boldsymbol{X_k}\hat{\boldsymbol{\alpha}}_k\|_2 \tag{6}$$

where $k \in [1, 2, ...c]$ and e_k is the residual error.

5.2 Sparse representation

Human action recognition utilizing SRC is inspired by the work of Wright and Yang [15] which used sparse representation to recognize faces. In the SRC, the testing samples are obtained by a sparse linear combination of the training samples. The unknown sample is identified by finding the label with the lowest residual error. Given a matrix X of training samples for c classes and an optional error tolerance $\epsilon > 0$, the test sample of the kth class y is represented from the training set $\mathbf{X}_k$ with the coefficients $\boldsymbol{\alpha}_k$. According to the sparse representation of y in terms of dictionary constructed from training samples of all c classes, we can retrieve α_k as follows :

$$\begin{aligned} &\arg\min_{\boldsymbol{\alpha}} \| \boldsymbol{\alpha} \|_1 \\ &\text{subject to } \| \mathbf{X}\boldsymbol{\alpha} - \mathbf{y} \|_2^2 \leq \epsilon \end{aligned} \tag{7}$$

Subsequently, we classify y for k= 1, 2, ..., c as follows:

$$identity(\mathbf{y}) = \arg\min_k (r_i = \| \mathbf{y} - \mathbf{X}_k \boldsymbol{\alpha}_k \|_2^2) \tag{8}$$

The calculation of identity (y) defines the label of the test sample y from all distinct c classes.

5.3 Kernel based extreme learning machine

KELM was developed in [17] to solve regression and multiclass classification tasks. An extreme learning machine was initially dedicated to match a Single-hidden Layer Feedforward Neural Network (SLFNN) in order to overcome the learning slowness. We have the n training samples $\{\mathbf{x}_i, l_i\}_{i=1}^n$ where $\mathbf{x}_i \in \mathbf{R}^d$ and $l_i \in [1 \ldots c]$ is it corresponding label. $\mathbf{t_i} = [t_{i1}, \ldots, t_{ic}]^T$ is the network target binary vector that denotes the sample belonging, where only one component is non null. For example, if $t_ik = 1$, it implies that the sample belongs to class k. The responses of the SLFNN to $\mathbf{x}_i$, $\mathbf{h}_i = [h_{il}, \ldots, h_{ic}]^T$ is:

$$h_{ik} = \sum_{j=1}^{D} \boldsymbol{\alpha}_{kj} f(\mathbf{w}_j \mathbf{x}_i + e_j) \tag{9}$$

where D is the number of the hidden nodes, f(.) is a linear activation function for the network output layer, $\mathbf{w}_i \in \mathbf{R}^d$ and $\boldsymbol{\alpha}_{ki} \in \mathbf{R}^D$ are the weight vectors that connect ith hidden node to the input and output nodes respectively and e_i is the bias for the ith hidden node. For all n equations we have: $\mathbf{h} = \boldsymbol{\alpha}\ \mathbf{F}$, where $\boldsymbol{\alpha}$ is the network output weights $\in \mathbf{R}^{D \times c}$ and $\mathbf{F}$ is the matrix of hidden layer outputs of all training sets x_i, which is written as:

$$\mathbf{F} = \begin{bmatrix} f(\mathbf{w}_1 \mathbf{x}_1 + e_1) & \ldots & f(\mathbf{w}_D \mathbf{x}_1 + e_D) \\ \ldots & \ldots & \ldots \\ f(\mathbf{w}_1 \mathbf{x}_n + e_1) & \ldots & f(\mathbf{w}_D \mathbf{x}_n + e_D) \end{bmatrix}$$

$T = [\ t_1,..,\ t_n\ ]^T$ is the matrix that contains the network target vectors. The output weights α is analytically computed as:

$$\boldsymbol{\alpha} = \mathbf{F}'\mathbf{T}^T \tag{10}$$

where $\mathbf{F}'$ is the Moore-Penrose generalized inverse of the matrix **F:**

$$\mathbf{F}' = \mathbf{F}(\mathbf{F}\mathbf{F}^T)^{-1} \tag{11}$$

A positive regularization term $1/\rho$ is added to the diagonal elements of $\mathbf{F}\mathbf{F}^T$, so we have:

$$\boldsymbol{\alpha} = (\mathbf{F}\mathbf{F}^T + \frac{\mathbf{I}}{\rho})^{-1}\mathbf{F}\mathbf{T}^T \tag{12}$$

The kernel matrix for the ELM is used as follows:

$$\Omega_{ELM} = \mathbf{F}\mathbf{F}^T: \Omega_{ELMj,s} = f(\mathbf{x}_j).f(\mathbf{x}_i) = K(\mathbf{x}_j, \mathbf{x}_i).$$

Therefore, the output of KELM classifier is :

$$h(\mathbf{x}) = \begin{bmatrix} K(\mathbf{x}, \mathbf{x}_1) \\ ... \\ K(\mathbf{x}, \mathbf{x}_n) \end{bmatrix} (\frac{\mathbf{I}}{\rho} + \Omega_{ELM})^{-1}\mathbf{T}^T \tag{13}$$

where $\mathbf{I} \in \mathbf{R}^{D \times D}$ is an identity matrix .

The predicted class label of the testing sample $\mathbf{y} \in \mathbf{R}^d$ is the index number of the network output node which has the highest value. Considering $\mathbf{f}_k(\mathbf{y})$ as the output function of the kth hidden node, where $\mathbf{f}(\mathbf{y}) = [f_1(\mathbf{y}),\ldots, f_c(\mathbf{y})]T$, the predicted class of y is calculated as :

$$label(\mathbf{y}) = \arg\max_{k=1...c} f_k(\mathbf{y}) \tag{14}$$

5.4 Classifier fusion

The fusion of various classifiers is a known robust technique as it is usually more robust and accurate than a single learner system. For benchmarking, we consider nine fusion approaches. The Majority vote serves in collecting all the votes of the different classifiers and selecting the label that is the most frequently occurring value. The maximum approach chooses the most confident classifier with the highest classification score. The sum function calculates the sum of score output elements of classifiers and outcomes the label with the highest value. The minimum method gives the class which has a minimum objection by different classifiers. The mean of the output classifier scores consists in choosing the label with the highest mean value. The product fusion technique consists in multiplying the vector elements of the classifier score outputs, and the final decision corresponds to the class with the highest probability. The Decision Template (DT) is a simple and robust classifier fusion method that compares the classifier output combination with a representative template for each class. The decision templates are the

averages of all classifier decision outputs during the training step that ties to the belonging of training samples to each class. These templates are later used to output the final class based on similarity measure. The Dempster-Shafer (DS) technique is like the DT and we can place both methods in the same group. The decision templates are generated from the training data to represent the most characteristic decision profile for each class. In the testing step, the DS method compares the DT to give the label with the largest similarity.

Naive-Bayes is a powerful technique for combining the confidence outputs of different classifiers [21, 22]. This method consists in calculating the a posteriori probability of each possible class w_k given the output labels s_i of the different classifiers. We assume that we have c classes and L classifiers D_i match the data $\mathbf{y} \in \mathbf{R^d}$. Each classifier generates a label s_i, $i \in [1,L]$, so we have the output vector $s = [s_1,\ldots s_L]$. We define $p(s_i)$ as the probability that classifier D_i labels $\mathbf{y}$ in class s_i. Naive Bayesian approach computes the a posteriori probability that y is labeled as w_k as follows:

$$p(w_k/\mathbf{s}) = \frac{p(\mathbf{s}/w_k)p(w_k)}{p(\mathbf{s})} \qquad k = 1...c \tag{15}$$

where $p(w_k)$ is the a priori probability of the hypothesis w_k, $p(\mathbf{s} / w_k)$ is the likelihood and $p(\mathbf{s})$ is the evidence used for normalization, which can be neglected. The equation that describes the support for class w_k can be written as :

$$\mu_k(\mathbf{y}) \propto p(w_k) \prod_{i=1}^{L} p(s_i/w_k) \tag{16}$$

The $\mathbf{c^*c}$ confusion matrix CM^i is defined for each classifier D_i where cm^i_{k,s_t} is the number of data elements having a true class w_k , and labeled by D_i as class w_s. Assuming that the training dataset $\mathbf{X}$ contains a total of n elements and n_s elements from class w_s, the probability for class w_k is given by $\frac{n_k}{n}$ and the probability $p(s_i/w_k)$ can be written in the form $\frac{cm^i_{k,s_t}}{n_k}$. In this way and according to (16), we obtain:

$$\mu_k(\mathbf{y}) \propto \frac{1}{n_k^{L-1}} \prod_{i=1}^{L} cm^i_{k,s_t} \tag{17}$$

The highest value $\mu_k(\mathbf{y})$ is used to label y in class w_k

6. EXPERIMENTAL RESULTS

In order to verify the validity of our method, we have carried out experiments based on the two Kinect v2 and UTD-MHAD datasets.

Table 1. UTD-MHAD dataset actions

Arm cross	Arm curl
Baseball swing	Basketball shoot
Bowling	Boxing
Catch	Clap
Draw circle CCW	Draw circle CW
Draw triangle	Draw X
Jog	Knock
Lunge	Pickup and throw
Push	Sit to stand
Squat	Stand to sit
Swipe left	Swipe right
Tennis serve	Tennis swing
Throw	Walk
Wave	

6.1 Kinect v2 dataset

We harnessed a public multimodal dataset defined in [14], with collected data from a Kinect v2 camera and a wearable inertial sensor simultaneously. This database contains three modalities of depth, skeleton joint position and inertial signals. In this paper, we use the depth images modality. As shown in Figure 3, the dataset includes the following ten actions: right hand high wave, right hand catch, right hand high throw, right hand draw x, right hand draw tick, right hand draw circle, right hand horizontal wave, right hand forward punch, right hand hammer and hand clap (two hands). These actions were effected by three female subjects and three male ones and each subject rehearsed the actions five times. Consequently, we obtained 300 depth video sequences in total. We performed the subject-specific experiment as in [14] to divide the data into training and testing steps. For each subject, the first two repetitions were chosen in training, and the remaining repetitions in testing. Table 2 compares our proposed approach with that used in [14], where the authors applied the DMM and CRC techniques to recognize actions from depth samples.

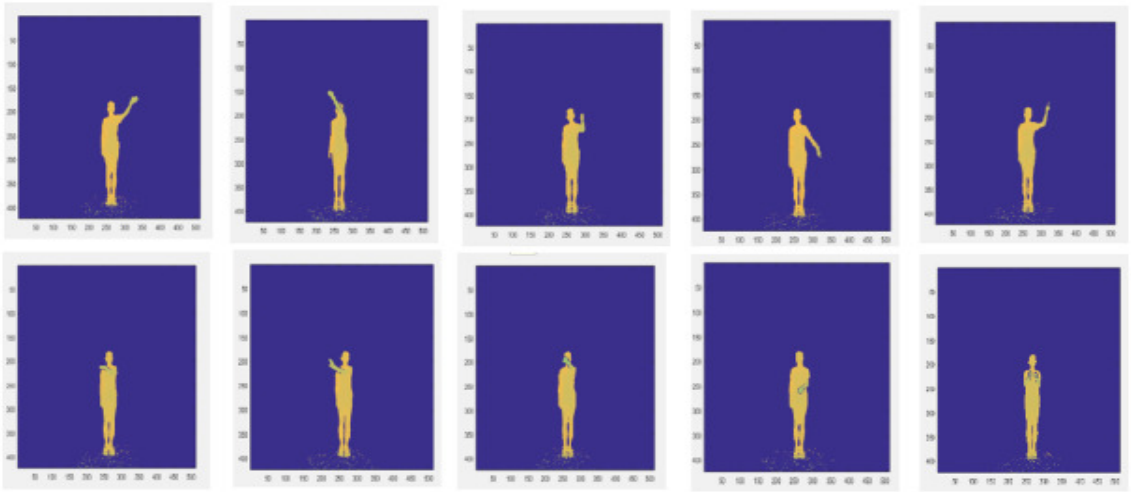


Figure 3. Kinect v2 dataset depth images

6.2 UTD-MHAD dataset

The UTD-MHAD dataset, described in [20], is acquired by a Kinect camera and an inertial sensor that collect multimodal data. UTD-MHAD encloses 27 actions which are mostly an arm or a leg based activity, as shown in Table 1.

Table 2. Results comparison on kinect v2 dataset

Approach	Description	Recognition rate (%)
[14]	DMM	79.4 %
Our method	classification fusion	93.9 %

Table 3. Results comparison on UTD-MHAD dataset for the half-subject experiment

Approach	Description	Recognition rate (%)
[20]	Kinect(only)	66.1
[40]	DMM	73.4
[41]	GF + LF	84.89
[34]	Depth + deep CNN	87.9
[18]	Decision-level fusion (LOGP)	88,4
Our	Classification fusion	90.5

Each action was carried out by four females and four males, and each subject performed four repetitions, so we have 861 video sequences in total after eliminating three corrupted videos. The UTD-MHAD is a multimodal database that includes color and depth videos, skeleton joint positions and inertial sensor signals (acceleration, angular velocity and magnetic strength). We executed two types of experiments on the UTD-MHAD dataset. The first is nominated half-subject where subjects 1, 3, 5 and 7 were used for training and subjects 2, 4, 6 and 8 for testing. The comparison of the existing work with our approach using the UTD-MHAD dataset for the half-subject experiment is illustrated in Table 3. The second experiments are the subject-specific settings in [19]. As each subject performs an action four times, the first two repetitions are used for training and the two remaining repetitions for testing. Table 4 depicts our obtained results compared to the method implemented in [19] based on the Kinect depth feature only.

6.3 Evaluation protocol

The experiment was carried out using a computer i7 3.4GHZ with a RAM of 16 GB.

Table 5 lists the recognition rate results of nine methods for score level fusion of the CRC, the SRC and the KELM. Figure 4 reports the classification results using the SRC, the CRC and the KELM each alone and then the fusion of these three classifiers based on the naive-Bayes fusion method.

Table 4. Results comparison on UTD-MHAD dataset for the subject-specific experiment

Approach	Description	Recognition rate (%)
[19]	Kinect (only)	85.1
Our method	classification fusion	91.6

6.4 Discussion

Recognizing human actions with a good recognition rate is an key computer vision requirement. In our paper, we propose a fusion approach based on Kinect v2 and UTD-MHAD datasets to improve accuracy. In the first step, we start by extracting the HOG and the LBP from the DMM representation of depth video sequences. Then, we concatenate these features using PCA to reduce dimension. Finally, the naive Bayesian approach is applied to fuse the classification score outputs of the CRC, the SRC and the KELM classifiers. As detailed in Table 5, we test different fusion methods on the Kinect v2 and UTD-MHAD datasets. These findings point to the usefulness of naive-Bayes as a score level fusion approach as they give a great recognition rate for both datasets. Figure 4 highlights how important our score level fusion method is for improving the recognition rate compared to approaches using each classifier alone. It is apparent from Table 2 that our method on the Kinect v2 dataset outperforms previous methods [14] by around 15%. The results of the two types of experiments on UTD-MHAD can be seen in Table 4 and Table 3. In Table 4, our technique for the subject-specific experiment demonstrates a clear advantage over the work in [19], which improves the accuracy by 6.5%. We observe from Table 3 that our fused method for the half-subject experiment outperforms previous work based on depth data with an accuracy of 90.5%. We are aware that our work may have two limitations. The first is that we have not considered conjointly the depth skeleton joint position and inertial multimodalities data and the fusion between them which can improve the classification rate. The second negative factor regarding our algorithm is that it has a limit with real-time constraints. An acceleration of our action recognition method can be suggested by using field-programmable gate array or graphics processing unit to speed up the application.Despite this, we can still state that our approach outperforms several previous methods based on only depth data, and it can be employed in computer vision applications.

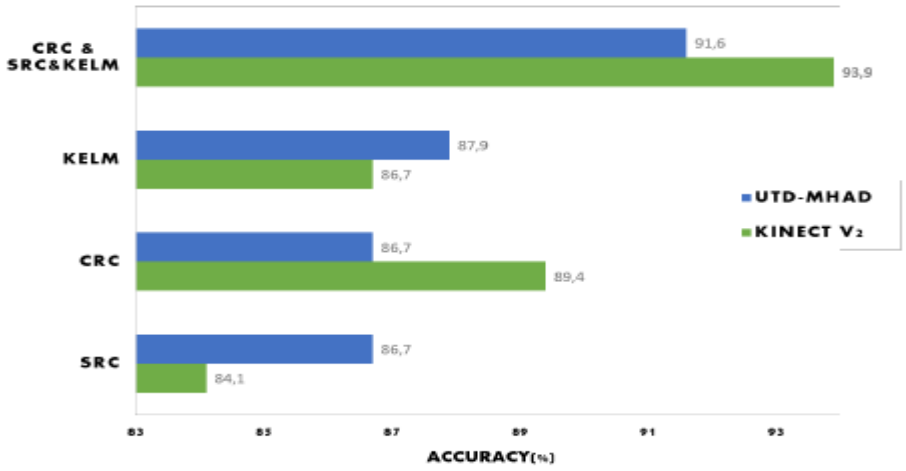


Figure 4. Recognition accuracy of our proposed approach

Table 5. comparison of fused methods of CRC, SRC and KELM classifiers

fused method	accuracy on kinect v2(%)	accuracy on UTD-MHAD (%)
Majority vote	87.8	86.9
Maximum	87.8	86.9
Minimum	88.9	87.9
Sum	86.7	87.4
Average	88.9	87.4
Product	68.9	79.8
Decision template	86.7	89.3
Dempster-shafer	85.5	85
Naive-bayes	93.9	91.6

7. CONCLUSION

This paper has given an account of our proposed probabilistic score level fusion based on the Bayes theorem for human action recognition tested on the Kinect v2 and UTD-MHAD datasets. We have exploited the depth video sequence in these datasets to calculate the DMM. To represent the DMM, we have used the HOG and LBP feature descriptors. The concatenation of the DMM-HOG and the DMM-LBP using the PCA technique has been then performed. Finally, we have applied a naive Bayesian approach to fuse the SRC, CRC and KELM classification scores, that has been shown to outperform different other fusion methods. Our results indicate that our system presents a good recognition accuracy compared to existing work. Future work will focus on multimodalities fusion at data, feature or score levels. We will also develop a co-design architecture to speed up the system.

ACKNOWLEDGEMENT

This work was carried out at the laboratory of Electronics and Microelectronics at the Faculty of Sciences of Monastir, Tunisia and the laboratory of SAMOVAR at Telecom SudParis, France.

REFERENCES

[1] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld, “Learning realistic human actions from movies,” in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 1–8, IEEE (2008).

[2] J. Sun, X.Wu, S. Yan, L.-F. Cheong, T.-S. Chua, and J. Li, “Hierarchical spatio-temporal context modeling for action recognition,” in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, 2004–2011, IEEE (2009).

[3] M. Selmi, M. A. El-Yacoubi, and B. Dorizzi, “Two-layer discriminative model for human activity recognition,” IET Computer Vision 10(4), 273–278, IET (2016).

[4] M. Selmi, M. El Yacoubi, and B. Dorizzi, "On the sensitivity of spatio-temporal interest points to person identity," in Image Analysis and Interpretation (SSIAI), 2012 IEEE Southwest Symposium on, 69–72, IEEE (2012).

[5] A. W. Vieira, E. R. Nascimento, G. L. Oliveira, Z. Liu, and M. F. Campos, "Stop: Space-time occupancy patterns for 3d action recognition from depth map sequences," in Iberoamerican Congress on Pattern Recognition, 252–259, Springer (2012).

[6] J. Wang, Z. Liu, J. Chorowski, Z. Chen, and Y. Wu, "Robust 3d action recognition with random occupancy patterns," in Computer vision–ECCV 2012, 872–885, Springer (2012).

[7] X. Yang, C. Zhang, and Y. Tian, "Recognizing actions using depth motion maps-based histograms of oriented gradients," in Proceedings of the 20th ACM international conference on Multimedia, 1057–1060, ACM (2012).

[8] O. Oreifej and Z. Liu, "Hon4d: Histogram of oriented 4d normals for activity recognition from depth sequences," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 716–723 (2013).

[9] C. Chen, R. Jafari, and N. Kehtarnavaz, "Action recognition from depth sequences using depth motion maps-based local binary patterns," in Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on, 1092–1099, IEEE (2015).

[10] M. F. Bulbul, Y. Jiang, and J. Ma, "Human action recognition based on dmms, hogs and contourlet transform," in Multimedia Big Data (BigMM), 2015 IEEE International Conference on, 389–394, IEEE (2015).

[11] J. Zhang,W. Li, P. O. Ogunbona, P.Wang, and C. Tang, "Rgb-d-based action recognition datasets: A survey," Pattern Recognition 60, 86–105, Elsevier (2016).

[12] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, 1, 886–893, IEEE (2005).

[13] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Transactions on pattern analysis and machine intelligence 24(7), 971–987, IEEE (2002).

[14] C. Chen, R. Jafari, and N. Kehtarnavaz, "Fusion of depth, skeleton, and inertial data for human action recognition," in Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on, 2712–2716, IEEE (2016).

[15] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," IEEE transactions on pattern analysis and machine intelligence 31(2), 210–227, IEEE (2009).

[16] M. I. Khedher, M. A. El Yacoubi, and B. Dorizzi, "Multi-shot surf-based person re-identification via sparse representation," in Advanced Video and Signal Based Surveillance (AVSS), 2013 10th IEEE International Conference on, 159–164, IEEE (2013).

[17] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 42(2), 513–529, IEEE (2012).

[18] M. F. Bulbul, Y. Jiang, and J. Ma, "Dmms-based multiple features fusion for human action recognition," International Journal of Multimedia Data Engineering and Management (IJMDEM) 6(4), 23–39, IGI Global (2015).

[19] C. Chen, R. Jafari, and N. Kehtarnavaz, "A real-time human action recognition system using depth and inertial sensor fusion," IEEE Sensors Journal 16(3), 773–781, IEEE (2016).

[20] C. Chen, J. Roozbeh, and K. Nasser, "Utd-mhad: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor," in ICIP, 168–172, IEEE (2015).

[21] H. Altınc¸ay, "On naive bayesian fusion of dependent classifiers," Pattern Recognition Letters 26(15), 2463–2473, Elsevier (2005).

[22] L. I. Kuncheva, "Combining pattern classifiers: methods and algorithms," John Wiley & Sons (2004).

[23] P. Zappi, T. Stiefmeier, E. Farella, D. Roggen, L. Benini, and G. Troster, "Activity recognition from on-body sensors by classifier fusion: sensor scalability and robustness," in Intelligent Sensors, Sensor Networks and Information, 2007. ISSNIP 2007. 3rd International Conference on, 281–286, IEEE (2007).

[24] M. Ye, Q. Zhang, L. Wang, J. Zhu, R. Yang, and J. Gall, "A survey on human motion analysis from depth data," in Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications, 149–187, Springer (2013).

[25] D. Weinland, R. Ronfard, and E. Boyer, "A survey of vision-based methods for action representation, segmentation and recognition," Computer vision and image understanding 115(2), 224–241, Elsevier (2011).

[26] M. Selmi and M. A. El-Yacoubi, "Multimodal sequential modeling and recognition of human activities," in International Conference on Computers Helping People with Special Needs, 541–548, Springer (2016).

[27] J. Han, L. Shao, D. Xu, and J. Shotton, "Enhanced computer vision with microsoft kinect sensor: A review," IEEE transactions on cybernetics 43(5), 1318–1334, IEEE (2013).

[28] J. K. Aggarwal and L. Xia, "Human activity recognition from 3d data: A review," Pattern Recognition Letters 48, 70–80, Elsevier (2014).

[29] C. Chen, R. Jafari, and N. Kehtarnavaz, "A survey of depth and inertial sensor fusion for human action recognition," Multimedia Tools and Applications 76(3), 4405–4425, Springer (2017).

[30] S. Tang, X. Wang, X. Lv, T. X. Han, J. Keller, Z. He, M. Skubic, and S. Lao, "Histogram of oriented normal vectors for object recognition with a depth sensor," in Asian conference on computer vision, 525–538, Springer (2012).

[31] W. Zhu, C. Lan, J. Xing, W. Zeng, Y. Li, L. Shen, X. Xie, and others, "Co-occurrence feature learning for skeleton based action recognition using regularized deep lstm networks.," in AAAI, 2, 8 (2016).

[32] Z. Liu, C. Zhang, and Y. Tian, "3d-based deep convolutional neural network for action recognition with depth sequences," Image and Vision Computing 55, 93–100, Elsevier (2016).

[33] I. Lillo, J. C. Niebles, and A. Soto, "Sparse composition of body poses and atomic actions for human activity recognition in rgb-d videos," Image and Vision Computing 59, 63–75, Elsevier (2017).

[34] J. Imran and P. Kumar, "Human action recognition using rgb-d sensor and deep convolutional neural networks," in Advances in Computing, Communications and Informatics (ICACCI), 2016 International Conference on, 144–148, IEEE (2016).

[35] C. Chen, R. Jafari, and N. Kehtarnavaz, "Improving human action recognition using fusion of depth camera and inertial sensors," IEEE Transactions on Human-Machine Systems 45(1), 51–61, IEEE (2015).

[36] X. Wang, T. X. Han, and S. Yan, "An hog-lbp human detector with partial occlusion handling," in Computer Vision, 2009 IEEE 12th International Conference on, 32–39, IEEE (2009).

[37] I. Dimitrovski, D. Kocev, S. Loskovska, and S. Dˇzeroski, "Hierarchical annotation of medical images," Pattern Recognition 44(10), 2436–2449, Elsevier (2011).

[38] M. Noridayu, R. A. R. Abdul, R. Ava, and R. Dhanesh, "Feature fusion in improving object class recognition," Citeseer (2012).

[39] N. Manshor, A. R. A. Rahiman, A. Rajeswari, and D. Ramach, "Feature fusion in improving object class recognition," Citeseer (2012).

[40] N. E. D. Elmadany, Y. He, and L. Guan, "Human action recognition using hybrid centroid canonical correlation analysis," in Multimedia (ISM), 2015 IEEE International Symposium on, 205–210, IEEE (2015).

[41] E. Escobedo and G. Camara, "A new approach for dynamic gesture recognition using skeleton trajectory representation and histograms of cumulative magnitudes," in Graphics, Patterns and Images (SIBGRAPI), 2016 29th SIBGRAPI Conference on, 209–216, IEEE (2016).

INTENTIONAL BLANK

QUANTITATIVE ANALYSIS IN HEURISTIC EVALUATION EXPERIMENTS OF E-COMMERCE WEBSITES

Xiaosong Li, Ye Liu, Zizhou Fan and Will Li

Computer Science Practice Pathway,
Unitec Institute of Technology, Auckland, New Zealand

ABSTRACT

This paper reports a pilot study on developing an instrument to predict the quality of e-commerce websites. The 8C model was adopted as the reference model of the heuristic evaluation. Each dimension of the 8C was mapped into a set of quantitative website elements, the websites were scraped to get the quantitative website elements, and the score of the dimension was calculated. A software was developed in PHP for the experiments. In the training process, 10 experiments were conducted and quantitative analysis was regressively conducted between the experiments. The conversion rate was used to verify the heuristic evaluation of an e-commerce website. The results showed that the mapping revisions between the experiments improved the performance of the evaluation instrument, therefore the experiment process and the quantitative mapping revision guideline proposed was on the right track. The experiment results and the future work have been discussed.

KEYWORDS

E-commerce Website, Heuristic Evaluation, Regression Experiments, 8C framework, Quantitative Analysis

1. INTRODUCTION

E-commerce websites have increased greatly in the new era; they face many competitors. Research revealed that efforts put into usability design and modification improved the performance of usability on websites greatly [1]. To help website developers and other stakeholders understand how to develop e-commerce websites properly and maximize profit, many evaluation methods have been developed [1, 2]. One approach is called user based testing [1], which takes into account subjective perception, both in terms of website content and design. This perception varies with the expertise, the cognitive skills and the end goal of each user [1]. If an automatic approach is used to evaluate website content and design from the user's perspective, that should standardize the evaluation process and make the evaluation consistent and objective.

7C framework was introduced to evaluate the quality of e-commerce website content and user interface design [3], which is considered as a useful reference model for developers, analysts, managers, and executives, when designing and/or evaluating the interface channels between the customer and the web based application. However, it is insufficient to completely address the new generation of web applications [4]. Collaboration and user-generated content are important features in the new generation websites. The 7C framework was extended into the 8C framework by adding collaboration as the 8th element in the model and the meaning of each of the eight

David C. Wyld et al. (Eds) : CNDC, DPPR, AIAA, WiMNET, ACITY - 2017
pp. 71– 80, 2017. © CS & IT-CSCP 2017 DOI : 10.5121/csit.2017.71607

design elements was updated as well, so that they are effective in representing the interface design elements of new generation websites [4].

Usually when a website is evaluated against the 7C framework, subjective perception is used. For example, in [5], a checklist consisting of 63 checkpoints was developed based on research literature and expert opinions to evaluate a group of 4 and 5-star luxury hotel websites against the 7C framework. This approach again could be inconsistent and subjective. An automatic approach could improve this.

The heuristic evaluation method is a technique for evaluating the usability, with the inspection being carried out mainly by evaluators from principles established by the discipline [6]. In most applications the results tend to be qualitative, however, these qualitative results do not allow us to determine how usable it is or how it becomes an interactive system. Hence, the need for quantitative results may also be very necessary in order to determine the effort that would be needed to get a sufficiently usable system [6].

The accurate prediction of a numerical target variable is an important task in machine learning. Quantitative heuristic analysis has been used in machine learning to predict various values in the data mining and inductive rule learning communities, where a strong focus lies on the comprehensibility of the learned models [7]. In [7], a heuristic rule learning algorithm that learns regression models is used where a region around the target value predicted by the rule is dynamically defined. In [8], a unified measure of web usability was used based on a multiple regression model, and then the estimated index is used to measure its impact on community bank performance. Results showed that banks with higher usability score perform significantly better than those with lower score.

Conversion rate (CR) is the percentage of users who take a desired action. The typical example of conversion rate is the percentage of website visitors who buy something on the site, For the purpose of managing user interface design and tracking the effectiveness of user experience efforts, the conversion rate is usually very important [9]. The conversion rate measures what happens once people are at your website, which is greatly influenced by the design and is a key parameter to track for assessing whether a user experience strategy is working. Lower conversion rates? You must be doing something wrong with the design. Higher conversion rates? You can praise your designers [9]. This suggests that there is a proportional relationship between the conversion rate of an e-commerce website and its user interface design. It is reasonable to use the conversion rate to measure the quality of the user interface of an e-commerce website.

This paper presents a pilot study on developing an instrument to predict the quality of e-commerce websites. The objective of the resulting instrument is to provide a meaningful estimation on the quality of a given e-commerce website. The 8C model was adopted as the reference model of the heuristic evaluation. Each dimension of the 8C model was mapped into a formula consisting of a set of quantitative website elements, the websites were scraped to get the quantitative website elements, and the score of the dimension was calculated based on the formula. Another formula was defined to calculate the total score for the website based on the scores from each dimension.

A software was developed in PHP for both training and testing experiments. An experimental process and its quantitative mapping revision guideline were proposed and used. In the training process, 10 experiments were conducted and quantitative analysis was regressively conducted between the experiments. The conversion rate was used in this study to test and verify the heuristic evaluation of an e-commerce website. 100 websites from five different categories were selected as the training data. 7 websites ordered by the conversion rate were used as testing data

to test the results at the end of each experiment in the training process and 15 websites ordered by the CR were used as the testing data.

In the rest of this paper, the design of the experiment is described first, then the experiments and the results are presented and discussed, after that a summary and future work are given lastly.

2. THE EXPERIMENT ENVIRONMENT AND DESIGN

This study considered the seven dimensions defined in the 7C framework and the additional dimension "collaboration" introduced in the 8C framework. For the web 2.0 features, only those features easy to be obtained via web scraping were considered such as website forum, blog and Ajax. Table 1 presents the key meaning of each dimension in 8C [4].

Table 1. The key meaning of each dimension in 8C.

Dimensions	Meanings
1: Context	How the site is organized, and how the content is presented to the users?
2: Content	What are offered by the site?
3: Community	Non-interactive communication; Interactive communication.
4: Customization	Refers to the site's ability to tailor itself (tailoring) or to be tailored.
5: Communication	Site-to-user communications.
6: Connection	Refers to the extent of formal linkage from one site to others.
7: Commerce	Deals with the interface that supports the various aspects of e-commerce.
8: Collaboration	Generally in the form of feedback forms, forums, and bulletin boards.

Quantitative usability estimation is typically associated with the calculation of metrics that assess dimensions of software quality [6]. Measuring the user experience offers so much more than just simple observation. Metrics add structure to the design and evaluation process, give insight into the findings and provide information to the decision makers [10].

Figure 1. The mapping management UI and the relations in a mapping

A software written in PHP was developed for both training and testing experiments. Figure 1 shows the software mapping management user interface with the mapping relations between the ***Context*** dimension and the selected HTML tags/keywords in ***Experiment 7***. For an e-commerce website to be experimented, only the home page was considered in this study.

Two major approaches were used to identify the website quantitative elements and calculate the metrics for each dimension: finding keywords and scraping HTML tags, where a keyword could be an important text or a JavaScript/CSS keyword. Each keyword or HTML tag is associated with a numeric weight, which determines the importance of the relation, higher weight means more important. The mapping relations between each dimension and the selected keyword or HTML tag are defined before each experiment, which can be adjusted in the subsequent experiments based on the experiment results.

Let *NR* be the total number of the relations in a mapping between a dimension and the selected HTML tags/keywords; *RSi* be the score of relation *i*; *Wi* be the associated weight of relation *i*; if the relation *i* is a keyword, *RSi* will be *Wi*; if relation *i* is an HTML tag, *RSi* will be calculated by the following formula:

$$RSi = \frac{STagNi}{TTagNi} * Scalar * Wi \qquad (1)$$

Where *STagNi* is the number of the occurrence of the selected HTML tag for relation *i*; *TTagNi* is the total number of HTML tag on the selected page; Scalar is set as 100 to make the score a meaningful magnitude. The total score *TS* is the sum of the scores for all 8 dimensions in 8C framework.

$$TS = \sum_{i=1}^{8} (RSi) \qquad (2)$$

An experimental process and its revision guideline were proposed and used. Initially, in ***Experiment 1***, only the keywords/HTML tags that can intuitively reflect the meaning of a dimension as defined in the 8C framework were selected as the relations for the mapping of that dimension heuristically. The weights for the relations also were selected in the similar way heuristically.

Then the scores for all the training websites were calculated respectively according to formula (2). The training websites were ordered based on their CRs (CR) first, and then the training websites were ordered again based on their scores. If the score order is different from the CR order, the mappings for all the 8 dimensions were reviewed and revised in the following three aspects:

1. Check if any relation score is dominating the dimension score based on the overall performance of the training websites, if yes, adjust the weight of that relation to make the relation score of a meaningful magnitude.
2. Check if the score of any dimension is dominating the total score based on the overall performance of the training websites, if yes, scale all the scores in that dimension to make the dimension score of a meaningful magnitude.
3. Recheck all the mappings against the 8C model and make adjustment accordingly. This may involve adding or deleting relations.

The above would result in the new mappings for the next experiment. This process went through regressively for 10 experiments. As an example, Table 2 shows the mappings for ***Collaboration*** dimension in ***Experiment 1***, ***Experiment 6*** and ***Experiment 8***.

Table 2. The mapping for Collaboration in three experiments.

Experiment 1		**Experiment 6**		**Experiment 8**	
Relation Name	Relation Weight	Relation Name	Relation Weight	Relation Name	Relation Weight
Forums	3	Forums	3	Forums	3
Bulletin boards	3	Bulletin boards	3	Bulletin boards	3
FAQ	3	FAQ	3	FAQ	3
		Feedback	5	Feedback	5
				Review	5
				Suggestion	5
				Comment	5

3. THE EXPERIMENT ENVIRONMENT AND DESIGN

100 websites from five different categories (Electronics, Publishing & entertainment, Home and garden, Books, Industrial equipment), 20 from each category were selected as the training data. The five categories were selected from [11], where the CRs for 25 retail categories were listed. ***Electronics*** and ***Publishing & entertainment*** were associated with high level CR; ***Home & garden*** and ***Books*** were associated with middle level CR; and ***Industrial equipment*** were associated with low level CR

Table 3. The categories of training data.

Categories	**Conversion Rates**
Electronics	Around 23%
Publishing & entertainment	Around 20%
Home & garden	Around 14%
Books	Around 13%
Industrial equipment	Around 7%

The top 10 e-commerce websites based on CR for 2010 were listed in [12], only 7 of them were valid for the experiments, and they all were used to test the results at the end of each experiment for all the 10 experiments. Table 4 shows the 7 testing websites.

Table 4. The testing data.

Website Names	**Conversion Rates**
Woman Within	25.3%
Blair	20.4%
1800petmeds	17.7%
qvc	16%
ProFlowers	15.8%
Oriental Trading Company	14.9%
Roamans	14.4%

After each experiment, the training websites were ordered again based on their scores. If the score order is different from the CR order, the mappings for all the 8 dimensions in 8C model were reviewed and then revised if needed, this resulted in the new mapping for the next experiment. Figure 2 shows the absolute score differences between the expected order and the actual order. [9]

suggests that there is a proportional relationship between the CR of an e-commerce website and its user interface design. It is reasonable to assume that the less the difference, the more accurate the evaluation. The differences in each experiment for all the 7 training websites were averaged and Figure 3 shows the average for all the experiments except ***Experiment 9***. As the scores of ***Experiment 10*** were obtained by scaling the scores in ***Experiment 9*** by 10%. It was observed that the trends of the curve going down along the experiments. This suggested that the mapping revisions between the experiments improved the performance of the evaluation instrument and it is positive.

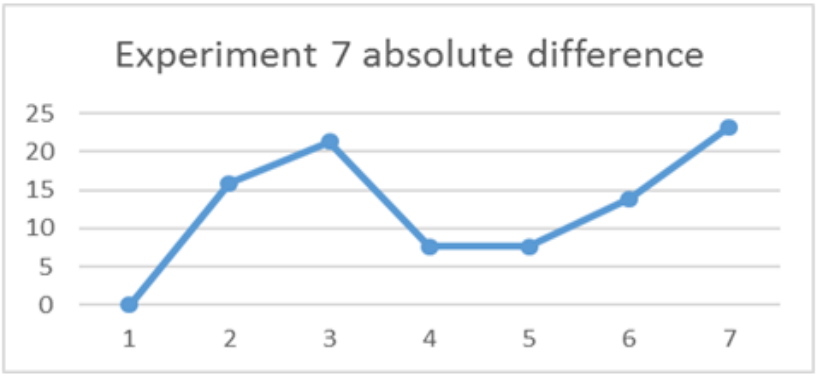


Figure 2. The absolute score differences between the expected order and the actual order for one experiment

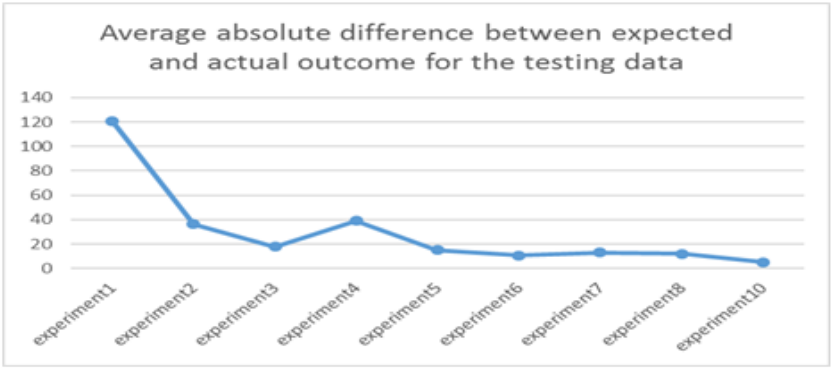


Figure 3. The average absolute difference between expected and actual outcomes for nine experiments

Table 5. The dimension contribution analysis of Experiment 8 & 10.

	Experiment 8			**Experiment 10**		
Attribute Category	Attribute Number	Contribution to the total score	Standard deviation	Attribute Number	Contribution to the total score	Standard deviation
Context	28	40.94%	15.43	28	14.71%	1.55
Content	11	5.93%	7.08	11	8.21%	2.62
Community	18	9.04%	7.77	18	12.60%	3.13
Customization	5	7.85%	6.00	5	11.22%	2.43
Communication	13	12.47%	6.55	13	17.79%	2.65
Connection	2	5.15%	4.81	2	7.44%	1.97
Commerce	10	14.68%	8.30	10	15.80%	2.55
Collaboration	7	3.93%	3.94	7	12.22%	3.48

In ***Experiment 8***, it was observed that some of the dimensions' scores dominated the total score of the website. Table 5 shows the dimension contribution analysis of ***Experiment 8 & 10***, where the number of attributes number is the number of relations in the mapping for each dimension (Attribute Category) of the 8C; contribution to the total score is the sum of the scores in a dimension for all the training websites divided by the total score of all the training websites in an experiment. ***Context*** made much more contribution (40.94%) than the others did. On the other hand, some were too small to influence the total score, such as Content (5.93%) ***Connection*** (5.93%) and ***Collaboration*** (5.93%). The standard deviation can provide some ideas on whether the attributes in a dimension is informative. For example, standard deviation for ***Collaboration*** was the smallest one in ***Experiment 8***, however, there were 7 attributes in this dimension. This suggested that the meaning of the attributes might be overlapping. So standard deviation for each dimension over all the training websites should be considered in the review process after each experiment in the future study.

In this study, scaling the scores for the dimensions were attempted to balance the influences of all the dimensions. For a website, let TS be its total score, and let score codes and scale parameter codes be defined in Table 6.

Table 6. Codes used in the scale formula.

Score Code	Meaning of the code	Scale Parameter	Scale Number
SC1	Score of Context	P1	1
SC2	Score of Content	P2	4
SC3	Score of Community	P3	4
SC4	Score of Customization	P4	4
SC5	Score of Communication	P5	4
SC6	Score of Connection	P6	4
SC7	Score of Commerce	P7	3
SC8	Score of Collaboration	P8	9

Formula (3) was used to calculate *TS* in ***Experiment 9***, the resulting scores were much larger than the other experiments, so the results were divided by 10 for further scaling, which were recorded as ***Experiment 10***.

$$TS = \sum_{i=1}^{8} (SCi * Pi) \qquad (3)$$

The right column of Table 5 shows the contribution of each dimension after the scaling in ***Experiment 10***. This time the contributions of the dimensions are much balanced.

The verifying data was obtained from [13], which listed top 15 e-commerce websites based on CR for 2014. All of them were valid for the experiments and were used to check the mappings used in all the experiments except ***Experiment 9*** as ***Experiment 10*** can represent ***Experiment 9***. Table 7 shows the order of the 15 verifying websites. Figure 4 shows the average absolute difference between expected and actual outcomes for the 15 verifying websites. It was observed that the trend of the curve was going down along the experiments, which was consistent with the testing results of Figure 3. This suggested that the experiments were on the right track and the results were positive. The resulting instrument from ***Experiment 10*** could be used to evaluate a given e-commence website and provide meaningful estimation on the quality of the website.

Table 7. The 2014 data.

Website Names	Conversion Rates
Play.Google	30.00%
MovieMars	22.95%
DollarShaveClub	20.00%
1800Contacts	18.40%
1800Flowers	16.90%
Coastal	14.50%
Keurig	13.00%
FTD	11.70%
ProFlowers	11.70%
PureFormulas	10.74%
FreshDirect	10.50%
TheGreatCourses	10.04%
1800PetMeds	10.00%
AmeriMark	10.00%
OvernightPrints	9.95%

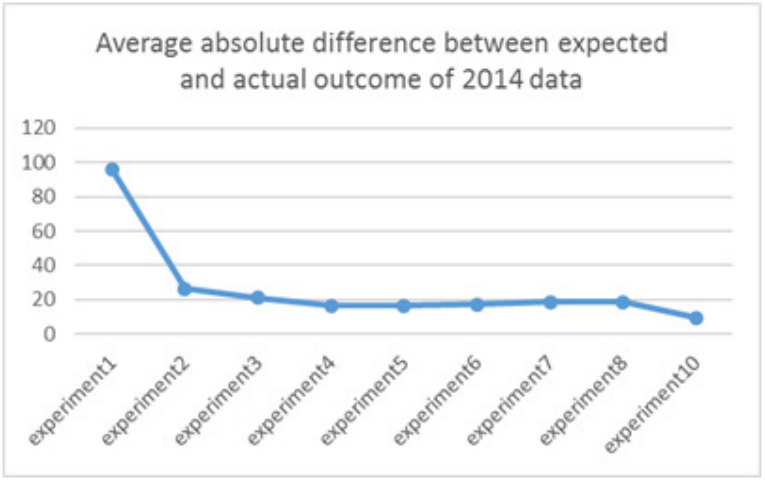


Figure 4. The average absolute difference between expected and actual outcomes for 2014 data.

Table 8. The average score of each category in each experiment.

No	Electronics	Entertainment	Home	Books	Industrial
1	165.27	142.92	157.36	152.41	148.19
2	144.16	124.11	124.71	129.84	124.40
3	108.77	83.19	85.73	93.20	88.15
4	98.73	85.15	90.07	91.09	94.00
5	90.08	77.28	82.28	87.61	83.50
6	91.27	81.50	84.73	86.80	89.00
7	163.41	141.10	145.65	136.69	148.75
8	159.83	138.54	142.54	134.37	145.52
9	467.53	395.98	380.85	358.28	403.04
10	46.77	39.89	40.21	36.00	40.34

Table 8 shows the average score of each category in each experiment. According to Table 3, websites in ***Electronics*** category should have the highest scores; websites in ***Industrial***

Equipment category should have the lowest scores; and ***Books*** are in the middle. In Table 8, the ***Electronics*** websites always have the highest score in all the experiments, ***Books*** websites are in the middle sometimes, particularly in ***Experiment 10***. These are consistent between the two tables (Table 3 and Table 8). However, ***Industrial Equipment*** websites usually do not have the lowest scores. This suggests that the website design and usability could have an impact on an e-commerce website's CR, however, there are other factors as well, such as the product nature, those relevant factors should be taken into consideration as well in an e-commerce website evaluation. In addition, the experiment results are dynamic; they are impacted by the network environment. The quantitative mappings might not be available temporarily for those popular websites due to heavy network traffic sometimes, and those popular websites are likely the websites with high scores. ***Industrial Equipment*** websites are not as popular as book websites or entertainment websites, so they are less impacted by network traffic; on the other hand, book websites or entertainment websites might get lower scores than their real scores due to network traffic, this issue should be addressed in the future experiment.

4. SUMMARY AND FUTURE WORK

This paper presented a pilot study on developing an instrument to predict the quality of e-commerce websites. The objective is to provide a meaningful estimation of a given e-commerce website. The 8C model was adopted as the reference model of the heuristic evaluation. Each dimension of the 8C was mapped into quantitative elements by means of web scraping. A software was developed in PHP for both training and testing experiments. 10 experiments were conducted and quantitative analysis was regressively conducted between the experiments. The conversion rate was used to test and verify the heuristic evaluation. It was observed that the trends of the curve for the differences between the expected and actual outcomes was going down along the experiments for both of the testing data and verifying data. This suggested that the mapping revisions between the experiments improved the performance of the evaluation instrument, therefore the experiment process and the revision guideline proposed in Section 2 was on the right track.

However, there are limitations in this study. The experiments only had been done on the home page of each website, although home page is very important for a website and it can provide rich information about the website, it is not sufficient for an e-commerce website, in some cases, the shopping cart or product list are not on the home page. Due to technique incapacity, not all the website features can be mapped into quantitative elements. The experiment results could be impacted by the network environment although that impact is not significant.

The above should be considered in the future work. In addition to that, the mapping revision process could be more robotic by improving the revision guideline (algorithm), for example, the standard deviation for each dimension over all the training websites could be considered in the review process after each experiment in the future work. The evaluation framework should not be limited to the 8C model; it could be extended to include other factors. [13] proposed a number of ways to improve the CR of an e-commerce website, which should be considered in the future study.

REFERENCES

[1] Li, Fangyu, & Yefei Li, (2011) "Usability evaluation of e-commerce on B2C websites in China.", Procedia Engineering 15 pp5299-5304.

[2] Bezes, Christophe (2009) "E-commerce website evaluation: a critical review."

[3] Rayport, Jeffrey F., & Bernard J. Jaworski (2002) Introduction to e-commerce. McGraw-Hill/Irwin marketspaceU.

[4] Yang, T. Andrew, Dan J. Kim, Vishal Dhalwani, & Tri K. Vu, (2008) "The 8C framework as a reference model for collaborative value Webs in the context of Web 2.0." In Hawaii International Conference on System Sciences, Proceedings of the 41st Annual, pp319-319. IEEE.

[5] Hamidizadeh, Mohammad R., Mohammad E. Fadaeinejad, & Fayegh Mojarrad, (2011) "Design of internet marketing based on 7Cs model." In 2011 International Conference on Social Science and Humanity.

[6] González, Marta, Llúcia Masip, Antoni Granollers, & Marta Oliva, (2009) "Quantitative analysis in a heuristic evaluation experiment." Advances in Engineering Software 40, no. 12. pp1271-1278.

[7] Janssen, Frederik & Johannes Fürnkranz, (2011) "Heuristic rule-based regression via dynamic reduction to classification." In IJCAI Proceedings-International Joint Conference on Artificial Intelligence, vol. 22, no. 1, pp1330.

[8] Acharya, Ram N., Albert Kagan, Srinivasa Rao Lingam & Kevin Gray, (2011) "Impact Of Website Usability On Performance: A Heuristic Evaluation Of Community Bank Homepage Implementation." Journal of Business & Economics Research (JBER) 6, no. 6.

[9] Nielsen, Jakob, (2013) Conversion Rates. https://www.nngroup.com/articles/conversion-rates/, last accessed 2017/06/28.

[10] Tullis, Thomas & Albert, William. Measuring the user experience. Morgan Kaufmann; 2008.

[11] Burstein, Daniel. (2015) Ecommerce Research Chart: Industry benchmark conversion rates for 25 retail categories. http://www.marketingsherpa.com/article/chart/conversion-rates-retail-categories/, last accessed 2016/07/29.

[12] Chopra, Paras. (2010) Top 10 eCommerce Websites (by Conversion Rate). https://vwo.com/blog/top-ecommerce-websites-conversion-rate/, last accessed 2016/07/29.

[13] Saleh, Khalid. (2017) The Average Website Conversion Rate by Industry. https://www.invespcro.com/blog/the-average-website-conversion-rate-by-industry/, last accessed 2017/05/29.

AUTHORS

Dr. Xiaosong Li obtained her PhD (1999) in Computer Science from University of Auckland in New Zealand. Her research interests include Graphical User Interface, E-Commerce Websites, Machine Learning and etc. She joined Unitec in 2002 where she is a Senior Academic Staff Member.

DEVELOPMENT OF A RADIAL BASIS FUNCTION NETWORK TO ESTIMATE THE HEAD GENERATED BY ELECTRICAL SUBMERSIBLE PUMPS ON GASEOUS PETROLEUM FLUIDS

Morteza Mohammadzaher[1], Mojataba Ghodsi[1] and Abdullah AlQallaf [2]

[1]Department of Mechanical and Industrial Engineering,
Sultan Qaboos University, Muscat, Oman
[2] Department of Electrical Engineering, Kuwait University
Kuwait City, Kuwait

ABSTRACT

This paper proposes radial basis function network (RBFN) models to estimate the head of gaseous petroleum fluids (GPFs) in electrical submersible pumps (ESPs) as an alternative to widely used empirical models. Both exact and efficient RBFN modelling approaches were employed. RBFN modelling essentially tend to minimise the modelling error, the discrepancy of estimated and real outputs within the modelling data. This may lead to overfitting and lack of model generality for operating conditions not reflected in modelling data. This critical matter was addressed in RBFN design process, and highly accurate RBFNs were developed and cross validated.

KEYWORDS

Electrical Submersible Pump(ESP), Radial Basis Function Network (RBFN), Model, Petroleum, Gaseous, Head Estimation

1. INTRODUCTION

ESPs are widely used to lift large volume of fluid from downhole at different well conditions [1, 2]. As a vital task, size of these pumps should be chosen correctly as over- or under-sizing leads to premature equipment failure or low petroleum fluid recovery. In order to facilitate size selection, manufacturers normally provide curves depicting generated head versus liquid volumetric flow rate for each ESP size, the size of ESPs is selected based on manufacturer curves.

However, the aforementioned curves are not valid for gaseous fluids; where, ESPs are utilised to pump two-phase fluid with high gas content [1].The solution is to develop models to estimate the generated head by ESPs on GPFs. This paper focuses on head estimating models and exclude other types of models developed for GFPs in ESPs, e.g. the ones which estimate surging or stability border [3], gas bubble size [4] or in-situ gas volume fraction [5].

David C. Wyld et al. (Eds) : CNDC, DPPR, AIAA, WiMNET, ACITY - 2017
pp. 81– 90, 2017. © CS & IT-CSCP 2017 DOI : 10.5121/csit.2017.71608

Head-estimating models have been developed and investigated for decades using analytical, numerical and empirical approaches [6]. Analytical models have been derived based on mass and momentum balances [7, 8]. Use of unrealistic assumptions and oversimplification of complex physics of two-phase fluids have weaken the reliability of analytical models. Numerical models are not trusted too as they are normally formulated based on one-dimensional two-fluid conservations of mass and momentum along streamlines and require the prediction of surging initiation in ESPs [9]. On the contrary, empirical models of GFPs in ESPs are widely trusted and used in practice [10, 11].

2. EMPIRICAL MODELS

In this section, homogenous model and a number of empirical head-estimating models of GPFs in ESPs are briefly introduced. The parameters of the presented empirical models have been identified using the data collected from experiments on diesel fuel/carbon dioxide mixtures, reported in [12]. Aforementioned mixtures are similar to petroleum fluids [13]. A number of other empirical models have been also reported in the literature in which their parameters have been identified based on the data of experiments on air/water mixtures [14-16]. These models have been excluded from this paper due to dissimilarity of the tested fluids and GPFs.

2.1. Model 1

The hoariest model is the homogenous model. This model is in fact a brief analytical model rather than an empirical model, based on oversimplification of two-phase physics of GPFs. Homogenous model receives an input from the curve provided by the manufacturer: the generated head by ESP if pure liquid was pumped instead of GPF (Hl) [17]. This head is modified with assumption that the fluid motion is homogenous i.e. liquid and gas have equal speeds:

$$\hat{H}_m = \left((1-\alpha)\rho_l + \alpha\rho_g\right) H_l, \tag{1}$$

where ρ, H and indices l, g *and* m stand for density, head, liquid, gas, and mixture, respectively. α is gas void fraction. ^ shows the head is estimated (not experimentally measured).

2.2. Model 2

The second model was developed by Turpin et al in 1986 [18]:

$$\hat{H}_m = H_l \exp\left(346430\left(\frac{q_g}{p_{in} q_l}\right)^2 - 410\left(\frac{q_g}{p_{in} q_l}\right)\right), \tag{2}$$

where q_l and q_g are liquid and gas volumetric flow rates in gallons per minutes (*gpm*) , p_{in} is intake pressure in *psi*.

2.3. Model 3

This model was proposed by Sachdeva et al, in 1992 [19]:

$$\hat{H}_m = \frac{K_1}{\rho_m g} {p_{in}}^{E_1} \alpha^{E2} {q_l}^{E3}. \tag{3}$$

where g stands for gravity acceleration. The values of E_1, E_2 and E_3 and K_1 are listed in [10] for multiple stages of electrical submersible pumps. As an example, for 8 stages of I-42B radial ESP, K_1=1.1545620, E_1=0.943308, E_2=-1.175596 and E_3=-1.300093. Similar to Model 2, (3) is

convertible to a linear equation through taking algorithm. Parameters of a linear equation can be often identified straightforwardly using least square of error algorithm [20]

2.4. Model 4

This model was presented by Zhou and Sachdeva in 2010 [10]:

$$\hat{H}_m = H_{\max} K_2 (Cp_{in})^{\alpha E_4} (1-\alpha)^{E_5} \left(1 - \frac{q_m}{q_{\max}}\right)^{E_6}, \tag{4}$$

where K_2 is a unit-less coefficient, C is pressure unit factor, e.g. 1, 1000 or 0.145 for *psi, ksi* or *kPa*. H_{max} and q_{max} are nominal maximum head and flow rate of the ESP; q_m is mixture or GPF flow rate where $q_m = q_l + q_g = q_g / \alpha$.

Mathematical structure of this model seems more meaningful than Model 3; as if gas void fraction and flow rate equal zero, estimated head is definitely H_{max}. According to [10], for 8 stages of I-42B radial ESP, K_2=1.971988, E_4=1.987838, E_5=9.659664 and E_6=0.905908.

2.5. Summary and Limits of Empirical Models

All presented models have three input variables amongst p_{in} , ρ_l, ρ_g, q_l, $q_{g,}$, q_m or α. Two other potential input variables, pump rotational speed and temperature are missing in all empirical models of GPFs in ESPs. In fact, the parameters of the presented models have been identified based on the data collected at a fixed rotational speed of 3500 *rpm*; thus, the models are valid merely at this speed. The estimated head can be adapted for other rotational speeds using 'affinity laws' detailed in [2, 10].

3. MODEL DEVELOPMENT

In this research, a radial basis function network (RBFN) was developed to estimate the head of mixtures of carbon dioxide/ diesel fuel pumped by eight stages of an I-42B radial ESP. The reason to choose an RBFN as the head-estimating model is the fact that RBFNs are universal approximators with significant mathematically proven modelling capabilities [21]. Inspired by existing empirical models, a single output of H_m and triple inputs of p_{in}, q_m and α were opted for the RBFN model. The same experimental data, which were used to identify the parameters of empirical models 1-4, were utilised to develop and test the RBFN. Thus, the proposed model and presented empirical models are comparable. The experimental data cover a wide range of gas void fractions [0 0.5], intake pressures [50 to 400] *psi* and heads [1 55] *ft*. 110 sets of input/output data are available, in this research, 93 were used for modelling and 17 sets for testing the RBFN.

An RBFN has two layers, the first layer receives inputs array (**U**) and produces the 'layer output' (**O**).The second layer receives **O** and produces the 'network output' (**Y**). In this problem, with three inputs and one output, the input and output arrays are $\mathbf{U}_{3\times Q}$ and $\mathbf{Y}_{1\times Q}$. Q is the number of data sets which are fed to the RBFN at once. For instance, if the inputs of the modelling data are fed into the model altogether, then Q=93.

The first layer has an array of weights ($\mathbf{W}_{R\times 3}$) and a scalar namely Spread (S). The components of layer output, $\mathbf{O}_{R\times Q}$, are calculated as following:

$$\mathbf{O}_{ik} = \exp\left(-\left(\underbrace{S\sum_{j=1}^{3}\left(\mathbf{W}_{ij}-\mathbf{U}_{jk}\right)^2}_{\substack{\text{distance between input}\\ \text{and weight arrays}}}\right)^2\right). \quad (5)$$

The second layer has an array of weights ($\mathbf{X}_{1\times R}$) as well as an array of biases ($\mathbf{B}_{1\times Q}$). The output array is calculated as following:

$$\mathbf{Y}_{1\times Q} = \mathbf{X}_{1\times R}\times\mathbf{O}_{R\times Q}+\mathbf{B}_{1\times Q}, \quad (6)$$

Combination of (5) and (6) is the structure of the RBFN; the next task is to identify its unknown parameters R, $\mathbf{W}$, S, $\mathbf{X}$ and $\mathbf{B}$ using the modelling data including input and output vectors of $\mathbf{U}_{3\times 93}$ and $\mathbf{Y}_{1\times 93}$.

From (5), it is clear that the range of $\mathbf{O}$ components is [0 1]; also, if i^{th} row of $\mathbf{W}$ and k^{th} column of $\mathbf{U}$ are identical, $\mathbf{O}_{ik}$ will be at its maximum, 1;or simply, maximum values of $\mathbf{O}$ components happen if the rows of $\mathbf{W}$ are same as the columns of $\mathbf{U}$. From (6), it can be seen larger components of $\mathbf{O}$ are more influential on the network output. As a result, in order to maximise the effect of the modelling data on parameter selection, it has been suggested to set $\mathbf{W}=\mathbf{U}^{\mathbf{T}}$, consequently $R=Q$. Then, $\mathbf{O}$ can be calculated with $\mathbf{U}_{3\times 93}$ of the modelling data and an S according to (5).

$\mathbf{B}$ and $\mathbf{X}$ can also be found from linear equation of (6). In practice, Eq. (6) in the form of (7) was solved to find $\mathbf{B}$ and $\mathbf{X}$:

$$\mathbf{Y}_{1\times 93} = \left[\mathbf{X}\ \ \mathbf{B}\right]_{1\times 186}\begin{bmatrix}\mathbf{O}\\ \mathbf{I}\end{bmatrix}_{186\times 93}, \quad (7)$$

where I is a unique matrix with size of 93×93.

By this point, it has been shown how to find all unknowns of (5) and (6) except for S. The developed model is called an 'exact' RBFN. Such a model evidently provides exact estimation for the modelling data; however, a serious concern about exact RBFNs is inaccuracy of estimation outside the operating points where the modelling data have been collected from. A large spread (S) (S >>1 in (5)) can smoothen the model output and generalise the network [22].

Here is a pseudo-algorithm of exact RBFN modelling (to find R, $\mathbf{W}$, $\mathbf{X}$, $\mathbf{B}$ and S using the input and out arrays of the modelling data, $\mathbf{U}_{3\times 93}$ and $\mathbf{Y}_{1\times 93}$)

1. Set $\mathbf{W}_{93\times 3} = \mathbf{U}^{\mathbf{T}}{}_{93\times 3}$
2. Choose a large S to generalise the developed RBFN
3. Calculate $\mathbf{O}_{93\times 93}$ from (5) with $\mathbf{U}_{3\times 93}$ (from the modelling data), $\mathbf{W}_{93\times 3}$ and S defined at steps 1 and 2.
4. Form (7) with $\mathbf{Y}_{1\times 93}$ (from the modelling data) and $\mathbf{O}$ calculated at step 3.
5. Solve (7) to find $\mathbf{X}_{1\times 93}$ and $\mathbf{B}_{1\times 93}$

A straightforward non-iterative parameter identification algorithm is an advantage of exact RBFNs; however, this method creates models with too many parameters: 466 in this research. While, only 93 modelling data sets, in total 372 pieces of input/output data, are available. Excessive number of parameters and focus of the algorithm on exact fitting of the model to the modelling data increases the risk of 'overfitting' or lack of generality (see [23, 24]). Spread as the only tool to generalise the algorithm has shown to be insufficient for this purpose [23].

An alternative is to use efficient RBFNs which may have much fewer parameters than exact RBFNs. In exact RBFNs, all columns of input data, **U**, are transposed and used as the rows of weight array, **W**. In efficient RBFNs, some columns of **U** are selected and transposed to form **W.** Thus, **W** array is smaller. In order to select **U** columns to be used as **W** rows, first, every single column is transposed and tried as a single-row **W**. The column of **U** which leads to the smallest modelling error (see the second appendix of [25] about the modelling error) is selected, transposed and used as the first row of **W**. At the next iteration, another column of **U** in which the merger of its transpose to **W** leads to the largest drop in the modelling error is chosen and added to **W.** This continues till the number of **W** rows (R) reaches its pre-defined maximum (R_{max}) or the modelling error reaches its predefined target (E_t). It should be noted that if a too small modelling error (e.g. 0) is targeted, overfitting is more likely to happen.

Here is a pseudo-algorithm of efficient RBFN modelling:

1. **W**=null, $\mathbf{U}_{rem}$=**U**, $\mathbf{U}_{opt}$=null, E=1000 (a large number)
2. Choose a large S to generalise the developed RBFN
3. Choose R_{max} and target modelling error, E_t
4. Set R=1
5. Set k=1
6. Add transpose of k^{th} column of $\mathbf{U}_{rem}$ as the R^{th} row of **W**
7. Calculate **O** from (5) with $\mathbf{U}_{3\times 93}$ (from the modelling data), $\mathbf{W}_{R\times 3}$ and S defined at steps 6 and 2.
8. Solve $\mathbf{Y}_{1\times 93} = [\mathbf{X}\ \mathbf{B}]_{1\times(R+93)} \begin{bmatrix} \mathbf{O} \\ \mathbf{I} \end{bmatrix}_{(R+93)\times 93}$ to find $\mathbf{X}_{1\times R}$ and $\mathbf{B}_{1\times R}$ (**Y** and **O** are available from the modelling data and step 7)
9. Find the Modelling Error (ME) from comparison of $\mathbf{Y}_{model}$ (calculated from (5) and (6)) and **Y**
10. if $ME<E$, then $E=ME$ and $\mathbf{U}_{opt}=\mathbf{U}_k$
11. $k=k+1$
12. if $k \leq 93-R$ then go to 6
13. Remove $\mathbf{U}_{opt}$ from $\mathbf{U}_{rem}$
14. $R=R+1$
15. if $R \leq R_{max}$ and $E>E_t$ then go to 5

4. RESULTS AND DISCUSSION

Both exact and efficient RBFN modelling methods were employed to develop models for GPFs pumped by eight stages of an I-42B radial ESP, using 93 sets of experimental data as detailed in the previous section. Accuracy of the models were tested with 17 data sets not used for modelling, 'test data'. Test data include an input array of ${}^{T}\mathbf{U}_{3\times 17}$ and an output vector of ${}^{T}\mathbf{Y}_{17\times 1}$. Upper left index of T refers to 'test'. The estimated outputs of a model with inputs of ${}^{T}\mathbf{U}_{3\times 17}$ is named ${}^{T}\hat{\mathbf{Y}}_{17\times 1}$. The 'test error' or TE, as defined in (8), was used to assess the accuracy of models.

$$TE = \frac{\sum_{l=1}^{17} \left| {}^{\mathrm{T}}\mathbf{Y}_l - {}^{\mathrm{T}}\hat{\mathbf{Y}}_l \right|}{17}, \tag{8}$$

A model should lead to an acceptably low *TE* to be cross-validated [25, 26]. Exact RBFN, with 466 parameters, provides a fairly good *TE*, 2.7683 *ft*, at a very high value of spread, *S*=125. At lower values of *S*, the modelling error may be misleadingly small. For instance, at *S*=1, if the modelling data are used in (8), the resultant error is 0.0645 *ft* where the generalisation [26] or test error is 44.8621 *ft*, an obvious case of overfitting to the modelling data.

The efficient RBFN, however, provides better results with fewer parameters and a lower spread. A spread of 20 and target modelling error (E_t) of 1.2 *ft* result in an efficient RBFN with *R*=74, total number of parameters of 390 and an excellent test error of 1.8648 *ft* , i.e. 3.45% of head range. As a result, this model is cross-validated too.

Table 1 compares the test error (*TE*) and number of parameters in empirical models 1-4 (M1-M4) and developed exact and efficient RBFNs.

Table 1. Test error (*TE*) in *ft* and total number of parameters for different models

	M1	M2	M3	M4	Exact RBFN	Efficient RBFN
Test Error	8.85	7.36	12.13	5.16	2.77	1.87
Number of Parameters	1	3	4	5	466	390

Figures 1-3 compare estimated heads of different models with the real head at three different operating areas. In this paper, an operating area is the collection of operations at a fixed intake pressure and a fixed gas void ratio, e.g. P_{in}=100 *psi* and α=0.2. Table 2 shows the mean of absolute estimation error for different operating areas. The results presented in Table 2 and Figs.1-3 have been calculated for the entire available experimental data in each operating area, not only the tests data.

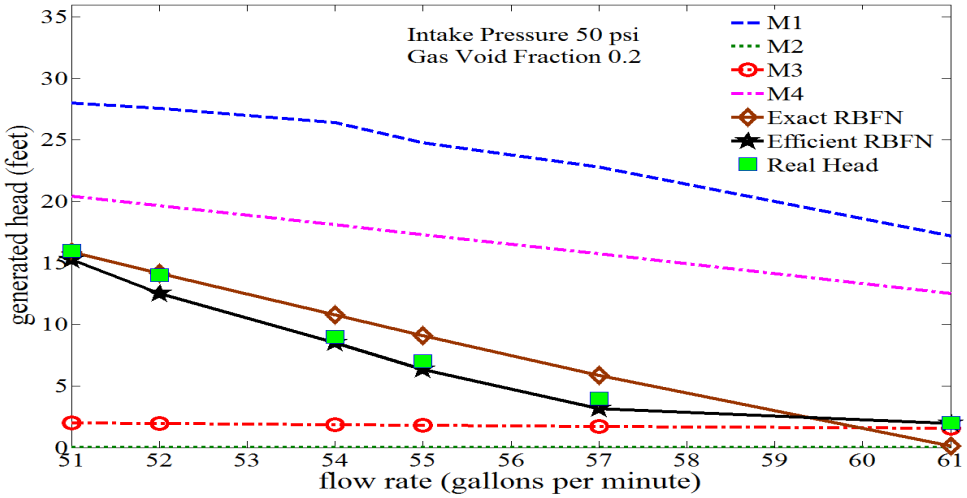


Figure 1. Real and estimated head (by six models) for a mixture of carbon dioxide and diesel fuel pumped by eight stages of an I-42B radial ESP; intake pressure is 50 *psi* and gas void fraction is 0.2

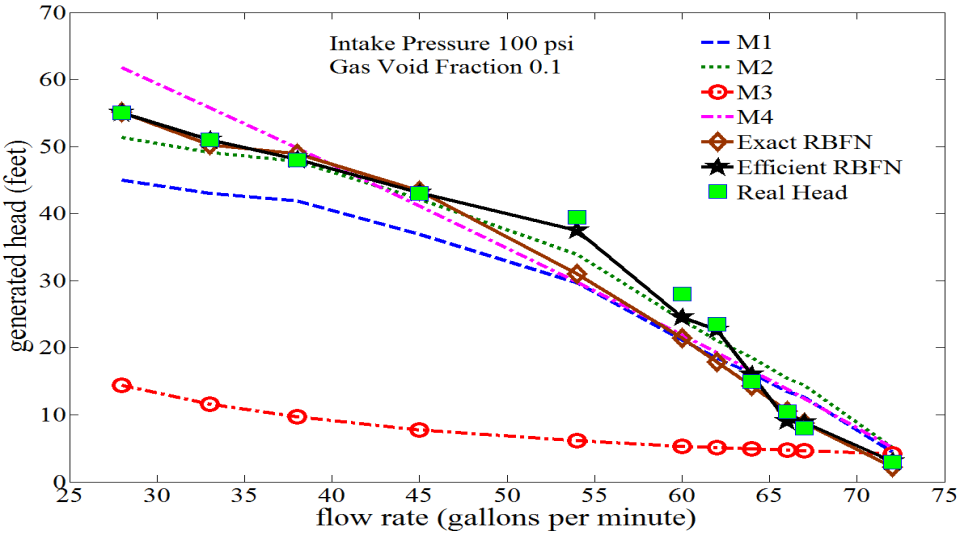


Figure 2. Real and estimated head (by six models) for a mixture of carbon dioxide and diesel fuel pumped by eight stages of an I-42B radial ESP; intake pressure is 100 psi and gas void fraction is 0.1

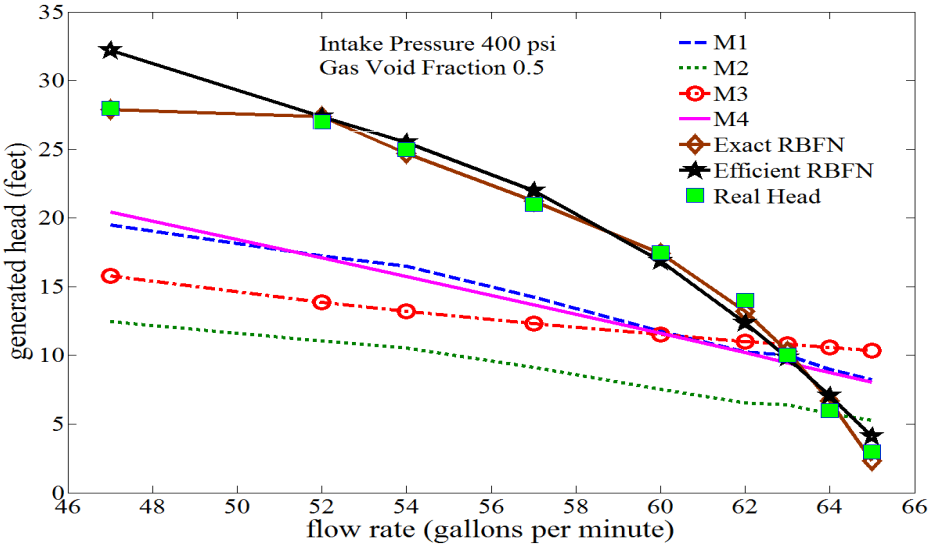


Figure 3. Real and estimated head (by six models) for a mixture of carbon dioxide and diesel fuel pumped by eight stages of an I-42B radial ESP; intake pressure is 400 psi and gas void fraction is 0.5

As to Table 2, in 13 operating areas covered by experiments, the efficient and exact RBFN models outperform all comparable empirical models, M1-M4, in 12 and 11 operating areas respectively. Only in one operating area, Model 3 presents a lower error than RBFNs, at pressure of 50 psi and gas void ratio of 0.4. In this operating area only 2 sets of experimental data are available. In both Table 2 and Figs. 1-3, it is observed that different empirical models may perform better in specific operating areas; thus, so called critical models [6] have been proposed to define the validity area of some empirical models. As an advantage, RBFN models are valid for the whole operating areas where the modelling and test data have been collected from.

Table 2. Mean of absolute head estimation error in *ft* for different models at various operating areas.

P_{in}	α	M1	M2	M3	M4	Exact RBFN	Efficient RBFN
50	0.10	4.66	14.7	24.1	5.00	1.42	0.33
50	0.15	12.1	12.0	11.5	7.32	4.62	1.85
50	0.20	15.8	8.63	6.85	8.63	1.32	0.71
50	0.30	16.4	6.63	5.06	4.34	4.97	1.68
50	0.40	17.1	3.00	1.39	2.44	2.31	1.81
100	0.10	5.66	3.21	22.5	4.24	2.29	0.90
100	0.15	6.28	4.61	13.3	5.94	1.98	1.06
100	0.20	8.25	4.96	10.4	6.68	1.29	1.30
100	0.30	10.1	10.3	7.76	4.65	3.00	1.00
100	0.40	11.7	6.40	3.81	2.89	1.33	0.43
400	0.30	5.47	3.73	9.92	5.84	0.80	1.09
400	0.40	4.45	2.95	8.27	4.30	0.57	0.52
400	0.50	5.69	9.04	7.51	5.79	0.39	1.18

5. CONCLUSION

This paper first presented existing empirical models, which estimate the head of gaseous petroleum fluids in ESPs. These include a simple analytical model (the homogenous model) and three empirical models.

Afterwards, the same data used to identify the parameters of aforementioned empirical models were used to develop and test the exact and efficient RBFN models to serve the same function as the empirical models. The developed models outperformed existing models, and the efficient RBFN particularly estimated head highly accurately with a test error equivalent to 3.45% of head range.

It was also shown that if some popular values were opted for RBFN design factors, e.g. spread of 1 and target modelling error of 0, the developed model would fail to fulfil cross-validation requirements due to overfitting. In the exact RBFN, a very large spread, 125, was shown to be able to reduce overfitting; this purpose was served with use of a large spread, 20, and a fairly large target modelling error for the efficient RBFN.

REFERENCES

[1] Y. Bai and Q. Bai, Subsea engineering handbook: Gulf Professional Publishing, 2012.

[2] M. Mohammadzaheri, R. Tafreshi, Z. Khan, M. Franchek, and K. Grigoriadis, "An intelligent approach to optimize multiphase subsea oil fields lifted by electrical submersible pumps," Journal of Computational Science, vol. 15, pp. 50-59, 2016.

[3] L. Barrios and M. G. Prado, "Modeling Two-Phase Flow Inside an Electrical Submersible Pump Stage," Journal of Energy Resources Technology, vol. 133, p. 042902, 2011.

[4] J. Zhu and H.-Q. Zhang, "Numerical Study on Electrical-Submersible-Pump Two-Phase Performance and Bubble-Size Modeling," SPE Production & Operations, 2017.

[5] J. Zhu and H.-Q. Zhang, "Mechanistic modeling and numerical simulation of in-situ gas void fraction inside ESP impeller," Journal of Natural Gas Science and Engineering, vol. 36, pp. 144-154, 2016.

[6] M. Mohammadzaheri, R. Tafreshi, Z. Khan, M. Franchek, and K. Grigoriadis, "Modelling of Petroleum Multiphase fluids in ESPs, an Intelliegnt Approach," presented at the Offshore Mediternean Conference, Ravenna, Italy, 2015.

[7] R. Sachdeva, "Two-phase flow through electric submersible pumps," University of Tulsa, 1988.

[8] D. Sun and M. Prado, "Modeling gas-liquid head performance of electrical submersible pumps," Journal of Pressure Vessel Technology, vol. 127, pp. 31-38, 2005.

[9] J. Zhu, X. Guo, F. Liang, and H.-Q. Zhang, "Experimental study and mechanistic modeling of pressure surging in electrical submersible pump," Journal of Natural Gas Science and Engineering, 2017.

[10] D. Zhou and R. Sachdeva, "Simple model of electric submersible pump in gassy well," Journal of Petroleum Science and Engineering, vol. 70, pp. 204-213, 2010.

[11] L. R. Pineda, A. L. Serpa, J. L. Biazussi, and N. A. Sassim, "Operational Control of an Electrical Submersible Pump Working with Gas-Liquid Flow Using Artificial Neural Network " presented at the IASTED International Conference on Intelligent Systems and Control Campinas, Brazil, 2016.

[12] J. F. Lea and J. Bearden, "Effect of gaseous fluids on submersible pump performance," Journal of Petroleum Technology, vol. 34, pp. 922-930, 1982.

[13] M. Ghodsi, N. Hosseinzadeh, A. Özer, H. R. Dizaj, Y. Hojjat, N. G. Varzeghani, et al., "Development of Gasoline Direct Injector using giant magnetostrictive materials," IEEE Transactions on Industry Applications, vol. 53, pp. 521-529, 2017.

[14] M. Romero, "An evaluation of an electrical submersible pumping system for high GOR wells," University of Tulsa, 1999.

[15] R. Cirilo, "Air-water flow through electric submersible pumps," University of Tulsa, Department of Petroleum Engineering, 1998.

[16] J. Duran and M. Prado, "ESP Stages Air-Water Two-Phase Performance-Modeling and Experimental Data," 2003.

[17] A. Qallaf and M. Mohammadzaheri, "A Fuzzy Model to Estimate Head of Gaseous Petroleum Fluids Driven by Electrical Submersible Pumps," presented at the Engineering and Technology, Computer, Basic and Applied Sciences, Sydney, Australia, 2017.

[18] J. L. Turpin, J. F. Lea, and J. L. Bearden, "Gas-Liquid Flow Through Centrifugal Pumps—Correlation of Data," presented at the The Third International Pump Symposium, College Station, Texas, USA, 1986.

[19] R. Sachdeva, D. Doty, and Z. Schmidt, "Performance of Axial Electric Submersible Pumps in a Gassy Well," in SPE Rocky Mountain Regional Meeting, 1992.

[20] M. Mohammadzaheri, L. Chen, A. Ghaffari, and J. Willison, "A combination of linear and nonlinear activation functions in neural networks for modeling a de-superheater," Simulation Modelling Practice and Theory, vol. 17, pp. 398-407, 2009.

[21] M. Mohammadzaheri, L. Chen, and S. Grainger, "A critical review of the most popular types of neuro control," Asian Journal of Control, vol. 16, pp. 1-11, 2012.

[22] M. Beale, M. Hagan, and H. Demuth. (2017). Neural Network Toolbox™ User's Guide. Available: https://www.mathworks.com/

[23] G. C. Cawley and N. L. Talbot, "On over-fitting in model selection and subsequent selection bias in performance evaluation," Journal of Machine Learning Research, vol. 11, pp. 2079-2107, 2010.

[24] M. Mohammadzaheri, A. Mirsepahi, O. Asef-afshar, and H. Koohi, "Neuro-fuzzy modeling of superheating system of a steam power plant," Applied Math. Sci, vol. 1, pp. 2091-2099, 2007.

[25] M. Mohammadzaheri, A. Firoozfar, D. Mehrabi, and M. Emadi, "A Virtual Temperature Sensor for an Infrared Dryer," presented at the 9th IEEE-GCC Conference and Exhibition, Manama, Bahrain, 2017.

[26] A. Lendasse, V. Wertz, and M. Verleysen, "Model selection with cross-validations and bootstraps—application to time series prediction with RBFN models," Artificial Neural Networks and Neural Information Processing—ICANN/ICONIP 2003, pp. 174-174, 2003.

AUTHORS

Morteza Mohammadzaheri received his PhD from School of Mechanical Engineering, University of Adelaide, Australia in 2011. He has published/presented more than 90 peer-reviewed articles in technical journals and conferences. He is now an Assistant Professor of Dynamic Systems and Control at the Department of Mechanical and Industrial Engineering of Sultan Qaboos University, Oman.

Mojtaba Ghodsi received his B.Sc. degree in Mechanical Engineering from Isfahan University of Technology in 1999, the M.Sc. degree in Applied Mechanics from Tehran Polytechnic in 2001 and continued his research as Ph.D. (2007) and JSPS postdoctoral fellow (2009) in Precision Engineering Department of the University of Tokyo, Japan. Currently, he is pursuing his career at Sultan Qaboos University, Oman in department of Mechanical and Industrial Engineering. His main research interests include Smart Materials for Actuators, Sensors and Energy Harvesting, NDT and development of Mechatronic systems and devices. Dr. Ghodsi is a Member of IEEE and International Society of Optics and Photonics (SPIE).

Abdullah AlQallaf is an Assistant Professor with the Department of Electrical Engineering, Kuwait University. He received his Ph.D. degree in Electrical Engineering from the University of Minnesota—twin cities, St. Paul, MN, in 2009. Alqallaf's research interests are Microwave Imaging Techniques, Multimedia Signal Processing, Communication, Bioinformatics and Medical Image Analysis. Alqallaf is an IEEE Senior Member and an IEEE Board Member-Educational and Professional Activities-Kuwait section.

SPECTRUM SENSING APPROACH BASED ON QOS REQUIREMENTS IN WHITE-FI NETWORKS

Nabil Giweli, Seyed Shahrestani, Hon Cheung

School of Computing, Engineering and Mathematics,
Western Sydney University, Sydney, Australia

ABSTRACT

Cognitive Radio (CR) technology opens the door for the opportunistic use of the licensed spectrum to partially address the issues relevant to the limited availability of unlicensed frequencies. Combining CR and Wi-Fi to form the so called White-Fi networks, has been proposed for achieving higher spectrum utilization. This paper discusses the spectrum sensing in White-Fi networks and the impacts that they have on the QoS of typical applications. It also reports the analysis of such impacts through various simulation studies. We also propose such a sensing strategy that can adapt to the IEEE 802.11e requirements. The proposed strategy aims to enhance overall QoS while maintaining efficient sensing. Simulation results of the proposed mechanism demonstrate a noticeable improvement in QoS.

KEYWORDS

Cognitive Radio, Spectrum Sensing, White-Fi, IEEE 802.11af, IEEE 802.11e, QoS.

1. INTRODUCTION

Traditionally, the radio spectrum is statically divided into frequency bands, most of which are licensed to organizations and companies usually referred to as primary users (PUs). A small number of frequency bands are unlicensed including the unlicensed Industrial, Scientific, and Medical (ISM) bands, which are used by a variety of indoor and short-range wireless communication systems, such as Wi-Fi, Bluetooth, and Zigbee. These free unlicensed frequencies are not sufficient to handle the rapidly growing number of wireless devices using these unlicensed bands. Also, modern applications running on these devices demand more bandwidth. These modern applications usually involve multimedia communications, e.g., media streaming, video conferencing, and interactive gaming. One of the promising solutions to increase radio frequency spectrum availability to these wireless devices is to add the Cognitive Radio (CR) capability to such devices. With CR capability, a wireless device can operate opportunistically, as a Secondary User (SU), over licensed frequency channels when they are unused, i.e., White Spaces (WS) or spectrum holes in the licensed bands. Television (TV)bands are the most attractive frequency ranges for such opportunistic use of the spectrum by SUs. This because the TV bands show high availability of WSs and their schedule use by the PUs can be obtained through Geolocation Database (GDB) services[1]. Moreover, the TV spectrum is located below 1 GHz. Compared to the higher ISM bands, these frequencies offer more desirable propagation characteristics. An IEEE-802.11 protocol with CR capability is often referred to as CR Wi-Fi, White-Fi, Wi-Fi Like or IEEE-802.11af. The IEEE 802.11af is the first draft standard for CR networks based on IEEE

David C. Wyld et al. (Eds) : CNDC, DPPR, AIAA, WiMNET, ACITY - 2017
pp. 91– 100, 2017. © CS & IT-CSCP 2017 DOI : 10.5121/csit.2017.71609

802.11 to operate in TV WS[2]. The White-Fi devices can operate either in ISM channels or TV WSs based on the IEEE 802.11af standard. Figure 1 shows the main approaches for assessing the potential operation bands along with their related basic conditions and required actions.

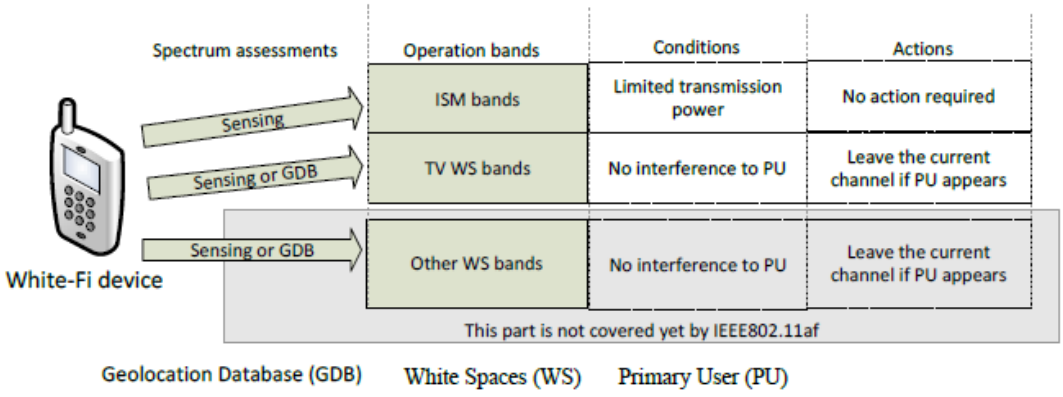


Figure 1. White-Fi potential operation bands and their required conditions and actions

Primarily, the CR capabilities and requirements are established at the Physical (PHY) and Medium Access Control (MAC) layers of wireless systems. Spectrum sensing is one of the most important functions in CR to identify the available spectrum holes and to protect the PU from interference. Conducting sensing has its impact on transmission delays and throughput. The Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) mechanism is used in wireless devices based on various 802.11 standards to share the ISM bands[3]. The concept is that a Wi-Fi device checks the channel occupancy before transmitting over it. Typically, this checking is accomplished by using a simple sensing technique, e.g., Energy Detection (ED), where the energy of the channel is measured and compared to a predefined threshold. If the measured energy level exceeds the threshold, the channel is in use by another device; otherwise, it is idle.

The remaining parts of this article are organized as follows. In section 2,some of the previous related work are reviewed andcompared to the contributions of this study. Correlation between QoS requirements and sensing is discussed in Section 3.Then the impacts of sensing duration on QoS are identified and analyzed by simulations in Section 4. In Section 5,the proposed solution for enhancing QoS in White-Fi networks by selecting sensing strategy based on QoS requirements is demonstrated. Conclusions and future works are described in Section 6.

2. RELATED WORK AND MOTIVATIONS

The success of White-Fi technology highly depends on the QoS level that it can offer for various communication applications. Although the QoS in CR networks, in general, has drawn more attention recently, White-Fi networks have received a minor portion of that attention. Articles that have been published to study QoS issues in White-Fi networks based on sensing are related to our work in this paper. The most related work is that where the sensing duration and its effect on QoS are considered.

The effect of the sensing duration on the access delay of an SU was analyzed when the Request-to-Send/Clear-to-Send(RTS/CTS)mechanism was enabled for spectrum access in[4]. The optimal sensing length was formulated as a function of the false alarm and miss detection probabilities and the number of SUs contending for the same access point. The ED was assumed as the sensing method and the maximum sensing duration used for analysis was 25ms.The sensing was only

conducted at the beginning of the contention period and before sending an RTS packet. The study found that the access delay of an SU depends on the sensing length, and its optimal value varies based on the number of contending SUs. The relation and the possible trade-off between the spectrum sensing duration and achievable throughput of SUs were addressed in several studies[5, 6]. The approximated analytical formulation of the optimally saturated throughputs for multiple SUs based on CSMA/CA mechanism was proposed in[14]. The discrete-time Markov chain was used to model this formulation for different false alarm probabilities and a varying number of SUs. Their numerical results showed the significant role of sensing in improving the SUs throughputs and how the saturated throughput was affected by the sensing false alarm probability. Their analysis was approximate and under typical parameters as their primary aim was to conduct an initial study on the performance of CR techniques used in 802.11-based networks. In another similar study by other authors, the authors' aim was to find the optimal sensing duration that achieves the maximum throughput under unsaturated traffic conditions in[6].The discrete-time Markov chain model was used by the authors to model their proposed MAC structure for analyzing the performance of SUs. The sensing was conducted by SUs only at the beginning of the contention period when one access point and several SUs exist. The optimal sensing duration was investigated within the range of 0.5 ms to 3 ms for different number of SUs (5, 10, 15 and 20) of the various queuing probabilities and contention window sizes. Other conditions, such as the false alarm probability, detection probability and single-to-noise ratio (SNR), were assumed constant and the same for all SUs. Under the aforementioned assumptions, the optimal sensing duration was found to be around 2 ms for all the simulated scenarios under certain assumptions and conditions[6]. Therefore, an optimal sensing duration can be found for maximum throughput under unsaturated traffic conditions.

The studies mentioned above were solely based on the use of ED as their sensing method and relatively small sensing length. Practically, the ED method performs poorly in low SNR environments and cannot distinguish between PU signals and other SU signals. The use of higher accuracy sensing methods implies the need of longer sensing. In our study, a wider range of sensing durations is considered to reflect the potential use of more complex sensing methods.

3. CORRELATION BETWEEN QOS REQUIREMENTS AND SENSING

The IEEE 802.11e standard is proposed to enhance QoS in IEEE 802.11 networks [7]. As applications have different requirements, in 802.11e, the frames belonging to different applications are prioritized with one of the eight user priority (UP) levels. In contrast, previous IEEE 802.11 standards use the Distributed Coordination Function (DCF) mechanism at MAC layer where the best-effort service is provided equally to all traffic streams from different applications to access the medium. In IEEE 802.11e, the Hybrid Coordination Function (HCF) is used for prioritizing traffic streams to enhance QoS on top of the DCF. The HCF accommodates two medium access methods, i.e., a distributed contention-based channel access mechanism, called Enhanced Distributed Channel Access (EDCA), and a centralized polling-based channel access mechanism, called HCF Controlled Channel Access (HCCA). Based on the UP, the EDCA defines four Access Categories (AC); voice (AC_VO), video (AC_VI), best-efforts (AC_BE) and background (AC_BK). These categories are assigned different priorities ranging from highest to lowest respectively. The category AC_VO has top priority and is usually given to traffic carrying voice information. It is followed by the AC_VI category for video traffic and then the AC_BE category for data traffic. The category AC_BK has the lowest priority and is usually assigned to unnecessary data traffic.

Each AC has a Contention Window (CW) that has a specified minimum size and maximum size, i.e., $CW_{min}[AC]$ and $CW_{max}[AC]$. Also, an Arbitration Inter-frame Space (AIFS) value and a Transmit Opportunity (TXOP) interval are used to support the QoS prioritization [8]. Instead of

using fixed Distributed Inter-Frame Space (DIFS), also called DCF inter-frame space, the AIFS[AC] value is a variable value calculated based on the AC. For instance, possible valuesfor AIFS[AC] are; AIFS[AC_BK], AIFS[AC_BE], AIFS[AC_VI] or AIFS[AC_VO]. The AIFS value determines the time that a node defers access to the channel after a busy period and before starting or resuming the back-off duration. Hence, the time for a station to wait for the channel to become idle before it starts sending data is calculated based on the AC category of the data AC [9]. However, the Short Inter-frame Space (SIFS) value is used as the shortest Inter-frame Space (IFS) value for transmitting high priority frames, such as DATA Acknowledgment frames. Therefore, the higher priority frames access the operational channel earlier than other frames in the same transmission queue.

White-Fi users and SUs based on other different wireless technologies may coexist in the same available WSs. Under this situation, a White-Fi user needs to distinguish between three types of users, i.e., the PU, the other White-Fi users and the other non-White-Fi SUs. The simple sensing methods, e.g., ED, cannot distinguish between these different signals. Although advanced sensing, such as Matched Filter Sensing (MFS) method, may distinguish between signals when prior information about these signals is available, higher sensing duration is required.

In the case of White-Fi, increasing sensing duration will impact the effectiveness of the IEEE802.11e standard on improving QoS in IEEE802.11 networks. The impact of the sensing operation should be investigated under different settings of the associated parameters. For the frames belonging to the categories of AC_VI and AC_VO, the AIFS, and CW values are set smaller than for the frames of the categories AC_BE and AC_BK, to reduce the delays. In White-Fi networks based on sensing, increasing sensing duration for more accurate sensing can result in compromising the IEEE 802.11e mechanism, as shown by the simulation results in the following sections.

4. SIMULATION STUDIES: IMPACTS OF SENSING DURATION

To study effects of sensing function on the frame transmission delays that impacts the QoS of applications running on a White-Fi device, a simulation tool Modeler 18.0 from Riverbed (formerly Opnet) [10] is used. Modeler 18.0 supports different types of applications and network traffics. However, CR networks are not implemented yet in Modeler 18.0. Hence, we have customized the standard Wi-Fi node to include a sensing function with different sensing periods to simulate the behavior of a White-Fi node. The settings of some common parameters of all simulation scenarios are shown in Table 1. The IEEE 802.11e is supported in all scenarios with the settings illustrated in Table 1. Simulations are conducted under different sensing durations and for different application categories. The main three types considered are the voice traffic, video conferencing traffic and email traffic. The result values are captured under bucket mode with a sample mean of 100 values per a result statistic. In the Bucket mode, the data is collected at all of the points over the time interval or sample count into a “data bucket” and generates a result from each bucket. The wireless delay, simply called delay in this article, represents the end-to-end delay of all the data packets that are successfully received by the MAC layer and forwarded to the higher layer in a node. The media access delay is the sum of delays, including queuing and contention delays of all frames transmitted via the MAC layer.

For each frame, the media access delay is calculated as the duration between the time when the frame is placed in the transmission queue until the time when the frame is sent to the physical layer for the first time. On other words, the media access delay is the time of processing a packet at the MAC layer. For a voice application, voice traffic is generated as IPv4 unicast traffic flows between the nodes. For generating email and video conferencing traffic, a server is used to run these applications in the infrastructure network scenarios.

Table 1 Simulation parameters settings

Parameter	value
Data rate	26 Mbps / 240Mbps
Buffer Size	256000 bits
Maximum Transmitter A-MSDU size	3839 bytes
Maximum Acceptable A-MSDU size	8191 bytes
EDCA Parameters:	
Voice:	CW_{min}= (PHYCW_{min} + 1) / 4-1 CW_{max}= (PHYCW_{min} + 1) / 2-1 AIFSN = 2 TXOP = One MSDU
Video:	CW_{min}= (PHYCW_{min} + 1) / 2-1 CW_{max}= PHYCW_{min} AIFSN = 2
Best Effort:	CW_{min}=PHYCW_{min} CW_{max}=PHYCW_{max} AIFSIN=3
Background:	CW_{min}=PHYCW_{min} CW_{max}=PHYCW_{max} AIFSIN = 7

4.1. Analyzing the sensing duration effect on different type of applications

In this subsection, the effect of sensing duration is analyzed for various applications. Three scenarios are implemented for that purpose; voice scenario, email scenario and video scenario. The sensing is conducted for all frames except response frames in all scenarios. For voice scenario, four nodes Ad Hoc network are used. Moreover, IPv4 unicast voice traffic is generated amongst all the four nodes. Several simulations are conducted for different sensing durations from 1 ms to 300 ms. Also, the simulation is run when the sensing is neglected, i.e., 0 ms. The average delay for each of the sensing duration starts to increase sharply from the beginning till the 900 seconds of simulation time. It becomes more stable afterward for all sensing durations.

In the email scenario, an email server is added for simulating heavy email traffic between the server and other four nodes. Instead of operating as an ad hoc network in the previous scenario, the network in this scenario operates in an infrastructure mode with the server also acting as an AP. Several simulations are run under different sensing periods from 1 ms to 300 ms. The simulated operation time of the network is an hour for each of these sensing durations. The average measured delay is 0.35 seconds when the sensing length is 300 ms. The average delay is between 0.3 and 0.1 seconds for sensing durations between 250 ms and 100 ms while it is less than 0.1 seconds for sensing lengths less than 50 ms. These results show that the delay is less with heavy email traffic than the voice traffic.

The video application scenario is similar to the email application one, except that the added server is used for providing video conferencing application to the other four nodes. The server is used to generate high-resolution video conference traffic between the four nodes through the server. The simulations are conducted for different sensing durations from 1 ms to 300 ms. The average delay in this scenario has not exceeded 0.1 seconds even when the sensing length is 300 ms. The average delay in this scenario does not follow proportional relation between the average delay and sensing duration.

4.2. Observations and comparisons

The comparison between the three traffic types; voice, email, and video, clearly illustrates that the voice traffic experience the highest average delay as shown in Figure1. Compared to the other traffics, the voice traffic is more sensitive to the length of the sensing period. The average delay in email traffic is proportional to the length of the sensing duration, but with a smaller slope. Thus, the email traffic is less affected by the sensing duration than voice traffic. Moreover, email applications are not sensitive to delay. The video traffic is less sensitive to sensing length compared to the other traffics and the average delay and the sensing duration is not in a constant proportionality relation. The results in Figure1 demonstrate that the same sensing strategy has different effects on the traffics from different applications.The diverse traffics used in these scenarios follow different aggregation settings. It can also be seen in Figure1 that sensing durations less than 50 ms cause an average delay of fewer than 0.2 seconds, which is acceptable in several applications. However, a longer sensing time could be required for achieving higher sensing accuracy. Although IEEE802.11e was enabled in the simulations, the results show that the voice and video traffic gain no benefit from it because of the impact of extended sensing time. Therefore, sensing duration and frequency should be conducted in a way that preserves the aim of improving QoS of different applications by using IEEE802.11e. In the next section, our proposed solution for enhancing QoS is discussed and evaluated by simulations.

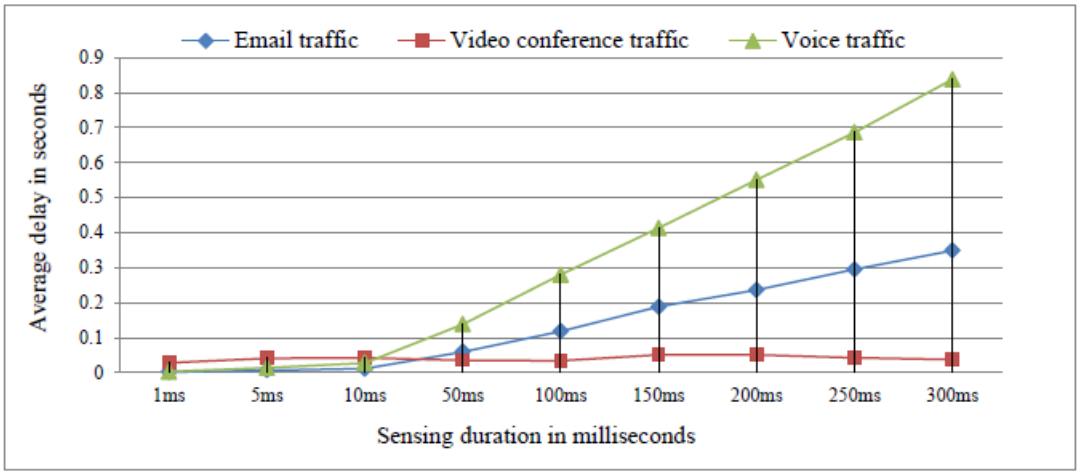


Figure1. Comparison between the average delays of different applications traffics for different sensing durations

5. SELECTING THE PROPER SENSING STRATEGY FOR ENHANCING QOS

The ED method is commonly used for sensing on CSMA/CA based networks because of its simplicity and low overhead. However, in a CR environment, the main ED drawback, in particular, is its inability to recognize PU signal among other signals, reduce its effectiveness. The MFS method can overcome this drawback and provide higher detection accuracy at the cost of more sensing time, complexity and power consumption. Therefore, the sensing method should be selected during operation by trading off between the sensing requirements and its implications. Such approach imposes that the White-Fi device supports a set of different effective sensing techniques where each one of these methods is suitable for a particular set of operation requirements.

Towards enhancing QoS in White-Fi networks, a mechanism of selecting the proper MAC operation settings and sensing strategy that suits the QoS application requirements is proposed.

The required QoS for different applications is classified into four levels based on the four ACs in IEEE 802.11e as shown in Table 2. The sensing strategies are classified into four types; Coarse (C), Moderate (M), Fine (F) and Extra Fine (EF) sensing. Each sensing type will be conducted based on the AC of the frame to be transmitted as shown in Table 2. The proposed sensing duration S_d range for each sensing type is chosen based on our previous study and classification of different sensing methods [11]. In the C sensing type, the sensing duration S_d is less or equal 1 ms to give higher priority to AC_VO frames with less impact on the delay. However, the sensing method that can be used within this short time is the blind sensing method, such as ED. Consequently, the C sensing type cannot distinguish between PU and other SU signals, and poor accuracy is expected at low SNR. Under the M sensing type, the sensing methods that can be used for an S_d larger than 1 ms up to 5 ms are similar to those used in the C sensing type with a slight improvement in sensing accuracy, particularly at low SNR. Therefore, AC_VI frames have less priority than AC_VO to win the contention window and more delay is predicted. For the F sensing type, the sensing duration S_d is larger than 5 ms up to 50 ms. Thus, sensing methods that can differentiate PU singles from other signals can be used. On the one hand, the F sensing type enables more utilization of WSs. On the other hand, it causes higher delay. In addition, conducting F sensing before AC_BE frames maintains the desired priority of these frames. As AC_BK frames have the lowest priority, EF sensing should be carried out before them for a sensing duration S_d larger than 50 ms. In the EF sensing type, sophisticated sensing methods can be used to achieve high sensing accuracy even under low SNR. Hence, higher spectrum utilization can be achieved under the cost of higher delay. When F and EF sensing recognize the appearance of PU in the current WS channel, the White-Fi device must scan for other vacant channels and leave the currently occupied one. Otherwise, when the current channel is found busy by other SUs, the device can continue use and share the current channel with other SU.

Table 2 Sensing strategy based on frame access category

Frame Access Categories	QoS requirements	Sensing type	Sensing duration (S_d) in ms	Sensing characteristics
Voice (AC_VO)	Highest priority (low latency, e.g., voice call, audio streaming)	Coarse (C) sensing	$S_d \leq 1$	Only blind sensing can be used. Cannot distinguish between PU and SU. Poor performance in low SNR.
Video (AC_VI)	Second highest priority (e.g., video conferencing, streaming)	Moderate (M) sensing	$1 < S_d \leq 5$	More advanced sensing methods but still not capable of distinguish between PU and SU. Moderate sensing accuracy.
Best Effort (AC_BE)	Low priority (traffic less sensitive to latency, e.g. web surfing)	Fine (F) sensing	$5 < S_d \leq 50$	Some sensing methods that can distinguish between PU and SU can be used. High sensing accuracy.
Background (AC_BK)	Lowest priority (no strict latency) (e.g., print jobs, email, etc.)	Extra Fine (EF) sensing	$S_d > 50$	Sensing methods that can distinguish between PU and SU can be used. High sensing accuracy even in low SNR.

In the case of C and M sensing, the sensing outcome cannot be certain about the PU presence. Hence, as long as the appearance for the PU is not certain, other factors should be considered before conducting handoff procedure to another available channel.

5.1 Evaluation of the sensing selection strategy

In this section, we demonstrate the QoS enhancement that can be achieved by considering the application requirements in selecting the proper sensing strategy. Two scenarios were

implemented to compare between fixed sensing approach and selecting the proper sensing based on the QoS requirement approach. The network in both scenarios was the same with four nodes and one server. Three applications, i.e., voice, video and email, were configured to run simultaneously on all nodes for both scenarios. The simulated environment was implemented to reflect a real-life scenario where the wireless device is used to run simultaneously IP telephony voice application, high resolution video conferencing and heavy load email application. The sensing was conducted for all frames except response frames in both scenarios. The nodes in the first scenario were implemented with the same sensing strategy in Section.4.1.Hence, fixed sensing duration was used for all different AC frames. In the second scenario, the nodes were implemented to use different S_d based on the AC of the frame to be sent as proposed in Section 5.

The first scenario was simulated under four different sensing durations in each run where the sensing duration was changed to one of these values: 1 ms, 5 ms, 50 ms and 100 ms. As each one of these values fallen in different sensing type proposed in Table 2. In the second scenario, the nodes were implemented to select S_d based on the AC of the frame, such as S_d = 1 ms for AC_VO, S_d = 5 ms for AC_VI, S_d = 50 ms for AC_BE and S_d = 100 ms for AC_BK. The achieved average throughput for BE frames shown in Figure 3demonstrates that the sensing based on selection approach can achieve higher throughput compared to fixed sensing approach even when the sensing duration is fixed to 1 ms. The average media access delay that was experienced by voice frames is illustrated in Figure 4. The sensing based on selection experienced average media access delay less than the fixed sensing when the sensing length was 50 ms or 100 ms. The fixed approach with low sensing duration, e.g., 1 ms or 5 ms, resulted in less medium access delay but with less achieved average throughput compared to the sensing based on selection. Moreover, the fixed small sensing duration most likely results in inefficient sensing and less protection to PU signals that may not comply with requirements. Therefore, the proposed sensing based on selection approach can enhance the archived QoS of White-Fi devices based on spectrum sensing.

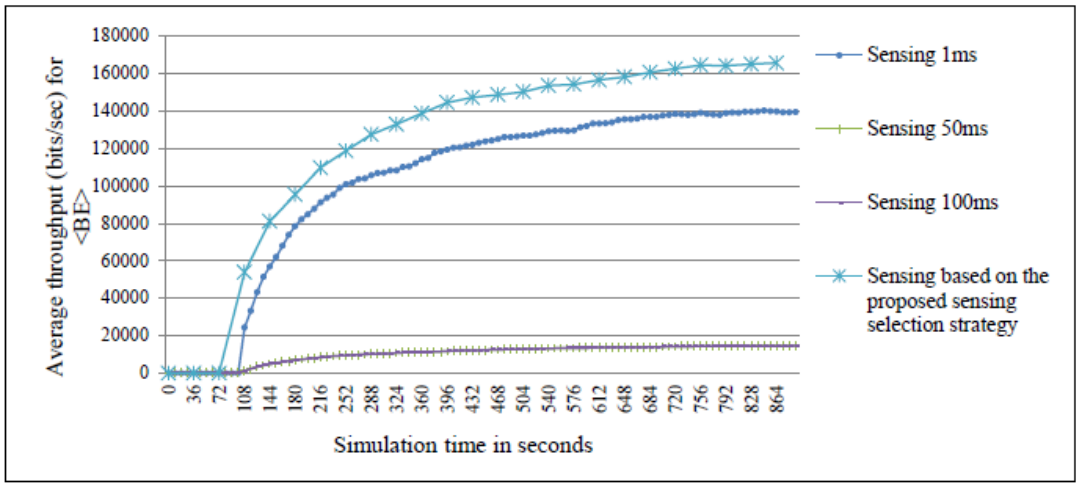


Figure 3. Average throughput for different sensing strategies when voice, video and email applications are running

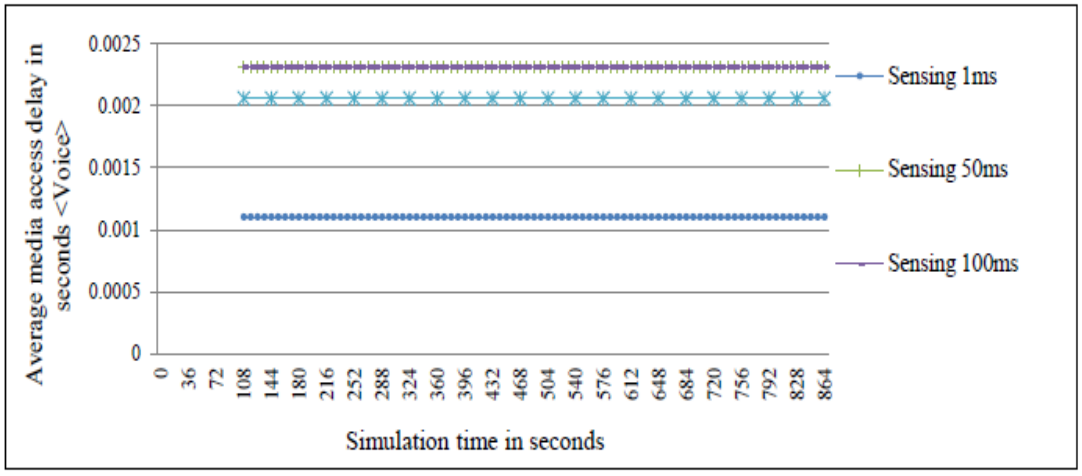


Figure 4. Average voice application media access delay for different sensing strategies when voice, video and email applications are running

6. CONCLUSIONS AND FUTURE WORK

Increasing sensing duration is required to achieve advanced sensing accuracy in CR networks. In this work, we studied the need of different sensing durations and its impact on QoS of various applications in the White-Fi networks. Our simulation results show that voice traffic is more affected by the sensing operation than video and email traffics. That means the IEE802.11e mechanism performs poorly when sensing duration is increased for higher accurate sensing. To address this issue, a sensing approach that selects the sensing parameters based on the 802.11e frames categories is proposed. The proposed sensing strategy resulted in QoS improvement while attempting to preserve higher PU protection and spectrum utilization. Our future research is aiming to develop our sensing strategy by considering more factors for better performance.

REFERENCES

[1] D. Gurney, G. Buchwald, L. Ecklund, S. Kuffner, and J. Grosspietsch, "Geo-location database techniques for incumbent protection in the TV white space," in New Frontiers in Dynamic Spectrum Access Networks, 2008. DySPAN 2008. 3rd IEEE Symposium on, 2008, pp. 1-9.

[2] IEEE, "IEEE Standard for Information technology - Telecommunications and information exchange between systems - Local and metropolitan area networks - Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 5: Television White Spaces (TVWS) Operation," IEEE Std 802.11af-2013 (Amendment to IEEE Std 802.11-2012, as amended by IEEE Std 802.11ae-2012, IEEE Std 802.11aa-2012, IEEE Std 802.11ad-2012, and IEEE Std 802.11ac-2013), pp. 1-198, 2014.

[3] IEEE, "IEEE Standard for Information Technology - Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks - Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," IEEE Std 802.11-2007 (Revision of IEEE Std 802.11-1999), pp. 1-1076, 2007.

[4] L. Dong-Jun and W. Hyuk, "Spectrum Sensing Time Minimizing Access Delay of IEEE 802.11-Like MAC in Cognitive Radio Networks," Communications Letters, IEEE, vol. 15, pp. 1249-1251, 2011.

[5] D. T. C. Wong and F. Chin, "Sensing-Saturated Throughput Performance in Multiple Cognitive CSMA/CA Networks," in Vehicular Technology Conference (VTC 2010-Spring), 2010 IEEE 71st, 2010, pp. 1-5.

[6] H. Ko, J. Lee, and C. Kim, "The Optimal Spectrum Sensing Time for Maximizing Throughput of 802.11-Based MAC Protocol for Cognitive Radio Networks Under Unsaturated Traffic Conditions," Wirel. Pers. Commun., vol. 77, pp. 1397-1414, 2014.

[7] IEEE, "IEEE Standard for Information technology--Local and metropolitan area networks--Specific requirements--Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications - Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements," IEEE Std 802.11e-2005 (Amendment to IEEE Std 802.11, 1999 Edition (Reaff 2003), pp. 1-212, 2005.

[8] I. Inan, F. Keceli, and E. Ayanoglu, "Modeling the 802.11 e enhanced distributed channel access function," in Global Telecommunications Conference, 2007. GLOBECOM'07. IEEE, 2007, pp. 2546-2551.

[9] G. Bianchi, I. Tinnirello, and L. Scalia, "Understanding 802.11 e contention-based prioritization mechanisms and their coexistence with legacy 802.11 stations," Network, IEEE, vol. 19, pp. 28-34, 2005.

[10] Riverbed, "Riverbed Modeler," 18.0.1 ed: Riverbed Technology Inc., 2014.

[11] N. Giweli, S. Shahrestani, and H. Cheung, "Selecting the Sensing Method in Cognitive Radio and Future Networks: A QoS-Aware Fuzzy Scheme," in 2015 IEEE International Conference on Data Science and Data Intensive Systems, Sydney, Australia, 2015, pp. 497-504.

SCOPE: A LIGHTWEIGHT CRYPTOGRAPHIC APPROACH FOR PRIVATE COPE DATA CODING

Ngoc Hong Tran[1], Cao Vien Phung[2], Binh Quoc Nguyen[1], Leila Bahri[3] and Dung Hai Dinh[1]

[1]Vietnamese German University, Vietnam
[2]Technische Universität Carolo-Wilhelmina zu Braunschweig, Germany
[3]KTH, Sweden

ABSTRACT

In this paper, we investigate a simple but effective coding mechanism, namely Coding Opportunistically (COPE) [1], and the privacy violations which are likely to happen to COPE. The COPE data, from the source node through the network to its destination node, can be easily learned by the surrounding nodes, particularly, intersecting nodes and neighbour nodes. This leads to a serious consequence in leaking the node identity and its private data. In order to cope with the mentioned issues, we can apply cryptographic schemes. However, the security solutions often exploit the public key schemes which nowadays run more slowly and give the longer encrypted values as the effects of the key size increase to guarantee the algorithm complexity. Hence, while the coding mechanism aims to decrease the bandwidth consumption by aggregating (i.e., coding) multiple packets in the network, the security solutions increase the data quantity. However, a necessary needs of combining data coding mechanism and cryptographic algorithm is raised for both preserving privacy and optimizing bandwidth use at the same time as mentioned above. In this paper we thus propose SCOPE, a lightweight privacy preserving approach able to support nodes running the COPE protocol in a secret way, by adopting the Elliptic Curve Cryptography (ECC) homomorphic encryption algorithm. The proposal's effectiveness and efficiency are proved through a variety of experiments.

KEYWORDS

Network Coding, Peer-to-Peer, Homomorphic Encryption, Elliptic Curve Cryptography (ECC).

1. INTRODUCTION

In order to optimize the network performance, particularly reducing the bandwidth consumption in a wireless network, there exist several techniques, such as network load balancing [2], routing optimization [2] [3] [5]. Another approach is the *network coding* paradigm that aims at increasing the performance in the network. In network coding, intermediate nodes, in a transmission network, can combine multiple native packets from different sources using simple operations, such as the Exclusive-OR operator, into a coded packet and then broadcast the coded packet in a single transmission instead of simply forwarding each native packet one by one. Therefore, the number of transmissions is reduced, and the network capacity is increased.

David C. Wyld et al. (Eds) : CNDC, DPPR, AIAA, WiMNET, ACITY - 2017
pp. 101– 119, 2017. © CS & IT-CSCP 2017 DOI : 10.5121/csit.2017.71610

One of the first practical network coding techniques proposed as an effective forwarding architecture for wireless networks is the Coding Opportunistically (COPE) model [1]. COPE uses a two-hop coding pattern in order to identify the packets that can be coded together. Even though COPE is simple and effective, there still exist some short comings that lead to security and privacy issues. COPE nodes can overhear packets from their neighbours. That is, each node can read any content of packets transmitted in its radio range. Moreover, the intermediate nodes which is in charge of coding the packets are mostly likely to misuse the retrieved information for the malicious purposes. As an example, see the sample network represented in Figure 2. We can see that N_1 can listen to and get packets which N_2 and N_4 send to N_5. If N_1 is a malicious node, it can replay data from N_2 and N_4 for the other transactions, or modify the data. Therefore, *there is a need for protecting packets against transmitting nodes that are able to listen to transmissions from their neighbours, both in terms of confidentiality as well as of integrity.* In addition to this, COPE relies on opportunistic listening by which transmitting nodes can be aware of the packet queue lists of their neighbours to increase the chances of finding packets that can be coded. This introduces more privacy issues when considering intruders or malicious transmission nodes.

In order to address the privacy issues related to COPE, we propose in this paper a secure COPE (i.e., SCOPE) protocol that exploits cryptographic techniques to address the privacy issues related to the plain COPE. More precisely, and not to affect the desired goal of increasing the network capacity aimed from COPE, we adopt the additive homomorphic Elliptic Curve Cryptography (ECC) [12] that, in addition to providing the needed security requirements, is also lightweight and fits our considered scenarios for secure and still good performing network coding using COPE. Our experimental results demonstrate that the performance cost of SCOPE is negligible.

The rest of this paper is organized as follows: in Section 2 we review the related literature and position our suggested work. In Section 3 we provide background information on the required concepts related to both COPE and to ECC encryption. In Section 4 we formalize our security model, and in Sections 5 and 6 we present the SCOPE architecture and protocol, respectively. In Section 7 we provide the experimental evaluation of SCOPE. The security property will be expressed in Section 8. We finally conclude the paper in Section 9.

2. RELATED WORKS

Today, coding opportunity is one of the hot topics in network coding technique. As such, after COPE, several other interesting research works focused on designing network coding conditions that can increase the coding opportunities, hence better improve the network capacity. For instance, we find BEND [6] (BEND is not an acronym) with non-intersecting two-hop flows and DCAR (Distributed Coding-Aware Routing in Wireless Networks) [7] with intersecting more than two-hop flows. DODE (Distributed Opportunistic and Diffused Coding in Multihop Wireless Networks) [8] has combined the advantages of COPE, BEND and DCAR to still increase the coding chances. DODEX (Distributed and Diffused Encoding with Multiple Decoders) [9] with multiple encoder has been developed from DODE. On the other hand, Re-encoding of a coded packet is agreed in DODEX+ (Distributed and Diffused Encoding with Multiple Encoders and Multiple Decoders) [13] to detect more codable flows. Besides that, some applications of linear programming to optimize the throughput of COPE [5] and BEND [3] to create more opportunities for coding and provide a better total network throughput. These issues have been addressed by considering the best paths which not only have more coding chances but also avoid wireless interference among network nodes in the network.

With the richness of the literature in terms of network coding schemes that aim at improving the coding opportunity rate, COPE, as well as most of its improved successors, suffer from privacy and security related issues.One of the first works providing an analysis to secure network coding

is in [14]. In this work, an eavesdropper able to see information in a single source scenario has been considered. Subsequently, in order to allow for secure network coding different models have been proposed [15]. The early works started by proposing a *wiretap* model, where the main idea consists at collecting subsets of nodes in a network coding system in wiretap networks such that each wiretapper has access to only one of these subsets [16].After that, the focus has shifted to addressing network coding security and privacy issues using different cryptographic schemes. The challenge, however, is to ensure the security and privacy of a network coding protocol (such as COPE), without much sacrificing the initial goal of increasing network capacity. It is indeed well known that cryptographic schemes do mostly come with huge costs both in terms of size and processing time.

Of the works available, we find the model proposed in [17], where the authors focus on the privacy preservation issue in terms of preventing traffic analysis and tracing in multi-hope wireless networks. The authors deploy the Paillier [22] homomorphic encryption scheme, which is based in the usage of large primes. The related consequence is on performance as operations done on large prime numbers is quite costly.

In [18], the authors have focused on the shortcomings of the Homomorphic Message Authentication Code (H-MAC) in terms of its vulnerability to pollution attacks, and in the context of being used to secure communications in transmission networks. The authors have proposed an improvement to minimize both data pollution and tag pollution attacks related to H-MAC, by introducing two types of tags, one related to verifying the integrity of the packet and the other related to checking the integrity of the H-MAC itself. In their analysis, they have showed that their method not only decreases the probability of tag pollution but it also results in decreasing the related bandwidth overhead. However, the overhead remains considerable. Many other works, such as [20] [21], have also studied the usage of H-MAC based techniques to secure network coding; however, the overhead remains quite high and the problem of data and tag pollutions often remain an issue. Moreover, we can also find the work in [19] where the authors have identified a flaw in H-MAC in general and have provided a corresponding inaccuracy in its formal security proof. This keeps it to doubt whether H-MAC based solutions are worthwhile in practice or not.

Differently from the available works in cryptographic secure network coding schemes, we proposed in this paper applying the lightweight ECC scheme to address the security and privacy issues in COPE.

3. BACKGROUND

COPE is a forwarding architecture for wireless mesh network. It is quite simple and effective in reducing the quantity of transmissions. Whereas, encryption algorithms are strong enough to secure data as well as to preserve the privacy against the curious attackers. We introduce herewith COPE and additive homomorphic encryption in the sense of highlighting their combination strength in tackling the security requirements mentioned in Sections 1 and 4.2.

3.1. COPE - Coding Opportunistically

Let us briefly describe COPE protocol. COPE has been the first practical network coding mechanism applied in wireless networks, aiming at reducing a significant quantity of transmissions. Thus the wireless throughput consumption can be optimized by coding several packets into one are forwarding them through a single transmission.

a) Without COPE. b) With COPE.

Figure1. Simple data transmission model for three nodes.

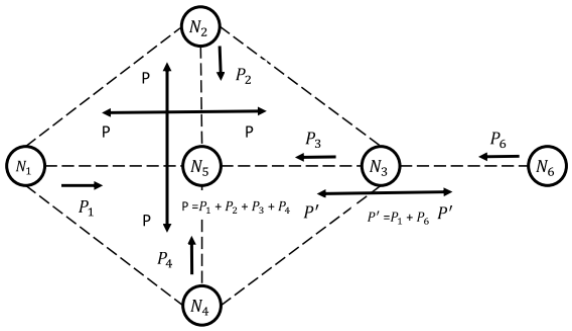


Figure 2. An example of network with COPE

3.1.1. COPE Protocol.Initially, with the standard COPE (see Figure 1), only 2-hop flows (i.e., including three nodes connected and involved in a transaction) are considered. Let consider 2-hop flows $N_1 \rightarrow N_2 \rightarrow N_3$ and $N_3 \rightarrow N_2 \rightarrow N_1$ in Figure 1. Two nodes N_1 and N_3 want to make a packet transfer to each other. But they locate out of the radio ranges of each other. Consequently, they cannot directly communicate. However, both of them can send their packets through node N_2 which places in the both of radio ranges. Then, N_2 becomes an intermediate node in charge of relaying packets between N_1 and N_3. Let us see Figure 1a which describes a protocol not applied with COPE. Let P_1, P_3 be packets sent, respectively, from N_1 to N_3 and vice versa. Hence, there are four unicast transmissions (i.e., P_1:$N_1 \rightarrow N_2$, P_3:$N_3 \rightarrow N_2$, P_3:$N_2 \rightarrow N_1$, P_1:$N_2 \rightarrow N_3$) in this protocol. In case of COPE protocol (see Figure 1b), the packets from N_1 and N_3 are considered as strings of bits, and aggregated by N_2, using the operator Exclusive--OR denoted as '+', to produce a coded packet $P=P_1+P_3$. Hence, N_2 is called ***encoder***, whereas N_1 and N_3 are ***decoders***. As a result, the number of transmissions drops down to three, including two unicast transmissions (i.e., P_1: $N_1 \rightarrow N_3$, P_3:$N_3 \rightarrow N_1$) and 1 broadcast transmission (i.e., P:$N_2 \rightarrow \{N_1, N_3\}$).

Example 1. .Let us give an example by considering Figure 2. A link connecting two nodes, depicted as a dash line, indicates that they are in radio ranges of each other, therefore they are so-called neighbours. Initially, nodes N_1, N_2,N_3 and N_4 send packets P_1, P_2,P_3 and P_4, respectively, to nodes N_6, N_4, $N_1$and N_2, via node N_5. N_5 has to wait to adequately get four native packets P_1, P_2, P_3, P_4, respectively, from N_1, N_2, N_3 and N_4. N_5 then aggregates them to a coded packet $P = P_1 + P_2 + P_3 + P_4$ and broadcasts P to N_1, N_2, N_3 and N_4 in a single transmission. Nodes N_1, N_2, N_3 and N_4 can decode to obtain its expected packet. They can do this, since each of nodes can catch the packets from its neighbours by overhearing their signals in the air. For example, N_3 can overhear P_2 and P_4 from N_2 and N_4. By XOR-ing P with its generated-and-sent packet P_3 and the overheard packets P_2 and P_4, N_3 obtains $P_1 = P + P_2 + P_3 + P_4$. Similarly, N_6 wants to send packet P_6 to N_5 via N_3. N_3 receives P_6, then, creates the aggregation $P' = P_1+P_6$, then broadcasts P' to its neighbours, that is, N_5 and N_6. N_6 receives and decodes P' to retrieve its expected packet $P_1=P'+P_6$. Similarly, N_5 decodes and retrieves its expected original packet P_6.

3.1.2. COPE header. A COPE header is inserted into the header of a packet, placed between the MAC and IP headers. The COPE header has a structure, as follows. A COPE header includes three blocks, that is, coding report, reception report, and ACK report. Coding report contains information of the XOR-ed native packets and their next hop. Reception report contains information of overlearned packets from neighbours including the source, the received last packet from that source, and a bitmap presenting the list of recently received packets from that source. ACK report contains the information of received or missed packets which the sending node has, including a neighbour IP, the last ACKed packet from that neighbour and a bit map of ACKed packet.

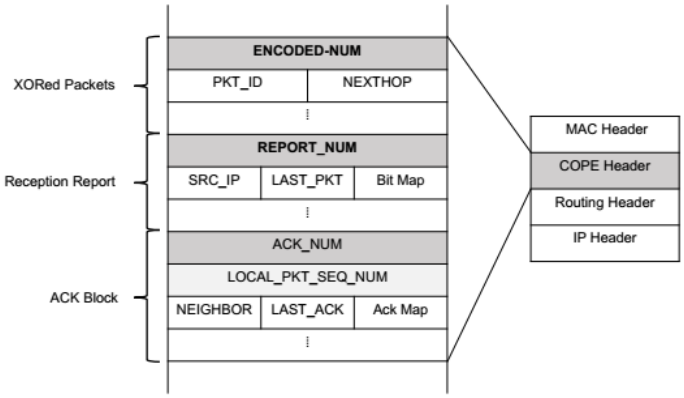


Figure 3. COPE Header Format[1]

3.2. Homomorphic Encryption

Homomorphic encryption is a form of encryption which allows specific types of computations to be carried out on ciphertexts and generate an encrypted result which, when decrypted, matches the result of operations performed on the plaintexts. Homomorphic encryption has three kinds, that is, additive homomorphic encryption, multiplicative encryption, and full homomorphic encryption having both of additive and multiplicative encryptions.

In this paper, the additive homomorphic encryption in [11] is considered as a brilliant candidate for solving the mentioned privacy issues of data coding as it can secure the coded data, as well as, preserve the result of executing the Exclusive-OR bit-wise operator. The property of the Exclusive-OR operator is "A+A=0".Let P_1 and P_2 be two considered messages, k be public key, P be the coded data of P_1 and P_2, C be the cipher text of P encrypted with k, $Dec_k(C)$ or P be the plain text ofC decrypted with k. Let us consider P_1, P_2 and P be three strings of bits. Encryption and decryption of the two messages as follows:

- Encryption:
  $C=Enc_k(P)= Enc_k(P_1+P_2)=Enc_k(P_1)+Enc_k(P_2)$.
- Decryption:
  $P=Dec_k(Enc_k(C))=Dec_k(Enc_k(P_1+P_2))=P_1+P_2$.

In our paper, we adopt Elliptic Curve Cryptography (ECC) [12] based on the binary finite field F^2_m. Assume each node n_i in the network has a pair of keys (k_i, k_i.B) where B is a base parameter of ECC and public to all over the network, k_i is the private key, and k_i.B is the public key. With ECC, the encryption of a message P_1 with k_i is $C_1=Enc_{ki}(P)=(r_i.B, P_i+r_i.B.k_i)$, and the decryption with k_i is $P_i=Dec_{ki}(C_i)=P_i+r_i.B.k_i-(r_i.B).k_i$, where $r_i$is a random number generated by the encryptor.

Hence, the addition of two encryption gets
$C = Enc_{ki}(P_1)+Enc_{ki}(P_2)=(r_1.B+r_2.B, P_1+P_2+r_1.B.k\_i+r_2.B.k_i)$, and the responsive decryption is $Dec_{ki}(C)=P_1+P_2+r_1.B._{ki}+r_2.B.k_i-(r_1.B+r_2.B).k_i=P_1+P_2$.

4. SECURITY MODEL AND REQUIREMENT

4.1. Security Model

Although COPE is simple but effective in reducing the amount of transmissions, COPE still discloses the risks of being attacked. Particularly, COPE header, as presented in Section 3.1.3.,containsmuch sensitive information relating to identity of recipient and sender of the considered packet. In this paper, we consider the honest-but-curious attackers.

Honest-but-curious attack. They correctly operate the protocol without modifying data. However, they try to learn as much personal information of the other users as possible to satisfy their curiosity. They do not use the learned data for any harmful purpose. These adversaries do not cause serious consequences, but their act can leave a back door for the other attacks if they do not preserve well that data. In this work, each node may be considered as an honest-but-curious attacker. They can learn the private information from the COPE header as well as infer the path on which the packet moves through. If they are intersecting nodes, they can try on the received packets. In case they are surrounding nodes of the packet's path, they can try to overhear the packets from that path.

Example 2.The adversary can get the aggregate package, e.g., $P=P_1+P_2+P_3$, at the same time itreceives another aggregate packet, i.e., $P'=P_1+P_2$,so it can infer the content of the packet $P_3= P - P'$.

4.2. Security Requirement

Based on the risk analysis from the above attack models, to avoid serious consequences of attacks, some security requirements need to be guaranteed to be done on the aggregate packet moving through the network.

(1) ***Packet data security.*** The packet payload cannot be read by the intermediate nodes on the network path. It should be accessed by only its destination. To ensure this requirement, a solution should adopt cryptographic algorithms into the problem. The COPE packets should be encrypted with the public key of the destination node, and are aggregated in a secret way. There are two parts to be encrypted, that is, packet payload and fields in the COPE header for coding condition evaluation. This will be presented in further details in Section 6.

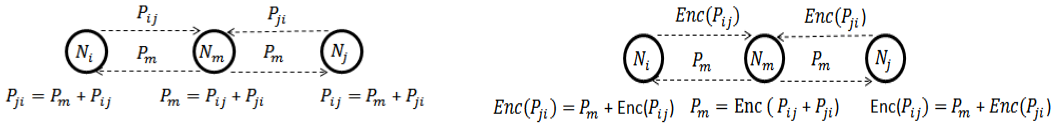
(2) ***Secure coding condition evaluation.*** The intermediate nodes only code the packets they receive if the packets satisfy the coding conditions of theirs. However, the coding condition evaluation process also leaks the packet flow information. To avoid the other party can see the evaluation process at the intermediate node, the security solution is needed. In this work, the cryptographic solutions are proposed to secure this coding condition evaluation process. This will be expressed in more details in Section 6.

(3) ***Performance optimization.*** As the network model is peer-to-peer, the peer devices own plenty of restrictions, that is, limited physical resource and performance. Hence there is a need of requirement, that is, to select a lightweight cryptographic algorithm to secure packets and make the protocol of encrypting packets more securely. In this work, the lightweight cryptography algorithm, i.e., ECC is adopted to reduce the performance cost, at the same time

keep the security complexity of an encryption algorithm. The details will be expressed in Section 6.

5. SCOPE ARCHITECTURE

This section describes a secure COPE architecture, namely SCOPE.SCOPE has three sides, that is, the source, the destination and the intermediate nodes. As in Figure 4, the source is N_i in charge of sending a packet to a certain destination. Whereas, the destination is N_j receiving the packet sent from N_i. Moreover, the middle side is N_m, as considered as an intermediate node, which takes packets and checks the conditions of SCOPE to see if it can aggregate the received packets by using the operator Exclude-OR (see Section6for the aggregate SCOPE conditions), then propagate the aggregate to the next node. For example, as in Figure 4-a, it is assumed that N_i sends a packet P_{ij} to N_j through N_m, and N_j sends a packet P_{ji} to N_i through N_m. N_m invokes an Exclusive-ORoperator (denoted as "+") over the two received packets, and obtains an aggregate value, that is, $P_m = P_{ij} + P_{ji}$.



a - COPE: Data coding without encryption *b - SCOPE:Encrypting coding data.*

Figure 4. COPE Header Format COPE vs SCOPE

In order to make N_m able to successfully compute P_m but also keep the aggregate value in a secret without learning any information from the packets. To achieve this goal, the additive homomorphic encryption is investigated to be used. Let us consider the scenario when the additive homomorphic encryption is applied into data coding mechanism as follows:

- *At source node N_i:*Let us consider Figure 4-b, node N_i tends to send data P_{ij} to N_j through N_m. Before it transmits P_{ij} to N_m, it encrypts P_{ij} by N_j's public key k_j and obtains an encryption $Enc_{Kj}(P_{ij})$. It propagates this encryption to N_m. In the meantime, N_j also wants to send data P_{ji} to N_i through N_m. N_j encrypts P_{ji} by N_i's public key k_i and obtains an encryption $Enc_{ki}(P_{ji})$. Then N_j send this encryption to N_m as well.

- *At intermediate node N_m:* N_m aggregates the two received encryption from N_i and N_j by performing the operator Exclusive-OR on the two cipher text, to receiveC = $Enc_{kj}(P_{ij})$ + $Enc_{ki}(P_{ji})$=$Enc_{(ki,kj)}(P_{ji} + P_{ij})$ where C is the encryption of the coded data of P_{ij} and P_{ji}. Then, N_m transfers C to both N_i and N_j.

- *At destination node N_j:* N_j again adds its $Enc_{kj}(P_{ij})$ to C and obtains $Enc_{kj}(P_{ij})$ + ($Enc_{kj}(P_{ij})$ + $Enc_{ki}(P_{ji})$)=$Enc_{ki}(P_{ji})$. N_j then decrypts $Enc_{ki}(P_{ji})$ with its private key to achieve the data from N_i to it, that is, P_{ji}. The same steps are similarly performed at N_i.

6. SCOPE SECRET PROTOCOL

6.1. Secure Coding Condition

Let us see Figure 4-b,node N_i sends a packet P_{ij} to its target, that is, N_j, through N_m. To make P_{ij} able to reach N_j, the work at N_m is crucial as it decides to forward the packet to N_j. Forwarding the packet is not simply receiving and sending the received packet P_{ij} to the network. It does not also mean that packets move through the intermediate nodes to reach their destination, not all of

intermediates will code (i.e., aggregate) the by-passed packets and propagate the result packets towards. As the objectives of coding protocol, to reduce the bandwidth cost, N_m needs to aggregate (i.e., code) several packets together and forwards only the aggregate packet. In order to support N_m in deciding the P_{ij} propagation, the coding conditions are built up for N_m to be evaluated. The necessary parameters for the coding condition evaluation are retrieved from the header of COPE packet. Only the packets satisfying the coding conditions will be aggregated into their suitable packet then sent towards to their destination. More specifically:

Definition 1. **Coding condition.** *Let F_i and F_j be two flows of packets crossing at node N_m, i.e., Fi ∩ Fj = {N_m}.N_m codes packets received from F_i and F_j in case the next hops of the packet at N_m on flow F_i (or flow F_j) are the previous hops of the packet at node N_m on flow F_j (or flow F_i), or are the neighbours of that previous hop. More formally:*

$$X_j = PH[N_m, F_j]$$
$$X_i = PH[N_m, F_i]$$
$$C(N_m, F_i, F_j) = \begin{cases} true, & ((NH[N_m, F_i] \subset NB_{X_j}) \vee (NH[N_m, F_i] = PH[N_m, F_j])) \\ & \wedge ((NH[N_m, F_j] \subset NB_{X_i}) \vee (NH[N_m, F_j] = PH[N_m, F_i])) \\ false, & otherwise. \end{cases}$$

Where

- *$C(N_m, F_i, F_j)$ is the coding condition.*
- *N_m is the intersecting node of two flows F_i and F_j.*
- *$NH[N_m,F_i]$ is the set of next hops of node N_m in flow F_i.*
- *$PH[N_m,F_i]$ or X_i is the set of previous hops of node N_m in flow F_i.*
- *NB_{Xi} is the set of all neighbours of nodes in X_i in flow F_i.*

Example 4.Let us see Figure 1-b. It is noted that in case of COPE, the set of neighbour nodes and the set of previous node contain only one element. It is assumed that there are two flows of packets, that is, F_1 and F_2 (F_1: $N_1 \rightarrow N_2 \rightarrow N_3$, and F_2: $N_1 \rightarrow N_2 \rightarrow N_3$). N_2 is clearly the intersecting node of the two flows, so it is also the intermediate node where the coding can cause. Hence, the set of previous hops of N_2 on F_1 is $PH[N_2, F_1] = \{N_1\}$, whereas the set of next hops of N_2 on F_1 is $NH[N_2, F_1] = \{N_3\}$.In the meanwhile, the set of previous hops of N_2 on F_2 is $PH[N_2, F_2] = \{N_3\}$, whereas the set of next hops of N_2 on F_2 is $NH[N_2,F_2] = \{N_1\}$. For each node in sets of previous nodes of N_2 in F_1 and F_2, N_1's neighbours is $NB_{N1} = \{N_2\}$, and the set of N_3's neighbours is $NB_{N3} = \{N_2\}$. Let us check the coding conditions in Definition 1,we see that the case $C(N_2, F_1, F_2)$ = true happens.

The above conditions are readable and stored in the header of each node. However, the coding condition evaluation process is done by the intersecting node (i.e., the intermediate node) N_m. Node N_m also needs the information from its surrounding nodes, especially the nodes are the sender and the recipients of the packets through it, as in Figure 4-b, that is, N_i and N_j. Hence, in case that N_m is not an honest and benign node, this assessment process can leak the packet flow information to the intermediate, and cause a serious consequence to the data security and privacy as presented in Section 4.2. This process thus should be done in a secure way. In this work, we adopt the homomorphic encryption as an effective way to secure this process, particularly, ECC is used for securing data. More specifically, all information owned by N_m are encrypted with a public key K_m. All data at N_i and N_j are respectively encrypted with the public keys of N_i and N_j, that is, K_i and K_j. It is noted that the atomic data which is encrypted is the ID of node's neighbours or previous hops or next hops. The encrypted atomic data is specified as in Definition 2.

Definition 2. Atomic Data Cipher*. Let K_2 be the public key of node N_2. Let D_2 is the atomic data to be encrypted with K_2. Therefore, $Enc_{K2}(D_2)$is the encryption (i.e., cipher) of the atomic data D_2 encrypted with the key K_2 of node N_2.*

Example 5.Let us continue Example 4. Let K_m be the public key of N_m, N_i be the previous hop of N_m on flow F_1. Hence, Enc_{Km} (N_i) is the encryption of N_i's ID. It is noted that N_i is also considered as its own ID.

Each of nodes in the network keeps one previous hop list, one next hop list, and one neighbour list. Hence, for evaluating the coding condition, the intersecting node, i.e., N_m, exploits its previous hop list and next hop list, at the same time, requests each of its previous nodes for sending it their neighbour list. These lists are also the input parameters of the coding evaluation process. The coding condition parameters all are lists of atomic data ciphers, and defined in Definition 3.The lists of encryption defined in Definition 4 are also stored in the SCOPE header instead of the plain text as in COPE header.

Definition 4. Coding Condition Parameter Cipher. *Let N_t, $F_i$be respectively the considered node and the consider flow of data. Let $NH[N_t, F_i]$, $PN[N_t, F_i]$, $NB_{PN[Nt,Fi]}$respectively N_t's previous hop list, N_t's next hop list and the neighbour node lists of N_t's previous nodes. From Definition 2 of atomic data cipher, for each element of the lists $NH[N_t, F_i]$, $PN[N_t, F_i]$, $NB_{PN[Nt,Fi]}$, the respective cipher of the lists are denoted as $Enc_{Kt}(NH[N_t, F_i]))$, $Enc_{Kt}(PN[N_t, F_i])$, $ENB_{PN[Nt,Fi]}$ and defined as follows:*

$$
\begin{aligned}
Enc_{K_t}(NH[N_t, F_i]) &= \{Enc_{K_t}(N_{(0,F_i)}), Enc_{K_t}(N_{(1,F_i)}), \cdots, Enc_{K_t}(N_{(n,F_i)})\} \\
Enc_{K_t}(PH[N_t, F_i]) &= \{Enc_{K_t}(N_{(n+1,F_i)}), Enc_{K_t}(N_{(n+2,F_i)}), \cdots, Enc_{K_t}(N_{(n+m,F_i)})\} \\
ENB_{PH[N_t,F_i]} &= Enc_{(K_{(n+1)},K_{(n+2)},\cdots,K_{(n+m)})}(NB_{\{N_{(n+1,F_i)},N_{(n+2,F_i)},\cdots,N_{(n+m,F_i)}\}}) \\
&= \bigcup_{u=n+1}^{n+m} Enc_{K_u}(NB_{N_{(u,F_i)}}) \\
&= \bigcup_{u=n+1}^{n+m} \{Enc_{K_u}(N_{(u,F_i,1)}), Enc_{K_u}(N_{(u,F_i,2)}), \cdots\}
\end{aligned}
$$

Example 7.Let us continue Example 4 and 5. Let N_2, K_2 be respectively the considered node and its own public key. Let F_1 be the considered flow. N_2's lists of next hops, previous hops, and its previous nodes' neighbour node lists on flow F_1as follows:$Enc_{K2}(NH[N_2, F_1]) = \{Enc_{K2}(N_{(3,1)})\}$; $Enc_{K2}(PH[N_2, F_1]) = \{Enc_{K2}(N_1)\}$; $Enc_{K2}(NB_{PH[N2, F1]}) = \{Enc_{K2}(N_1), Enc_{K2}(N_3)\}$ as N_2 on F_1 has two neighbours, that is, N_1 and N_3.

The coding condition evaluation is processed at the intersecting node but it needs a collaboration among multiple parties (i.e., the intersecting node, and its previous hop and next hop on the same flow) to support this evaluation process. For example, as in Figure 4-b, the coding condition evaluation is done by N_m but it needs a collaboration among N_i, N_m and N_j. However, as presented in Section 4.2., to prevent the risk of information violation, this collaboration needs to be processed secretly to avoid leaking anode's private information to the others. In this situation, the information of N_i and N_j must be kept against the reading of N_m. Moreover, the nature of each coding condition evaluation is a comparison among elements of the two lists. Hence, to meet the security requirement in Section 4.2., this comparison is securely processed among encryptions of elements of the lists. For example, as in Definition 1, one of coding condition is the comparison between the lists $NH[N_m, F_i]$ and $PH[N_m, F_j]$, whereas, the comparison between $NH[N_m, F_i]$ and $NB_{PH[Nm, Fi]}$ is a series of comparisons between the list $NH[N_m, F_i]$ and each of lists in $NB_{PH[Nm, Fi]}$ since $NB_{PH[Nm, Fi]}$contains many lists, each of lists contains the neighbour nodes relating to each of

N_m's previous nodes. As in Definition 4, the coding conditions contain the lists of encrypted elements. The comparison operators used for assessing the coding conditions are executed on the lists of encryptions. Thus, let us present one secure comparison between NH[N_m, F_i] and PH[N_m, F_j] done at N_m. The other comparisons in the coding conditions are similarly performed.
Let us consider the two original lists of NH[N_m, F_i] and PH[N_m, F_j] as follows:

$$NH[N_m, F_i] = \{N_{(0,F_i)}, N_{(1,F_i)}, \cdots, N_{(n,F_i)}\}$$
$$PH[N_m, F_j] = \{N_{(n+1,F_j)}, N_{(n+2,F_j)}, \cdots, N_{(n+m,F_j)}\}$$

As in steps in Table 1, first N_i is in charge of generating the encryption of NH[N_m, F_i] using its destination node's public key, that is N_j's public key (i.e., K_j), to obtain $Enc_{K_i}(NH[N_m, F_i]) = \{Enc_{K_i}(N_{(0,F_i)}), Enc_{K_i}(N_{(1,F_i)}), \cdots, Enc_{K_i}(N_{(n,F_i)})\}$ (step 1). In the meanwhile, N_j is in charge of generating the encryption of PH[N_m, F_j] with its destination node's public key, that is Ni's public key (i.e., K_i), to obtain $Enc_{K_j}(PH[N_m, F_i]) = \{Enc_{K_j}(N_{(n+1,F_j)}), Enc_{K_j}(N_{(n+2,F_j)}), \cdots, Enc_{K_j}(N_{(n+m,F_j)})\}$ (step 2). After generating the encryption lists, N_i sends the encryptions to N_m (Step 3), while N_j sends the encryptions to N_m (step 4).Hence, N_m can help transfer the received encryption lists to their destination nodes, that is, N_m sends Enc_{Ki}(NH[N_m, F_i]) to N_j(step 5), and forwards Enc_{Kj}(PH[N_m, F_i]) to N_i (step 6).At N_i, N_i continues to uses the its public key, i.e., K_i, to encrypt the encryption list from Nm, to obtain a twice-encrypted list, that is,

$Enc_{(K_i,K_j)}(PH[N_m, F_i]) = \{Enc_{(K_i,K_j)}(N_{(n+1,F_j)}), Enc_{(K_i,K_j)}(N_{(n+2,F_j)}), \cdots, Enc_{(K_i,K_j)}(N_{(n+m,F_j)})\}$
(step 7). Similarly, $N_j$again encrypts the received list of encryptions with its public key, i.e., K_j, and obtains the twice-encrypted list, that is,
$Enc_{(K_j,K_i)}(NH[N_m, F_i]) = \{Enc_{(K_j,K_i)}(N_{(0,F_i)}), Enc_{(K_j,K_i)}(N_{(1,F_i)}), \cdots, Enc_{(K_j,K_i)}(N_{(n,F_i)})\}$
(step 8). After that, N_i and N_j transfer the twice-encrypted lists to N_m (steps 9, 10).N_m then invokes the function *EqualList ()* as in Algorithm 1 (step 11). *EqualList()*evaluates if two lists are equal to each other. It inputs two lists, that is, $Enc_{(K_i,K_i)}(NH[N_m, F_i])$ and $Enc_{(K_i,K_j)}(PH[N_m, F_i])$, and returns a boolean result, that is, true or false. True is returned as the two encryption lists are equal, and false as the two encryption lists are unequal.

Table 1. Private condition evaluation between two lists NH[N_t, F_i], PH[N_t, F_i].

1	N_i	Creates $Enc_{K_i}(NH[N_m, F_i]) = \{Enc_{K_i}(N_{(0,F_i)}), Enc_{K_i}(N_{(1,F_i)}), \cdots, Enc_{K_i}(N_{(n,F_i)})\}$
2	N_j	Creates $Enc_{K_j}(PH[N_m, F_i]) = \{Enc_{K_j}(N_{(n+1,F_j)}), Enc_{K_j}(N_{(n+2,F_j)}), \cdots, Enc_{K_j}(N_{(n+m,F_j)})\}$
3	$N_i \rightarrow N_m$	N_i sends Enc_{Ki}(NH[N_m, F_i]) to N_m
4	$N_j \rightarrow N_m$	N_j sends Enc_{Kj}(PH[N_m, F_j]) to N_m
5	$N_m \rightarrow N_i$	N_m forwards Enc_{Ki}(NH[N_m, F_i]) to N_j
6	$N_m \rightarrow N_j$	N_m forwards Enc_{Kj}(PH[N_m, F_j]) to N_i
7	N_i	N_i encrypts Enc_{Kj}(PH[N_m, F_j]) and obtains the new encryption list, that is, $Enc_{(K_i,K_j)}(PH[N_m, F_i])$
8	N_j	Nj encrypted Enc_{Ki}(NH[N_m, F_i]) and obtains the result, that is,

		$Enc_{(K_i,K_i)}(NH[N_m, F_i])$
9	$N_i \rightarrow N_m$	Ni transfers the $Enc_{(K_i,K_j)}(PH[N_m, F_i])$ to N_m
10	$N_j \rightarrow N_m$	N_j transfers the $Enc_{(K_i,K_i)}(NH[N_m, F_i])$ to N_m
11	N_m	• Evaluate the equality of two lists $Enc_{(K_i,K_j)}(PH[N_m, F_i])$ and $Enc_{(K_j,K_i)}(NH[N_m, F_i])$ by calling the function *EqualList()* as in Algorithm 1. • If the two lists are totally equal, the conditions is met.

More specifically, in the Algorithm 1, $N_m$traverses the two lists $Enc_{(K_i,K_j)}(PH[N_m, F_i])$ and $Enc_{(K_j,K_i)}(NH[N_m, F_i])$(lines 2,3) to evaluate the equality of elements of two lists by subtracting (or Exclusive-ORing)each element $\overline{x}$of $Enc_{(K_i,K_j)}(PH[N_m, F_i])$ by each element $\overline{y}$of $Enc_{(K_j,K_i)}(NH[N_m, F_i])$ (line 4).Let *count* be a temporary integer. If the subtractive result is equal to an encryption of 0 generated with the public key of N_i and N_j, i.e., $K_i$and K_j, that means $\overline{x}$ is equal to $\overline{y}$, *count* is increased by 1 (line 5). Then, if *count* is equal to the sizes of two lists, that is, sizeNH and sizePH (lines 9, 10, 11), the functions return true (line 12), it means two lists are equal, otherwise false is returned (line 14). In case the two lists are equal, it also indicates that the coding condition is met. Then, the other coding condition can be continued to be evaluated.

Algorithm 1 *EqualList()*

Input:

$Enc_{(K_i,K_j)}(NH[N_m, F_i]) = \{Enc_{(K_i,K_j)}(N_{(0,F_i)}), Enc_{(K_i,K_j)}(N_{(1,F_i)}), \cdots, Enc_{(K_i,K_j)}(N_{(n,F_i)})\}$
$Enc_{(K_i,K_j)}(PH[N_m, F_j]) = \{Enc_{(K_i,K_j)}(N_{(n+1,F_i)}), Enc_{(K_i,K_j)}(N_{(n+2,F_i)}), \cdots, Enc_{(K_i,K_j)}(N_{(n+m,F_i)})\}$

Output: true | false

```
1: count = 0;
2: for x̄ ∈ Enc_(Ki,Kj)(NH[Nm, Fi]) do
3:     for ȳ ∈ Enc_(Ki,Kj)(PH[Nm, Fj]) do
4:         if (x̄ + ȳ == Enc_(Ki,Kj)(0)) then
5:             count++;
6:         end if
7:     end for
8: end for
9: sizeNH = sizeof(Enc_(Ki,Kj)(NH[Nm, Fi]));
10: sizePH = sizeof(Enc_(Ki,Kj)(PH[Nm, Fj]));
11: if (count == sizeNH == sizePH) then
12:     return true;
13: end if
14: return false;
```

6.2. Secure Payload Coding

The fact that COPE header are protected against attacks of observing the flow of packet and intervening the packets' routines is protecting coding conditions and operations on them as presented in Section 6.1.Even though that is a meaningful and important security strategy, securing data payload also play a substantial role since the payload contains several sensitive information of users. Especially, the coding is done at the intersecting node. As discussed in Section 4.2., the intersecting node can be an adversary and the plain data can reveal the personal information to the intersecting node. Hence, the payload should be secured. In this work, ECC algorithm is used to encrypt data into the cipher. This solution makes the intersecting node unable

to read the data but at the same time still work on the encryption only. Hence, the sending node needs to encrypt the data before propagating the encryption to the intersecting node.

Definition 5. Coded Payload Cipher. *Let K_0, K_1,…, K_n be public keys of nodes N_0, N_1, …, N_n. Let $Enc_{K0}(P_0)$, $Enc_{K1}(P_1)$, …, $Enc_{Kn}(P_n)$ be the n payload encryption of packets:$P_0$with K_0, P_1 with K_1,…, $P_n$with K_n, where packetsare sent through the intermediate node N_m. It is assumed that the coding condition as in Definition 1 are met at N_m. The coded payload cipher made at $N_m$is formulated as follows:*

$$\begin{aligned} Enc_{(K_0,K_1,,K_n)}P &= Enc_{(K_0,K_1,,K_n)}\sum_{i=0}^{n}(P_i) \\ &= \sum_{i=0}^{n}(Enc_{K_i}(P_i)) = \sum_{i=0}^{n}(r_i.B, P_i + r_i.B.K_i)) \\ &= (\sum_{i=0}^{n}(r_i.B), \sum_{i=0}^{n}(P_i + r_i.B.K_i)) \end{aligned}$$

where r_0, r_1, …, r_n are random numbers generated at nodes creating the partial encryptions.

Example 8. Let us continue Example 7. It is assumed that N_i wants to send the packet P_{ij} to N_j through N_m on flow F_1, so it encrypts the payload of P_{ij} into $Enc_{Kj}(P_{ij})$ and forwards this encryption to N_m. In the meanwhile, N_j wants to send the packet P_{ji} to N_i through N_m on flow F_2, so it encrypts the payload of P_{ji} into $Enc_{Ki}(P_{ji})$ and forwards this encryption to N_m. At node N_m, after evaluating the coding condition as in Example 4, N_m code the two encryptions $Enc_{Kj}(P_{ij})$ and $Enc_{Ki}(P_{ji})$ by aggregating them, and get $Enc_{(Ki,Kj)}(P_{ij}+P_{ji})=(r_i.B+r_j.B,(P_{ij}+P_{ji})+(r_1.B.K_1+r_2.B.K_2))$ as the coded payload cipher.

Receiving packets from different neighbour nodes, after evaluating the coding condition, $N_m$detaches the encrypted payloads of all packets and code them. Then N_m put them into a new packet. Then, N_m propagates it towards the network. As the receiving node gets the coded packet from N_m, it can assess the coded payload cipher, then decodes and decrypts the cipher to obtain the data for it. This process is concerned as the coded payload assessment. The decoded payload is defined as in Definition 6.

Definition 6. Decoded Payload. *Let N_n be the destination node receiving the encrypted coded payload as defined in Definition 5. Let K_0, K_1,…, K_{n-1} be public keys of N_n's neighbour nodes N_0, N_1, …, N_{n-1}. Let* $Enc_{(K_0,K_1,,K_n)}\sum_{i=0}^{n}(P_i) = (\sum_{i=0}^{n}(r_i.B), \sum_{i=0}^{n}(P_i + r_i.B.K_i))$ *be the coded payload cipher of packets from N_0, N_1, …, N_n with the random number r_0, r_1, …, r_n generated by nodes generating the partial encryptions.More formally, the decoded payload by N_n, that is $Enc_{Kn}(P_n)$, is defined as follows:*

$$\begin{aligned} Enc_{K_n}(P_n) &= Enc_{(K_0,K_1,,K_n)}\sum_{i=0}^{n}(P_i)) + Enc_{(K_0,K_1,,K_{(n-1)})}\sum_{i=0}^{(n-1)}(P_i)) \\ &= (\sum_{i=0}^{n}(r_i.B), \sum_{i=0}^{n}(P_i + r_i.B.K_i)) + (\sum_{i=0}^{(n-1)}(r_i.B), \sum_{i=0}^{(n-1)}(P_i + r_i.B.K_i)) \\ &= ((\sum_{i=0}^{n}(r_i.B) + (\sum_{i=0}^{(n-1)}(r_i.B)), (\sum_{i=0}^{n}(P_i + r_i.B.K_i) + \sum_{i=0}^{(n-1)}(P_i + r_i.B.K_i))) \\ &= (r_n.B, P_n + r_n.B.K_n) \end{aligned}$$

$$Dec_{K_n}(Enc_{K_n}(P_n)) = (r_n.B).K_n + P_n + r_n.B.K_n = P_n$$

Example 9. Let us continue Example 8. N_j receives the coded payload cipher for it, that is, $Enc_{(Ki,Kj)}(P_{ij} + P_{ji}) = (r_i.B + r_j.B, (P_{ij} + P_{ji}) + (r_i.B.K_i + r_j.B.K_j))$. N_j still keeps the coded payload cipher of the packet it wants to send to Ni, that is, $Enc_{Ki}(P_{ji}) = (r_i.B, P_{ji} + r_i.B.K_i)$ where r_i is a random number generated by N_j. Hence, the decoded payload is $Enc_{Kj}(P_{ij}) = Enc_{(Ki,Kj)}(P_{ij} + P_{ji}) + Enc_{Ki}(P_{ji}) = (r_i.B + r_j.B, (P_{ij} + P_{ji}) + (r_i.B.K_i + r_j.B.K_j)) + (r_i.B, P_{ji} + r_i.B.K_i) = ((r_i.B + r_j.B + r_i.B), ((P_{ij} + P_{ji} + (r_i.B.K_i + r_j.B.K_j + P_{ji} + r_i.B.K_i)) = (r_j.B, P_{ij} + r_j.B.K_j)$. Then, the decryption is executed at N_j with the private key of N_j (i.e., K_j), to get the data needed for N_j, i.e., $Dec_{Kj}(Enc_{Kj}(P_{ij})) = (r_j.B).K_j + P_{ij} + r_j.B.K_j = P_{ij}$.

7. SCOPE EVALUATION

In this work, to prove for effectiveness and efficiency of the proposed secure protocol, experiments on different number of nodes and different key sizes of ECC encryption are done. These experiments are operated on PC with the physical resources in terms of CPU 1.8GHz Intel Duo-Core, RAM 4GB, HDD 16GB.

7.1. Throughput

We use NS-2 as a simulator for the experiment. We use *4* topologies in Figure 1, Figure 6a, Figure 6b, and Figure 6c. Flows in test scenarios are shown in the Table 2. Each topology has been generated in a flat area *1000m X 1000m*. The data traffic in the network are all CBR (Constant Bit Rate) flows sent over UDP (User Datagram Protocol) using *1000*-byte datagrams with an arrival interval of *0.01s* and traffic generation duration at source of *150s*. The routing protocol used is DSDV (Destination-Sequenced Distance-Vector) [10]. The results are collected with a confident interval of *95%*.

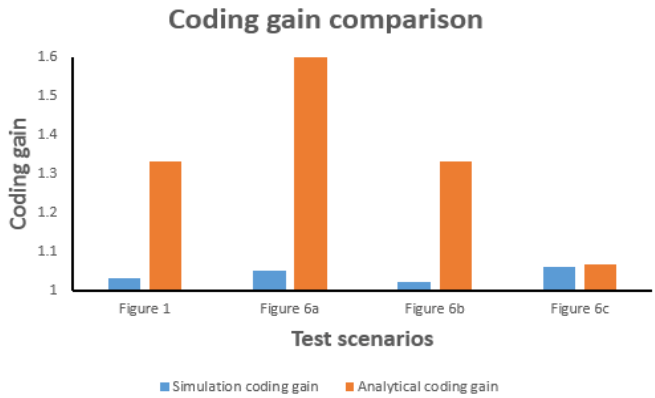


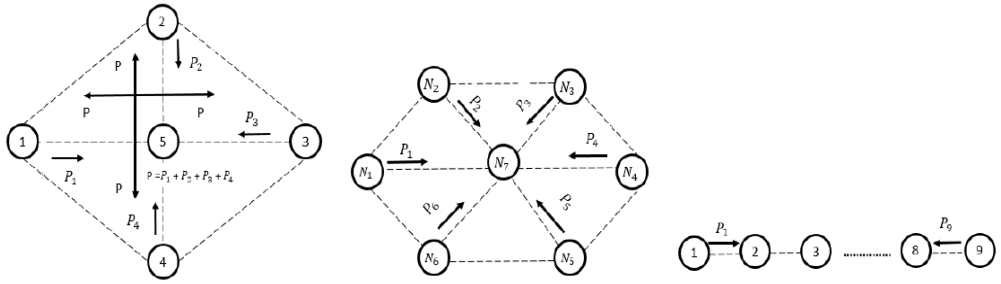
Figure 5.Comparison between analytical coding gain and simulation coding gain.

Results are presented in Figure 5, which compares between the coding gains obtained by simulation and the ones obtained by theoretical analysis [5]. We observe that the coding gains obtained by theoretical analysis are always greater than what obtained by simulation. For instance, the test case of Figure 1, the simulated coding gain is equal to *1.033* while the theoretical coding gain is $\frac{4}{3}$= *1.333*. For the test scenario of Figure 6a, the simulated and analytical coding gains are *1.050* and $\frac{8}{5}$= *1.600*, respectively. For the test case of Figure 6b, the coding gain is *1.020* for the simulation while the coding gain is $\frac{12}{9}$= *1.333* for the theoretical analysis. For the test scenario of Figure 6c, the simulated coding gain and the theoretical coding gain equal *1.060* and $\frac{16}{15}$= *1.067*, respectively. These deviations are because the theoretical analysis cannot take into account the collision in wireless network environment. The collision can lead to the delay increase that a packet sent from source node to destination node and so, some coding

opportunities are missed. Besides that, frame loss, frame retransmission, or framing error are also one of the problems derived from the collision, affecting to the coding gain results.

7.2. Time overhead

In this experiment, to assess the computing performance of SCOPE. We make a diversity of experimentson different parameters. Particularly, we create fourSCOPE scenarios (as in Figures 1, 6a, 6b, 6c), and change the key sizes of ECC additive homomorphic encryption algorithm. Each calculated value in the experimentsis the average of 20 times running the same experiments.



(a) COPE with 5 nodes, packets XOR-ed at node 5 (b) COPE with 7 nodes, packets XOR-ed at node 7 (c) COPE with 9 nodes, packets XOR-ed at node 2, 3, 4, 5, 6, 7 or 8.

Figure 6.SCOPE scenarios.

Table 2. Flows in test scenarios.

Scenario	Figure	Flows
1	1	F1: N1 → N2 → N3; F2: N3 → N2 → N1
2	6a	F1: N1 → N5 → N3; F2: N3 → N5 → N1; F3: N2 → N5 → N4; F4: N4 → N5 → N2
3	6b	F1: N1 → N7 → N4; F2: N4 → N7 → N1; F3: N2 → N7 → N5; F4: N5 → N7 → N2; F5: N3 → N7 → N6; F6: N6 → N7 → N3
4	6c	F1: N1 → N2 → N3 → ... → N9; F2: N9 → N8 → N7 → ... → N1

Figures 1 and 6 describes four scenarios, and table 2 presents the number of flows w.r.t. the scenarios. Figure 1 involves 3 nodes and 2 flows. Figure 6a involves 5 nodes and 4 flows. Figure 6b involves 7 nodes and 6 flows. Figure 6c involves 9 nodes and 2 flows. To evaluate the computing performance of SCOPE applied with ECC encryption algorithms, the selected ECC key sizes are varied in {163, 283, 409, 571} bits. These key sizes are guaranteed to be still secure by NIST.

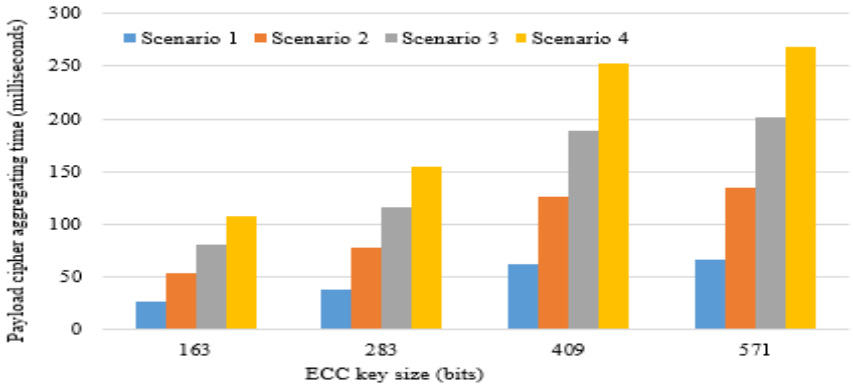


Figure 7. Time on aggregating the payload cipher at the intersecting node (milliseconds or ms) vs ECC key size (bits)

First, the time on aggregating the payload ciphers in the four scenarios. These payload ciphers are aggregated using the additive property of ECC homomorphic encryption. The number of payload cipher aggregation at the intersecting node in the four scenarios are respectively 2, 4, 6, 8 for each flow. In Figure 7, with the smallest ECC key size (i.e., 163 bits), the time on aggregating two payload ciphers of Scenario 1 is 26,8ms. With the 283-bit key size, the time on aggregating 8 payload ciphers of Scenario 4 is 107,2ms. In the worst case, that is, the largest key size (i.e., 571 bit), the time computed for aggregating 8 payload ciphers of Scenario 4 is 268,8ms. The times on different scenarios and key sizes are reasonable in the peer-to-peer environment.

Figure 8 is another experiment to compute the time cost for SCOPE transmissions. These time are measured to evaluate the time which a packet moves through a flow from the source node to the destination node. Hence, these times include the payload cipher aggregating times (as in the experiment of Figure 7) and transmission times. In this experiment, the number of payload cipher aggregations at the intersecting nodes are similar to the previous experiment. The ECC key sizes are also varied in {163, 283, 409, 571} bits. In the case of smallest ECC key size, that is, 163 bits, in scenario 1 involving 2 payload cipher aggregations, the time cost is 260,8ms. Whereas, in the case of largest ECC key size (i.e., 571 bits) and Scenario 4 involving 8 payload cipher aggregations, the time costs is 2.5s. The time overheads in this experiment in both cases are reasonable and prove that SCOPE is effective and efficient.

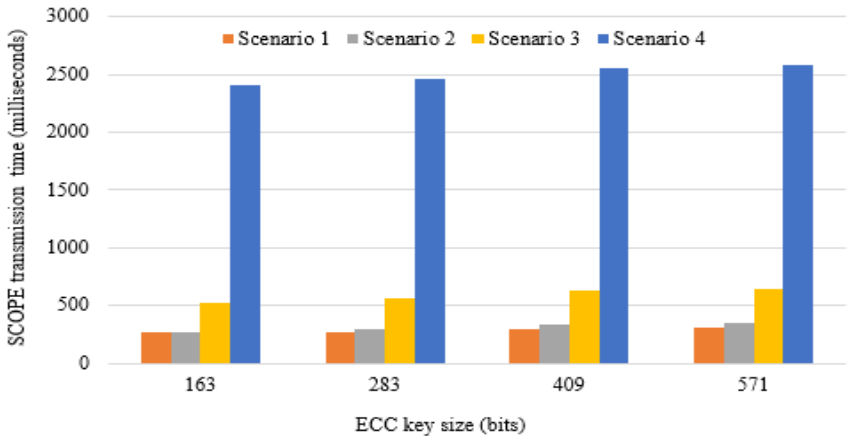


Figure 8. Time on SCOPE transmission including aggregation time (milliseconds or ms) vs ECC key size (bits)

In order to evaluate the cryptographic process of evaluating the coding conditions of SCOPE, we measure the time SCOPE spent on this evaluation (Figure 9). The key sizes are selected in the range {163, 283, 409, 571} bits. The number of coding conditions for Scenarios 1, 2, 3, and 4 are respectively 4, 20, 30, 32 conditions. The time on evaluating the coding conditions in Scenario 1 (i.e., with the lowest number of conditions, that is, 4) with the smallest key size (i.e., 163 bits) is 6.8ms. In the meanwhile, the time on coding conditions evaluation in Scenario 4 (i.e., the highest number of condition, that is, 32) with the largest ECC key size (i.e., 571 bits) is 115,2 ms. In the both cases of the smallest parameters and the largest parameters, the time overheads are still reasonable and prove the effectiveness and efficiency of SCOPE.

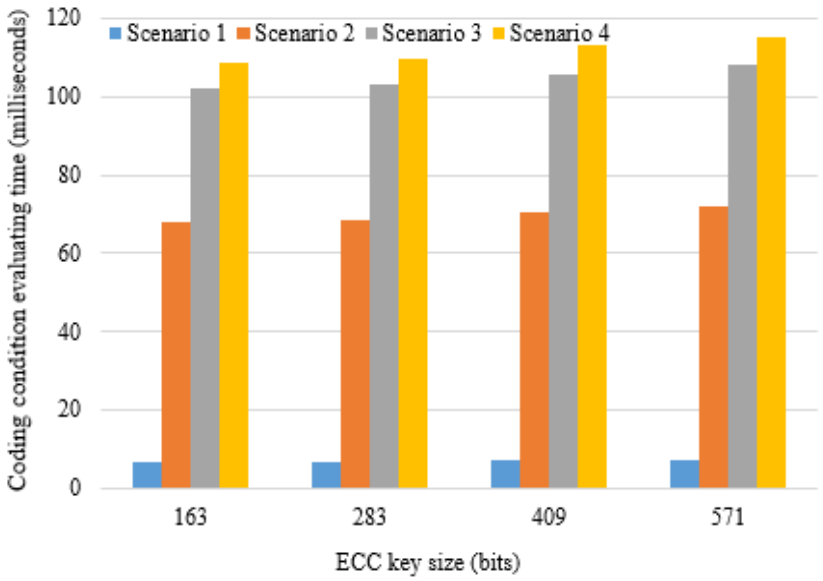


Figure 9. Time on evaluating coding condition at intersecting node (milliseconds or ms) vs ECC key size (bits)

8. SECURITY PROPERTY

In this section, a security proof is presented. More specifically, the expression how the proposal can cope with risks of honest-but-curious attack as in Section 4.1., as well as how the proposal meets the security requirements as in Section 4.2. As introduced in Section 4.1, this attack does not aim to poison or misuse the data for any dangerous purposes, but the adversaries try to learn or infer as much personal information as possible only to satisfy their curiosity. However, this attack possibly gets more dangerous as its consequence can leave a backdoor for another attack.

Packet data security. To avoid this risk, ECC homomorphic encryption is adopted to cipher the content of the data payload of packets. The payload then gets into a secret writing, i.e., unreadable. As a result, this carries the COPE packets a shield to deteradversaries from inferring any information inside the payload. Particularly, the ECC public key used for encrypting the payload is kept by only the destination node of the packet, so the other nodes cannot read the packet anyway. Only the destination node of the packet has the private key which can be used for decrypting the payload, then the payload can be read.

Information inferring protection. In this work, the addition property of the homomorphic encryption ECC is exploited for coding multiple packets into one packet at the intermediate node before the aggregate packet is propagated to the next hop of the intermediate node. More

specifically, the public key of the destination node of the packet is used for this secret aggregation phase. The intermediate nodes just follow up with the protocol steps, and aggregates the encrypted packets without being aware of the packets' content. This point helps the data safe from the intermediate node. They cannot read the data inside and cannot infer the information as they do not have the private key of the destination node.

Coding condition evaluation security. The coding conditions involve encryptions of the node IDs as presented in Section 6.1. The comparisons are executed on these encryptions. Thus the intermediate nodes cannot learn the node IDs inside the thresholds and operands. Additionally, we also protect the neighbour nodes by encrypting their IDs, and only their direct neighbours can know their IDs, but the two-hop nodes cannot know their IDs. The comparative results are also not recognized by the nodes. Hence, the coding conditions are secured during the evaluation phase.

Performance optimization. In this work, we empower our proposal's security with the ECC encryption algorithm. The ECC encryption algorithm is invoked based on the binary field with the binary operators. This reduces much the time consumption, and meets the computing performance requirements. Additionally, the ECC is still guaranteed to be secure by NIST. So, the proposal can ensure the computing performance to be optimized.

9. CONCLUSION

In this work, we propose a cryptographic approach, namely SCOPE, which is able to support nodes secretly executing the COPE protocol, by applying the lightweight homomorphic encryption ECC. Hence, the packets in SCOPE can move through the intersecting nodes without leaking any private information of the packets. Moreover, SCOPE can be also against the honest-but-curious attack at the intersecting nodes. SCOPE can keep all operations in evaluating the coding conditions or in aggregating the payload work securely. The proposal is also proved to be effective and efficient through the different experiments on a variety of ECC key sizes and different scenarios. This work will be improved to fit with more network coding protocols, such as, BEND, DCAR, etc. and to be immune to the malicious attack in the future work.

REFERENCES

[1] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Mdard, and J. Crowcroft, XORs in the air: Practical Wireless Network Coding, Proc. ACM SIGCOMM, pp. 243-254, 2006.

[2] C. V. Phung, Q. T. Minh, and M. Toulouse, Routing Optimization Model in Multihop Wireless Access Networks for Disaster Recovery, International Conference on Advanced Computing and Applications (ACOMP2016), Can Tho, Vietnam, Nov. 23-25, 2016.

[3] C. V. Phung, V. H. Nguyen, and T. M. T. Nguyen, BEND-Aware Routing Optimization in Wireless Mesh Networks, IEEE International Conference on Advanced Technologies for Communications (ATC), Ho Chi Minh, Vietnam, 2015.

[4] C. V. Phung, T. V. Vu, and T. M. T. Nguyen, DCAR Coding Gain Modeling and Analysis, NoF 2013 - Fourth International Conference on the Network of the Future, Pohang, South Korea, pp. 1-3, (IEEE) (2013).

[5] Sudipta Sengupta, Shravan Rayanchu, and Suman Banerjee, Network Coding-Aware Routing in Wireless Networks, IEEE/ACM Transactions on Networking, vol. 18, no. 4, Agust 2010.

[6] J. Zhang, Y.B. Chen, and I. Marsic, MAC-layer proactive mixing for network coding in multi-hop wireless networks, Computer Networks, Elsevier, vol. 54, pp. 196-207, 2010.

[7] J. Le, J.C.S. Lui, and D.M. Chiu, DCAR: Distributed Coding-Aware Routing in Wireless Networks, IEEE Transactions on Mobile Computing, vol. 9, no. 4, pp. 596-608, April 2010.

[8] T.V Vu, T.M.T. Nguyen, and G. Pujolle, Distributed Opportunistic and Diffused Coding in Multi-hop Wireless Networks, IEEE ICC workshop on Cooperative and Cognitive Mobile Networks (COCONET), Ottawa, Canada, June 2012.

[9] T.V Vu, T.M.T. Nguyen, and G. Pujolle, Distributed Opportunistic and Diffused Coding with Multiple Decoders in Wireless Mesh Networks, ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, Cyprus (2012).

[10] C.E. Perkins, and P. Bhagwat, Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers, SIGCOMM, London, England UK, 1994.

[11] X. Yi, R. Paulet, and E. Bertino, Homomorphic Encryption and Applications, Springer Publishing Company, Incorporated, 2014.

[12] V. Katiyar, K. Dutta and S. Gupta, A Survey on Elliptic Curve Cryptography for Pervasive Computing Environment, International Journal of Computer Applications, vol. 11, no. 10, pp. 41–46, December 2010.

[13] T. V. Vu, T. M. T. Nguyen, G. Pujolle and N. Boukhatem, DODEX+: A new network coding scheme for mesh networks in mobile cloud computing, Wireless Days (WD), Rio de Janeiro, Brazil, Nov. 12-14 , 2014.

[14] N. Cai, R.W. Yeung, Network coding and error correction, in: Proceedings of the 2002 IEEE Information Theory Workshop, 2002, 2002, pp. 119–122.

[15] Talooki, V.N., Bassoli, R., Lucani, D.E., Rodriguez, J., Fitzek, F.H., Marques, H. and Tafazolli, R., 2015. Security concerns and countermeasures in network coding based communication systems: A survey. Computer Networks, 83, pp.422-445.

[16] N. Cai, R.W. Yeung, Secure network coding, in: Proceedings. 2002 IEEE International Symposium on Information Theory, 2002, 2002, p. 323.

[17] Fan, Y., Jiang, Y., Zhu, H., Chen, J. and Shen, X.S., 2011. Network coding based privacy preservation against traffic analysis in multi-hop wireless networks. IEEE Transactions on Wireless Communications, 10(3), pp.834-843.

[18] Esfahani, A., Mantas, G., Monteiro, V., Ramantasy, K., Datsikay, E. and Rodriguez, J., 2015, September. Analysis of a Homomorphic MAC-based scheme against tag pollution in RLNC-enabled wireless networks. In Computer Aided Modelling and Design of Communication Links and Networks (CAMAD), 2015 IEEE 20th International Workshop on (pp. 156-160). IEEE.

[19] Li, C., Chen, L., Lu, R. and Li, H., 2015. Comment on "An Efficient Homomorphic MAC with Small Key Size for Authentication in Network Coding". IEEE Transactions on Computers, 64(3), pp.882-883.

[20] Li, X., Fu, F.W., Zhao, X. and Wang, G., 2015, August. Two improved homomorphic MAC schemes in network coding. In Fuzzy Systems and Knowledge Discovery (FSKD), 2015 12th International Conference on (pp. 2214-2219). IEEE.

[21] Esfahani, A., Mantas, G., Rodriguez, J., Nascimento, A. and Neves, J.C., 2015, June. A null space-based MAC scheme against pollution attacks to Random linear Network Coding. In Communication Workshop (ICCW), 2015 IEEE International Conference on (pp. 1521-1526). IEEE.

[22] Paillier, P., 1999, May. Public-key cryptosystems based on composite degree residuosity classes. In Eurocrypt (Vol. 99, pp. 223-238)

AUTHORS

Ngoc Hong Tran has been a full-time lecturer at the Vietnamese German University, Vietnam, since 2016. She obtained a Bachelor Degree in Information Technology, and a Master Degree in Computer Science, from University of Science, Vietnam National University - Ho Chi Minh City. She was awarded a PhD Diploma in Computer Science from University of Insubria, Italy. Moreover, she had been a full-time lecturer at the University of Science, VNU-HCM, from 2004 to 2016. She also worked in National Institute of Informatics, Tokyo, Japan, in 2009, and was an exchange scholar in Portland State University (PSU), Portland City, Oregon State, USA, during the Fall term in 2010. She worked in LAAS-CNRS, Toulouse, France, in 2012, and was an invited researcher in Singapore University of Technology and Design, Singapore, from March to April 2017.Her research interests are applied-cryptography, secure collaboration among multiple parties in distributed and centralized networks, mobile/MANET social networks, web service composition, network coding; privacy preserving data mining.

Cao Vien Phung is currently a Ph.D student in Information System Technology at Braunschweig University of Technology, Germany. He was born in Phu Yen, Vietnam. He attended Luong Van Chanh high school for the gifted in Chemistry, Tuy Hoa - Phu Yen, Vietnam in 2007. He received the Engineer Degree in Electronics and Telecommunications from the Posts and Telecommunications Institute of Technology (PTIT), Ho Chi Minh, Vietnam in 2012. He obtained the Master Degree in Computer Science from Paris 6 University, Paris, France in 2015.

Binh Quoc Nguyen is currently working as Research and Teaching Assistance of Computer Science study program of Vietnam German University. He attended Le Hong Phong high school for the gifted in 2003. He obtained the Bachelor Degree in Information Technology in 2009 and Master Degree in 2012 at University of Science, VNU of Ho Chi Minh City.

Leila Bahri is currently a postdoc in KTH, Sweden. She obtained her bachelor's degree in Computer Science from the School of Science and Engineering at Al Akhawayn University in Ifarne - AUI, Morocco (with highest honor / Summa Cum Laude). In 2011 she received her master's degree in software engineering from AUI with a thesis on the implementation and adoption of ITIL in an academic IT department. In 2009, she received a scholarship for short term studies in Japan from the Japan Student Services Organization (JASSO) by which she performed research for one year at a Robotics laboratory in Meijo University, Nagoya. In 2010, she received the Google University Excellence Award for being selected as best computer science student at her university for that year. Right after that she fulfilled the position of IT Service Desk Manager at the IT department of AUI until May, 2013. She was awarded the Ph.D diploma from University of Insubria, Italy in 2016. She worked as a posdoc in Koc University, Turkey, in 2016.

Dung Hai Dinh has workedat the Vietnamese German University since 2015, and is now currently a full-time Lecturer cum Academic Coordinator of the M.Sc. study program Business Information Systems. He began his study and academic career in Germany since 2004 and got his degree at University of Applied Sciences Gelsenkirchen, major in Business Engineering. He holds a Ph.D. Degree granted by Saarland University, Saarbrücken, Germany. Since 2009, he worked as a researcher at Chair of Operations Research and Business Informatics, with focus on quantitative methods and optimization algorithms for business decision-making problems.

AUTHOR INDEX