Lecture Notes in Electrical Engineering 513

Kuinam J. Kim · Hyuncheol Kim
Editors

Mobile and Wireless Technology 2018

International Conference on Mobile and Wireless Technology (ICMWT 2018)

Springer

Lecture Notes in Electrical Engineering

Volume 513

**** Indexing: The books of this series are submitted to ISI Proceedings, EI-Compendex, SCOPUS, MetaPress, Springerlink ****

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

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

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

For general information about this series, comments or suggestions, please use the contact address under "service for this series".

To submit a proposal or request further information, please contact the appropriate Springer Publishing Editors:

Asia:

China, *Jasmine Dou, Associate Editor* (jasmine.dou@springer.com) (Electrical Engineering)

India, *Swati Meherishi, Senior Editor* (swati.meherishi@springer.com) (Engineering)

Japan, *Takeyuki Yonezawa, Editorial Director* (takeyuki.yonezawa@springer.com)
(Physical Sciences & Engineering)

South Korea, *Smith (Ahram) Chae, Associate Editor* (smith.chae@springer.com)
(Physical Sciences & Engineering)

Southeast Asia, *Ramesh Premnath, Editor* (ramesh.premnath@springer.com)
(Electrical Engineering)

South Asia, *Aninda Bose, Editor* (aninda.bose@springer.com) (Electrical Engineering)

Europe:

Leontina Di Cecco, Editor (Leontina.dicecco@springer.com)
(Applied Sciences and Engineering; Bio-Inspired Robotics, Medical Robotics, Bioengineering; Computational Methods & Models in Science, Medicine and Technology; Soft Computing; Philosophy of Modern Science and Technologies; Mechanical Engineering; Ocean and Naval Engineering; Water Management & Technology)

Christoph Baumann (christoph.baumann@springer.com)

(Heat and Mass Transfer, Signal Processing and Telecommunications, and Solid and Fluid Mechanics, and Engineering Materials)

North America:

Michael Luby, Editor (michael.luby@springer.com) (Mechanics; Materials)

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

Kuinam J. Kim · Hyuncheol Kim
Editors

Mobile and Wireless Technology 2018

International Conference on Mobile and Wireless Technology (ICMWT 2018)

Springer

Editors
Kuinam J. Kim
iCatse
Suwon, Kyonggi-do
Korea (Republic of)

Hyuncheol Kim
Computer Science
Namseoul University
Cheonan, Ch'ungch'ŏng-namdo
Korea (Republic of)

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-981-13-1058-4 ISBN 978-981-13-1059-1 (eBook)
https://doi.org/10.1007/978-981-13-1059-1

Library of Congress Control Number: 2018945082

Printed on acid-free paper

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

This LNEE volume contains the papers presented at the iCatse International Conference on Mobile and Wireless Technology, and iTAIWAN workshop, which were held in Hong Kong, China, during June 25–27, 2018.

The conferences will provide an excellent international conference for sharing knowledge and results in Mobile and Wireless Technology. The aim of the conferences is to provide a platform to the researchers and practitioners from both academia as well as industry to meet the share cutting-edge development in the field.

The primary goal of the conference is to exchange, share, and distribute the latest research and theories from our international community. The conference will be held every year to make it an ideal platform for people to share views and experiences in related fields.

On behalf of the Organizing Committee, we would like to thank Springer for publishing the proceedings of the conferences. We also would like to express our gratitude to the 'Program Committee and Reviewers' for providing extra help in the review process. The quality of a refereed volume depends mainly on the expertise and dedication of the reviewers. We are indebted to the Program Committee members for their guidance and coordination in organizing the review process, and to the authors for contributing their research results to the conference.

Our sincere thanks to the Institute of Creative Advanced Technology, Engineering and Science for designing the conference Web page and spending countless days in preparing the final program in time for printing. We would also like to thank our organizing committee for their hard work in sorting our manuscripts from our authors.

We look forward to seeing all of you next year's conference.

Suwon, Korea (Republic of) — Kuinam J. Kim
Cheonan, Korea (Republic of) — Hyuncheol Kim

ICMWT 2018: Organizing Committee

General Co-chairs

Hyeuncheol Kim, NamSeoul University, Republic of Korea
Sangmin Park, Inchon National University, Republic of Korea
Xiaoxia Huang, University of Science and Technology Beijing, China
Nikolai Joukov, New York University and modelizeIT Inc, USA

Steering Committee

Nikolai Joukov, New York University and modelizeIT Inc, USA
Borko Furht, Florida Atlantic University, USA
Bezalel Gavish, Southern Methodist University, USA
Kin Fun Li, University of Victoria, Canada
Xiaoxia Huang, University of Science and Technology Beijing, China
Naruemon Wattanapongsakorn, King Mongkut's University of Technology Thonburi, Thailand

Publicity Chair

Dmitri A. Gusev, Purdue University, USA
Chan Shiau Wei, Universiti Tun Hussein Onn Malaysia, Malaysia
Azizi bin Miskon, National Defence University of Malaysia, Malaysia
Naruemon Wattanapongsakorn, King Mongkut's University of Technology Thonburi, Thailand
Suresh Thanakodi, National Defence University of Malaysia, Malaysia
Hongseok Jeon, ETRI, Republic of Korea
Tomas Cerny, Czech Technical University, Czech Republic
Dan (Dong-Seong) Kim, University of Canterbury, New Zealand

Financial Chair

Hara Paul Kim, Institute of Creative Advanced Technologies, Science and Engineering, Republic of Korea

Publication Chair

Nakhoon Baek, Kyungpook National University, Republic of Korea

Program Chair

Kuinam J. Kim, Institute of Creative Advanced Technologies, Science and Engineering, Republic of Korea

Organizers and Supporters

Institute of Creative Advanced Technologies, Science and Engineering (iCatse)
Korean Industry Security Forum (KISF)
Korea Information Assurance Society (KIAS)
National Defence University of Malaysia, Malaysia
River Publishers, The Netherlands
Korea Institute of Science and Technology Information (KISTI)
Electronics and Telecommunications Research Institute (ETRI)
Kyonggi University
Nam Seoul University
Software Networking Journal
Chinese Management Science Society (CMSS)

Program Committee

Azeddien M. Sllame, University of Tripoli, Libya
Michel Mainguenaud, Institut National des Sciences Appliquées—Rouen, France
Jeril Kuriakose, Manipal University Jaipur, India
Wen-Jyi Hwang, National Taiwan Normal University, Taiwan
Yadigar Imamverdiyev, Institute of Information Technology of Azerbaijan National Academy of Sciences, Azerbaijan
Jay Kishigami, Muroran Institute of Technology, Japan
Wing Kwong, Hofstra University, USA
Avinash Malik, University of Auckland, New Zealand
Fernando Moreira, Universidade Portucalense, Portugal
Moises Diaz, Universidad del Atlántico Medio, Spain

Mohd Faizal Abdollah, University Teknikal Malaysia Melaka, Malaysia
Paolo Crippa, Università Politecnica delle Marche, Italy
Pedro Nuno Miranda de Sousa, University of Minho/Department of Informatics, Portugal
Paulo Pinto, Universidade Nova de Lisboa, Portugal
Jitender Grover, IIIT-Hyderabad, India
Nasimuddin, Institute for Infocomm research, A-STAR, Singapore
Lin-huang Chang, National Taichung University of Education, Taiwan
Mohd Ashraf Bin Ahmad, Universiti Malaysia Pahang, Malaysia
Mehmet Celenk, Ohio University, USA
Reggie Davidrajuh, University of Stavanger, Norway
Sabin-Corneliu Buraga, Alexandru Ioan Cuza University of Iasi, Romania
Cláudio Alves, Universidade do Minho, Portugal
Marco Aiello, University of Stuttgart, Germany
Tantikorn Pichpibul, Panyapiwat Institute of Management, Thailand
Suprakash Gupta, Indian Institute of Technology (B.H.U), India
Siti Zulaiha Ahmad, UiTM Perlis Branch, Malaysia
Rui Pedro Marques, University of Aveiro, Portugal
Suksan Prombanpong, King Mongkut's University of Technology Thonburi, Thailand
Ankur Gupta, Model Institute of Engineering and Technology, Jammu, India
Galina G. Artyushina, Moscow Aviation Institute (National Research University), Russia
Yiliu Liu, Norwegian University of Science and Technology, Norway
Rika Ampuh Hadiguna, Andalas University, Indonesia
Muhammad Ibrahim Abdullah, COMSATS Institute of Information Technology, Lahore, Pakistan
Kittisak Jermsittiparsert, Political Science Association of Kasetsart University, Thailand
Ramayah Thurasamy, Universiti Sains Malaysia, Malaysia
Marco Anisetti, Università degli Studi di Milano, Italy
Virgilio Cruz Machado, Universidade NOVA de Lisboa, Portugal
Professor Seifedine Kadry, Beirut Arab University, Lebanon
Hardeep Singh, Ferozepur College of Engineering and Technology, India
Somlak Wannarumon Kielarova, Naresuan University, Thailand
Sangmin Park, Inchon National University, Republic of Korea
Alejandro Escudero-Santana, Universidad de Sevilla, Spain
Rafif Al-Sayed, The International Institute of Knowledge and Innovation, USA
Vu Truong Son Dao, International University—Vietnam National University HCMC, Vietnam
Paweł Błaszczyk, University of Silesia in Katowice, Poland
Md. Mamun Habib, BRAC University, Bangladesh
Sidi Wu, Waseda University, Japan
S. Sarifah Radiah Shariff, Malaysia Institute of Transport (MIITRANS), Universiti Teknologi MARA, Shah Alam, Malaysia

Virgínia Helena Arimateia de Campos Machado, Universidade Nova de Lisboa, Portugal

iTAIWAN Workshop Program Committee Members

Workshop Chairs

Lin-huang Chang, National Taichung University, Taiwan

Workshop Co-chairs

Jiun-Jian Liaw, Chaoyang University, Taiwan,
Hung-Chi Chu, Chaoyang University, Taiwan
T. H. Lee, National Taichung University, Taiwan

Program Committee

Lin-huang Chang, National Taichung University, Taiwan
Hung-Chi Chu, Chaoyang University, Taiwan
Jenq-Neng Hwang, University of Washington, USA
T. H. Lee, National Taichung University, Taiwan
Victor C. M. Leung, University of British Columbia, Canada
Jiun-Jian Liaw, Chaoyang University, Taiwan
huan-Pin Lu, Meiho University, Taiwan
Frode Eika Sandnes, Oslo University College, Norway
S. C. Wang, Chaoyang University, Taiwan
K. M. Yap, Sunway University, Malaysia

Organizers and Supporters

Institute of Creative Advanced Technologies, Science and Engineering (iCatse)
Chinese Management Science Society (CMSS)
Korean Industry Security Forum (KISF)
Korea Information Assurance Society (KIAS)
Kyonggi University, Republic of Korea
University of Science and Technology Beijing, China
King Mongkut's University of Technology Thonburi, Thailand
River Publishers, The Netherlands
Software Engineering Journal

Contents

Part I
Mobile and Wireless Technology

Design and Evaluation of a Bloom Filter Based Hierarchical Hybrid Mobility Management Scheme for Internet of Things

Ihn-Han Bae

Abstract Information-centric networking (ICN) has emerged as a promising candidate for the architecture of the future Internet of Things (IoT). ICN provides unique and location-independent content names, in-network caching and name-based routing. Consequently, locating the demanded device or content is one of the major challenges in ICN-based IoT networks. In this paper, we propose a BFH^2M^2: Bloom Filter based Hierarchical Hybrid Mobility Management that works in IoT with Edge-Fog Cloud architecture. The proposed BFH^2M^2 uses the hybrid location scheme that combines direct location with indirect location schemes depending on a moving device characteristics. Also, BFH^2M^2 uses time-based location update mixed with geographic-based location update, and it uses various Bloom filters: counting Bloom filter (CBF), mapping Bloom filter information base (MaBFIB) and attenuated Bloom filter (ABF) as data structures to effectively store and query the location and routing information of a node name. The performance of BFH^2M^2 is evaluated through an analytical model.

Keywords Bloom filter · Edge-Fog Cloud · Information-centric networking Internet of Things · Location mobility · Name-based routing

1 Introduction

The Internet of Things (IoT) holds the promise to improve our lives by introducing innovative services conceived for a wide range of application domains: from industrial automation to home appliances, from healthcare to consumer electronics, and many others facing several societal challenges in various everyday-life human contexts [1]. In the IoT paradigm, wireless sensor networks (WSNs) are considered the most important elements which collect information from their surrounding

I.-H. Bae (✉)
Catholic University of Daegu, Gyeongsan, Gyeongbuk 38430, Republic of Korea
e-mail: ihbae@cu.ac.kr

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_1

environment. WSNs provide a remote access when connecting with IoT elements. Apart from this, the collaboration among heterogeneous information systems exhibit common services [2].

IoT presents the requirements to manage security and privacy, and needs several changes for the future Internet. The innovative services and solutions considered for the future Internet require global interoperability and mobility support, but it also requires a separation of the global networks from the edge networks, since they are highly constrained in terms of computational capabilities, memory, communication bandwidth, and battery power. This poses vulnerabilities for both security and privacy problems are addressed in current Internet [3]. As IoT networks can include hybrid and heterogeneous devices in terms of mobile and non-mobile devices. While most of the IoT applications such as smart home, smart grid, smart building require mostly static devices. But other applications like smart transport, smart vehicles, and smart mobile networks involve more mobile devices as compared to static devices. Therefore, mobile devices are important part of IoT and thus their mobility management also become essential [4].

Most traditional mobility architectures, which strictly manage devices' mobility, accept a common concept of identifier/locator separation so that the device is identified by its identifier and that packets are routed accordingly to its locator [5]. Even if a mobile device changes its location, it can be accessed by updating a mapping between its identifier and its new locator. In these architecture, the mapping is managed for each device, since the devices' movements are independent of each other. The seamless access pays the serious cost of the fatal defeat that mapping cannot be aggregated. This results in the large size of routing information [6]. A promising solution is adopting Information Centric Networking (ICN) [7, 8] as a common mobility. ICN naturally supports client mobility and its simple name-based forwarding is expected to make the current complex mobility management mechanisms simple. This paper propose a BFH^2M^2: Bloom Filter based Hierarchical Hybrid Mobility Management that works in IoT environment with Edge-Fog Cloud architecture. The proposed BFH^2M^2 uses the hybrid location scheme that combines direct location with indirect location schemes depending on a mobile device characteristics. Also, BFH^2M^2 uses time-based location update mixed with geographic-based location update, and it uses various Bloom filters: counting Bloom filter (CBF), mapping Bloom filter information base (MaBFIB) and attenuated Bloom filter (ABF) as data structures to effectively store and query the location and routing information of node names (IDs). The performance of BFH^2M^2 is evaluated through an analytical model.

The remainder of this paper is organized as follows. In Sect. 2 presents an overview of existing mobility managements for IoT and the related element technologies: Edge-Fog Cloud and Bloom filters for BFH^2M^2 design. The system architecture and design of BFH^2M^2 is presented in Sect. 3. In Sect. 4, we undertake an analytical analysis study for the BFH^2M^2. We finally conclude and describe future works in Sect. 5.

2 Related Works

This section presents the major relevant categories of routing protocols that have been proposed for IoT. The Internet Engineering Task Force (IETF) proposed many routing protocols and constrained application protocol (CoAP) [9] that are suitable to the IoT. To avoid wasting network resources through Interests flooding, an alternative approach is to permit origin servers advertising their content offers frequently, whenever new content objects are available in repositories. Origin servers could represent their content offers using Bloom Filters (BFs) that can represent sets in a compact way. This leads to a smaller overhead needed for propagation of content advertisements. Also, BFs provide an efficient mechanism to store and disseminate information, enabling routing protocols to scale to large networks with a high density of nodes. Hence, various routing protocols relying on this Bloom filter have been proposed.

M. Ishino et al. [6] proposed a new routing-based mobility management scheme, Bloom Filter Based Routing (BFBR) for IoT devices. To reduce the size of routing information associated with IoT devices, the scheme adopt a Bloom filter as a data structure to store routing information at routers. The proposed routing-based mobility management scheme eliminates strict mobility management by IoT communication patterns.

Counting Bloom Filter Routing (CBFR) [10] uses space-efficient data structures known as Bloom filter to efficiently store routing tables on the networked devices. Each node in the collection tree stores the address of its direct and indirect child nodes in its local Bloom filter. A packet is forwarded down-tree only if the node's local filter indicates the presence of the packet's destination address among the node's descendants. In order to cater for the presence of mobile nodes, the scheme apply the concept of counting Bloom filters to allow for removal of elements from the filter by means of gradual forgetting.

In Named Data Networking (NDN), temporary copies of a content object might be cached on route to the nodes that provide the permanent copies of the content object. This possibility of in-network caching enables consumers to retrieve content objects from the caches that are closer than servers. In Bloom Filter-based Routing (BFR) [11], only origin servers perform BF-based content advertisement. Nevertheless, nodes receive the content advertisement of an origin server from all the paths on route to the origin server and populate their Forwarding Information Bases (FIBs) accordingly. Further, BFR adopt the multicast strategy for forwarding Interest. Therefore BFR forwards each Interest in parallel through all the paths towards the origin server of its demanded content object. The Interest could be satisfied from the caches before reaching the origin server.

2.1 Edge-Fog Cloud

IoT typically involves a large number of smart sensors sensing information from the environment and sharing it to a cloud service for processing. This leads to two

major issues for computing IoT-generated data: (i) the processing time of time-critical IoT applications can be limited by the network delay for offloading data to cloud, and (ii) uploading data from a large number of IoT generators may induce network congestion thus incurring further network delay [12].

Figure 1 shows the architecture of Edge-Fog Cloud. Edge of the network implies a number of characteristics that make the Fog a nontrivial extension of the Cloud: edge location, location awareness, low latency, geographical distribution, large-scale sensor networks to monitor the environment, very large number of nodes, support for mobility, real-time interactions, wireless access, heterogeneity, etc.

Fog computing [13] is virtualized platform that provides computation, storage and networking services between end devices and cloud computing data centers, typically, but not exclusively located at the edge of network. Computation, storage and networking resources are the building blocks of both Cloud and Fog.

Unlike the traditional cloud model, the core of the Edge-Fog Cloud has no computational capabilities and only serves as a repository for archiving all data in the cloud. A centralized store provides reliability and easy access to data by any computing resources in the cloud. Being at the core of the architecture, the data store is accessible by both Edge and Fog layers.

2.2 *Bloom Filters*

A Bloom filter (BF) [14] is a space-efficient solution used as a filter for identifying input memberships. It is probabilistic data structures used to test whether an element is a member of a set. False positive matches are possible, but false negatives are not.

A standard Bloom filter is defined as an array of bits used for representing a set, $S = \{x_1, x_2, \ldots, x_n\}$ of n elements. The hash function $H(x)$ is a key function to insert the element x into a set S of a Bloom filter. Let $H = \{h_1, h_2, \ldots, h_k\}$ be the set of hash functions to map element $x \in S$, each with range {1,…, m}. Each hash

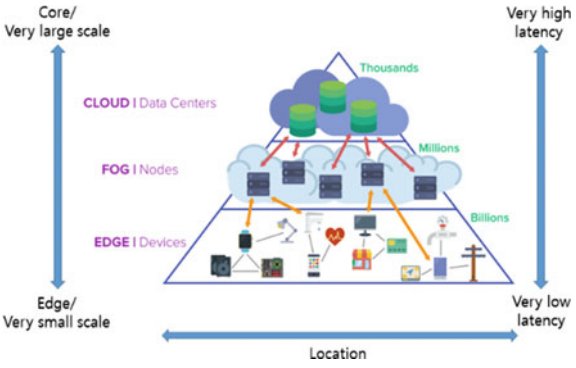


Fig. 1 Edge-Fog Cloud architecture

function yields a bit location and represent x by setting the bits of hash index $h_i(x)$ as 1 for $1 \leq i \leq k$. Initially, all m bits of the filter are set to zero, where m is the number of bits in the Bloom filter. In order to store element x into the Bloom filter using the hashing function $H(x)$, a logical OR operation is executed between the bit arrays. Conversely, to search for an element in the Bloom filter, the bit array is generated by hashing. An array of bits is compared with the Bloom filter to guarantee the membership of an element in the Bloom filter. If the bit locations $h_i(x)$ are set to 1 in an array, and are also 1 in the Bloom filter, then an element is assumed a member of the Bloom filter with some probability. An example of inserting three elements into a BF with a parameter set $\{n = 3, k = 3, m = 15\}$ is presented in Fig. 2.

Attenuated Bloom filter (ABF) [15] consist of multiple (d) layers of standard Bloom filters. The first layer of the filter contains the information for the current node, while the second layer contains the information i hops away in the i-th layer. By using ABF consisting of multiple layers, context sources at more than one hop distance can be delivered, while avoiding saturation of the Bloom filter by attenuating bits set by sources further away. Figure 3 shows the example of the information aggregation operation for a node with two neighbors.

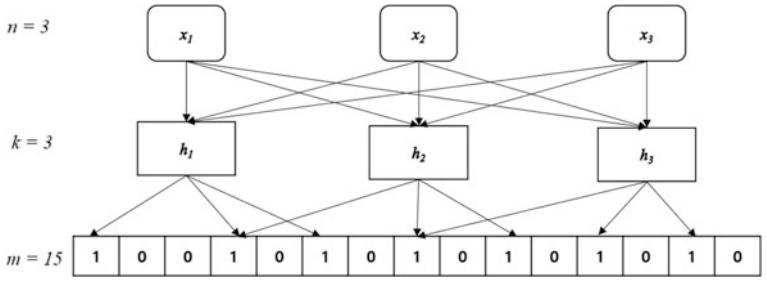


Fig. 2 Bloom filter with three functions

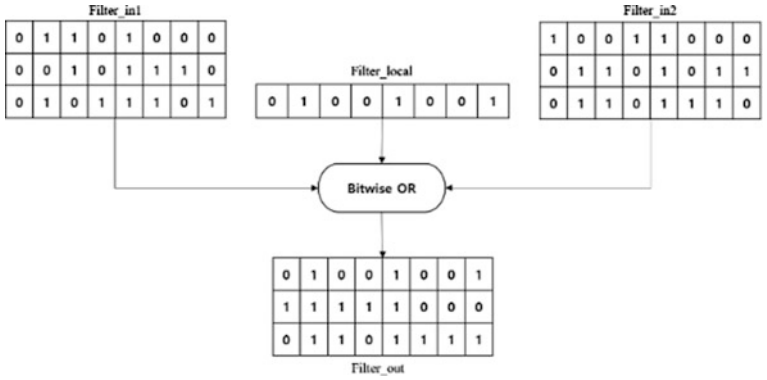


Fig. 3 An example of attenuated Bloom filter

3 BFH^{2}M^2 Design

In this section, we present an overview of our BFH^{2}M^2, where the Edge-Fog Cloud based system architecture that the proposed BFH^{2}M^2 is running environment, the necessary data structures at each layer for BFH^{2}M^2, and the process of our BFH^{2}M^2 based routing for IoT devices are described.

3.1 System Architecture

The system architecture of Edge-Fog Cloud model for supporting BFH^{2}M^2 consists of 3 layers: edge, fog and cloud as shown in Fig. 4, there are a router between layers. Edge nodes provide computing and storage capabilities at the edge of the wireless network in close proximity of mobile devices. Edge can minimize latency, conserve network bandwidth, reduce cross-domain traffic, and make better location-awareness and context-awareness decisions. Fog severs have functions as middleware. They are simpler and with less capabilities than cloud servers and data centers. Also, they containing the replicas and records of the required data can help in reducing transaction failures and support the quality of services to the users through periodic cloud updates, for available data. The top layer is at data centers of the cloud, which collects data and information from the fog servers on fog layer.

The mapping system stores the relationship between the locator and the identifier. It is comprised of local mapping systems within different routing areas. To avoid introducing new entities to the network, the mapping system can be deployed on routers. The edge routers initially obtain local routing information (LRI) of the

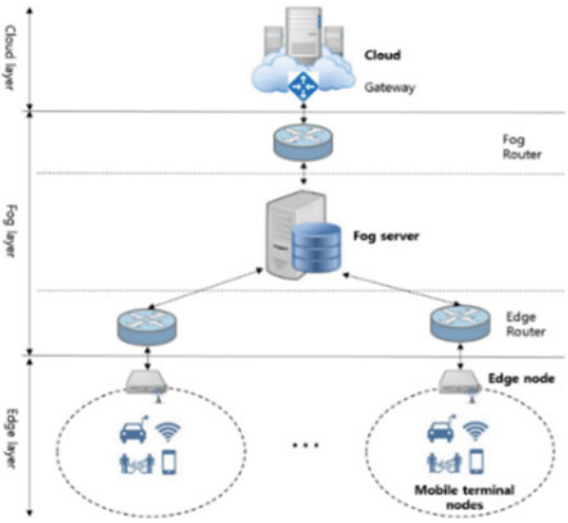


Fig. 4 System architecture for BFH^{2}M^2

IoT devices in each edge networks. Edge routers has been assigned the task to compute the probability of node using node properties that is incoming request rate and location of node in network. Then, the fog routers have some routing information for the hot nodes over underlying multiple edge networks, where the hot nodes represent that whether their location information are frequently requested from other nodes in IoT or their time interval of location movement to neighbor cells are less than the location update interval in time-based location update scheme. Therefore the edge routers support micro-community based locality over an edge network, and also the fog routers support macro-community based locality over the fog region, a group of edge networks. Each router has different type of BFs for resolving routing-based mobility by In-network computation. In-network computation include easy management of mobile nodes, less and refined cached data, simple data routing and forwarding and hence it can improve network-life, battery life at the cost of simple and optimal in-network computation algorithms.

3.2 Device Mobility Management

ICN-IoT introduces a unique identifier (ID) for every network object, separated from its dynamic access network locations represented by network addresses. The separation of naming and addressing allows ICN-IoT to support identity based routing. Naming the data makes, in-network caching easy and efficient as named data is easy to manage and search. Therefore, naming reduces information retrieval delay. Moreover, multicast forwarding outperforms because of aggregation of name base requests. In BFH^2M^2, each networking element is no longer identified by its location, but by a content name. ICN uses a hierarchical, structural, human-readable, unbound name. Each networking process is also based on content name. In proposed BFH^2M^2 scheme, the name of each mobile device within edge layer is made up of four parts: domain name, organization, device type and device ID. The domain name and organization is the prefix of the router that stores the device's mapping information. The device ID is the globally unique identity of the device. Example of proposed device naming scheme is /cu/eng/it535/user_terminal/smartphone, where /cu, /eng/it535, /user_terminal and /smartphone represent domain name, organization, device type and device ID components, respectively. Figure 5 shows the BF structure for proposed device naming format.

IoT applications involve not only nearly stationary devices but also more mobile devices, so a hybrid mobility management scheme that combines direct location scheme with indirect location scheme is needed depending to the IoT device properties: incoming request and location update rates. When a node updates its location servers according to direction location scheme, the node computes its responsible cells. Position updates are then sent to all responsible cells. In indirect location scheme [16], the location information is represented as a pointer to the pointer to the position of the node. The network load can be reduced with the indirect location scheme where the location servers on higher hierarchy levels only

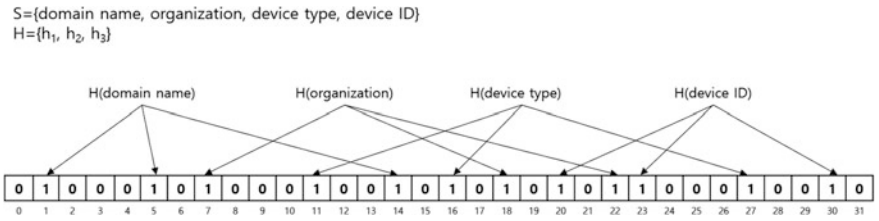


Fig. 5 Example of the BF structure for proposed device naming format

know the region of the next lower level, a node is located in. More precise location information in not necessary on higher levels.

Proposed BFH^{2}M^2 uses a hybrid location scheme that each layer within Edge-Fog Cloud uses direct location scheme or indirect location scheme. If a mobile node is hot, BFH^{2}M^2 uses geographic based direct location update scheme until fog layer. Otherwise, BFH^{2}M^2 uses time based direct location update scheme until edge layer. The higher layers that do not apply direct location use indirect location update scheme. Figure 6 shows the overview of device mobility process in proposed BFH^{2}M^2.

In BFH^{2}M^2, each layer uses different data structures for storing location information because it uses different location schemes. BF involves three operations: build up, lookup and update. BF with on-bit wide array provides fast lookup. CBF [17] with counters is utilized to support the insertion or deletion of pattern. BF is used for lookup operation and CFB is used for update operation. Therefore, BF is used for lookup the pointer to the location of the node in indirect location scheme, while CBF is used for performing location update in direct location scheme.

We assume that the devices, C and F in mobile edge networks are hot mobile device, and the edge routers initially have routing information of the IoT devices in each edge networks. Each edge router aggregates the routing information of cold IoT devices toward upstream Fog router periodically by the bitwise OR operation of its BFs and sends the aggregated BF to the upstream cloud gateway. When the hot

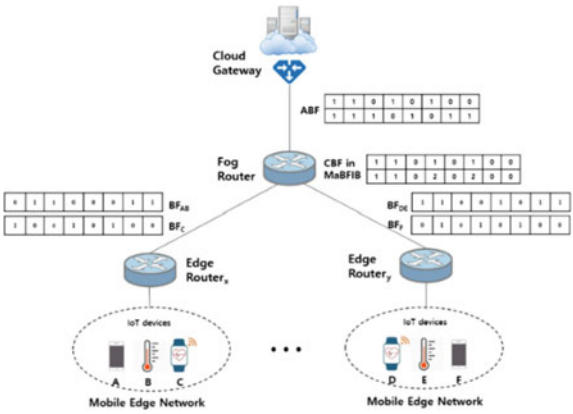


Fig. 6 Overview of device mobility process by BFH^{2}M^2 scheme

mobile devices geographically enter into local edge networks, the corresponding Edge router sends a location information message with the BF of device ID and local locator information (LLI) to the upstream Fog router, stores the aggregated BFs in the CBF of MaBFIB for In-network routing computation. After waiting for up to a tolerance time for mobility delay, the Fog router sends the corresponding BF of the CBF as the second layer from the bottom of ABF to the upstream cloud gateway. Figure 7 illustrates the structure of location information message, where *type* indicate whether the mobile device is hot or cold, and *mobility delay tolerance time* indicates the maximum time permitted until the mobility is done.

In case of cold device in example of Fig. 6, the Fog router aggregates BF_{AB} and BF_{DE} from the routers of lower layer: Edge $Router_x$ and Edge $Router_y$, then sends the aggregated BF, bit array (11101011) to the upstream cloud gateway. Otherwise, the Fog router aggregates BF_C and BF_F from the routers of lower layer: Edge $Router_x$ and Edge $Router_y$, stores the aggregated BFs in the CBF. Then the Fog router sends the corresponding BF of the CBF, bit array (11010100) to the upstream cloud gateway. Therefore, the message sequence diagram for the location query process in BFH^2M^2 scheme is show in Fig. 8.

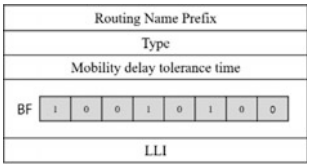


Fig. 7 Structure of location information message

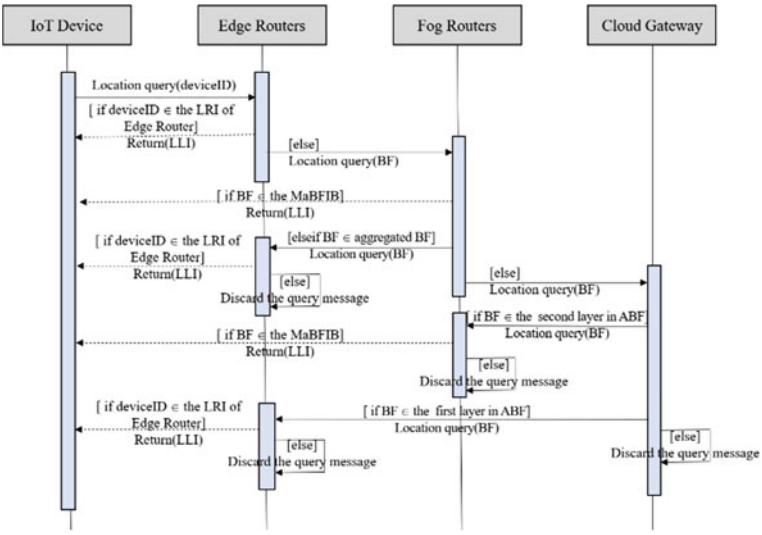


Fig. 8 Message sequence diagram for the location query process in BFH^2M^2

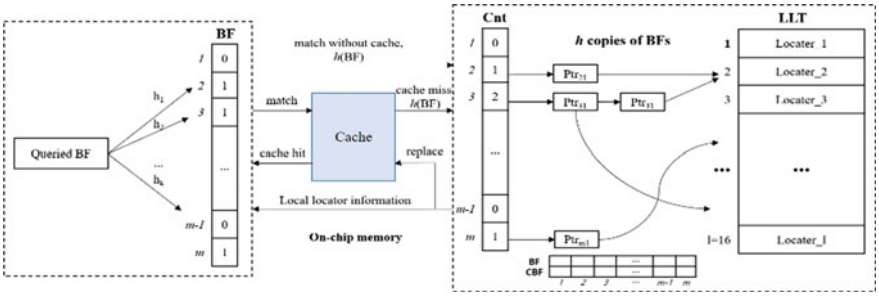


Fig. 9 MaBFIB structure for locating of hot devices in BFH^{2}M^2

The data structure called mapping bloom filter (MBF), which is an improvement of BF, support querying and mapping the set elements in the memory and lowers on-chip memory consumption [18]. Fog routers in BFH^{2}M^2 system maintain a data structure, MaBFIB that can effectively reduce the memory space and can accelerate the lookup process. The MaBFIB consists of two parts: on-chip such as cache, off-chip such as CBF and Local Locator Table (LLT) as illustrated in Fig. 9, where h(BF) represents the hash function of an Iterated Bloom Filter by the input of a BF.

Suppose that BFH^{2}M^2 is executed in the Edge-Fog Cloud based system with grid hierarchy which a fog server has 4 edge nodes and an edge node has 4 cells, the number of local locators of a fog router is 16, l = 16 as shown in Fig. 9. Then h represents of the number of hot IoT devices in a fog region.

4 Performance Evaluation

In this section, we analyze the performance of BFH^{2}M^2 through an analytical model. The notations to be used in the analysis are listed below.

λ/μ	Call to mobility ratio (CMR)
C_{h_update}	Update occurrence rate for a hot device within time based location update interval
C_{c_update}	Update cost for a hot device
C_{c_update}	Update cost for a cold device
P_{mh}	Probability that moving device is hot
P_{qh}	Probability that queried device is hot
P_{edge}	Probability that local edge router has local location information
P_{fog}	Probability that local fog router has routing information
C_{q_edge}	Query processing cost in edge router
$C_{h\_q\_fog}$	Query cost in fog router for a hot device
$C_{c\_q\_fog}$	Query cost in fog router for a cold device
$C_{c\_q\_cloud}$	Query cost in cloud gateway for a cold device

In the above notations, the CMR is computed by dividing the average number of calls made to the mobile device by the average number of cell crossings, the mobile device makes in a given period of time. We define a simple model for device requests where the requests are independent and distributed according to a Zipf-like distribution [19].

Let N be the total number of IoT devices in an Edge-Fog Cloud system. Let $P_N(i)$ be conditional probability that, given the arrival of a device request, the arriving request is made for device i. Let all the devices be ranked in order of their popularity where node i is the i-th most popular device. We assume that $P_N(i)$, defined for i = 1, 2,…, N, has a Zipf-like distribution given by Eq (1).

$$P_N(i) = \frac{1}{i^{\alpha}} \left(\sum_{i=1}^{N} \frac{1}{i^{\alpha}} \right)^{-1} \tag{1}$$

If we assume that if $P_N(i) \geq 0.015$, device i is hot. The cost of location mobility is evaluated by a total location management cost that adds the location update cost and location query cost.

$$C_{total} = C_{update} + \frac{\lambda}{\mu} C_{query} \tag{2}$$

$$\begin{aligned} C_{update} &= P_{mh}(r_{h\_update} C_{h\_update}) + (1 - P_{mh}) C_{c\_update} \\ C_{query} &= P_{mh}\{P_{edge} C_{q\_edge} + (1 - P_{edge}) C_{h\_q\_fog}\} + (1 - P_{mh})\{P_{edge} C_{q\_edge} \\ &\quad + P_{fog} C_{c\_q\_fog} + [1 - (P_{edge} + P_{fog})] C_{c\_q\_cloud}\} \end{aligned}$$

Table 1 shows the parameters and the values for analytical performance evaluation, where the cost used in analytical evaluation are normalized values.

Figure 10 shows the total cost for mobility management schemes over call to mobility ratio, where the performance BFH^{2}M^2 is compared with the performance of BFBR that is time-based location update is used and all upper routers of the

Table 1 Parameters and values for analytical evaluation

Parameter	Value
C_{h_update}	4
C_{c_update}	2
P_{mh}	0.3
P_{qh}	0.7
P_{edge}	0.3
P_{fog}	0.5
C_{q_edge}	1
$C_{h\_q\_fog}$	2
$C_{c\_q\_fog}$	3
$C_{c\_q\_cloud}$	4

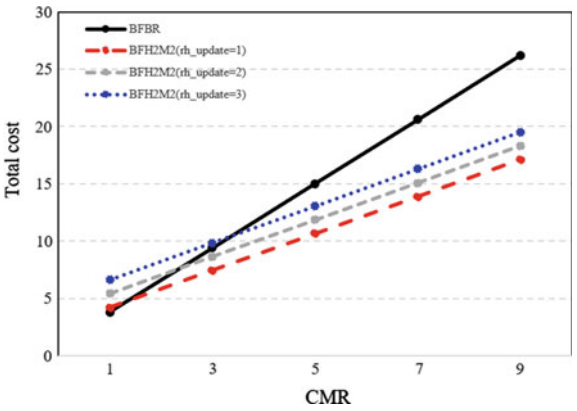


Fig. 10 Performance of mobility management schemes over CMR

bottom edge router is used indirect location scheme. As shown in the Fig. 10, the performance of proposed BFH^{2}M^2 is superior to BFBR except very small CMR. In case that CMR = 1, the performance of BFH^{2}M^2 approximately equals to or little worse than BFBR because BFH^{2}M^2 has a lot of update cost for hot devices. Also, the mobility of hot devices is the lower, the performance of our BFH^{2}M^2 is the better. Therefore, our BFH^{2}M^2 can support seamless mobility with fast and efficient connectivity among IoT devices.

5 Conclusion

The mobility management schemes for IoT devices should support seamless mobility because most of IoT applications involve not only static devices but also mobile devices. To provide fast, reliable connectivity and make data available at everywhere, network architecture should support seamless mobility and roaming. However, existing BF-based mobility management schemes did not support seamless mobility because time-based location update was used and all upper routers of the bottom edge router was used indirect location scheme. To resolving these problems of BF-based mobility management schemes, we proposed a hybrid mobility management scheme called BFH^{2}M^2 that combines geographic based direct location update scheme with time based direct location update scheme. The performance of proposed BFH^{2}M^2 was evaluated through an analytical model. As a result, we confirmed that our hybrid-based BFH^{2}M^2 is superior to time-based BFBR.

In our future work, we will research a mathematical model that decides hot devices depending on location update intervals and the cache structure design for routers in proposed BFH^{2}M^2.

Acknowledgements This paper was supported by research grants from Daegu Catholic University in 2017.

References

1. Militano L, Araniti G, Condoluci M, Farris I, Iera A (2015) Device-to-device communications for 5G Internet of Things. EAI Endorsed Trans Internet Things 1(1):1–15
2. Ghaleb SM, Subramaniam S, Zukarnain ZA, Muhammed A (2016) Mobility management for IoT: a survey. EURASIP J Wirel Commun Network 1–25
3. Jara AJ, Kafle VP, Skarmeta AF (2013) Secure and scalable mobility management scheme for the Internet of Things integration in the future internet architecture. Int J Ad Hoc Ubiquit Comput 13(3/4):228–242
4. Arshad S, Azam MA, Rehmani MH, Loo J (2018) Recent advances in information-centric networking based Internet of Things (ICN-IoT). IEEE communication survey & tutorials
5. Zhu Z, Wakikawa R, Zhang L (2011) A survey of mobility support in the internet. RFC 6301, Internet engineering task force (IETF)
6. Ishino M, Koizumi Y, Hasegawa T (2015) A routing-based mobility management scheme for IoT devices in wireless mobile networks. IEICE Trans Commun E98-B(12):2376–2381
7. Amadeo M et al (2016) Information-centric networking for the Internet of Things: challenges and opportunities. IEEE Netw 30(2):92–100
8. Lee J, Cho S, Kim D (2012) Device mobility management in content-centric networking. IEEE Commun Mag 50(12):28–34
9. Sheng J, Yang S, Yu Y, Vasilacos A, Mccann JJ, Leung K (2013) A survey on the IETF protocol suite for the Internet of Things: standards, challenges, and opportunities. IEEE Wirel Commun 20(6):91–98
10. Reinhardt A et al (2012) CBFR: bloom filter routing with gradual forgetting for tree-structured wireless sensor networks with mobile nodes. In: IEEE international symposium on a world of wireless, mobile and multimedia networks (WoWMoM), San Francisco, pp 1–9
11. Maradi A, Braun T, Salamatian K, Thomos N (2017) BFR: a bloom filter-based routing approach for information-centric networks. IFIP Network 1–9
12. Mohan N, Kangasharju J (2016) Edge-Fog Cloud: a distributed cloud for Internet of Things computations. In: International conference cloudification of the Internet of Things, Paris, pp 1–6
13. Bonomi F, Milito R, Zhu J, Addepalli S (2012) Fog computing and its role in the Internet of Things. In: Proceedings of the first edition of the MCC workshop on Mobile cloud computing, Helsinki, pp 13–16
14. Talpur A, Shaikh FK, Newe T, Sheikh AA, Felemban E, Khelil A (2017) Bloom filter-based efficient broadcast algorithm for Internet of Things. Int J Distrib Netw 13(12):1–12
15. Liu F, Heijenk G (2006) Context discovery using attenuated bloom filters. In: International conference on wired/wireless internet communications, Bern, pp 13–25
16. Bae I-H, Liu H (2007) FHLS: fuzzy hierarchical location service for mobile ad hoc networks. In: IEEE international fuzzy systems conference, London, pp 1–5
17. Dautov R, Distefano S (2017) Targeted context delivery to IoT devices using bloom filters. In: International conference on ad-hoc networks and wireless, Messina, pp 39–52
18. Li Z, Liu K, Liu D, Shi H, Chen Y (2017) Hybrid wireless networks with FIB-based named data networking. EURASIP J Wirel Commun Network 1(54):1–7
19. Lee B, Cao P, Pan L, Phillips G, Shenker S (1999) Web caching and Zipf-like distributions: evidence and implications. In: Proceedings IEEE INFOCOM, New York, pp 126–134

Probabilistic Classification Models for Reliable Vertical Handoff in Heterogeneous Wireless Networks

C. S. Jayasheela, Prakash B. Metre and Gowrishankar

Abstract *Cooperative Communication* refers to a communication framework in Next Generation wireless networks. In this framework, multiple wireless networks provide mutual cooperation, such that, the wireless network user is provided with requested services. The *Vertical Handoff* is one such technique, which is used in cooperative communication. In this technique, if a client cannot be provided with the requested Quality of Service (QoS), then, the client is migrated from the current wireless network to another available wireless network, which can deliver the requested QoS. To perform this vertical handoff, important parameters such as service cost, data transmission rate, the speed of the mobile device, network latency, the battery level of the mobile device, received signal strength information (RSSI) etc are usually considered. Many vertical handoff techniques in the literature have focused on effective network selection problem, wherein, the most suitable network is searched for performing vertical handoff. The suitability of a network depends on its ability to provide requested QoS. All these proposed techniques have given excellent performance in achieving this goal. But, most of these techniques have ignored the issue of *Reliability* in making vertical handoff decisions, wherein, the selected network to which handoff will be made should provide requested QoS for a prolonged period, and avoid frequent new handoff scenario. Frequent handoff can be expensive and inefficient for the client. Hence, the handoff decision should achieve the dual goal of suitability and reliability. In this work, a new vertical handoff scheme is proposed, which achieves the dual goal of suitability and reliability. Two probabilistic classification models namely Bayesian classification

C. S. Jayasheela (✉)
Department of ISE, BIT, VTU, Bangalore, India
e-mail: sheeladhan1@gmail.com

P. B. Metre
Department of ISE, AIT, VTU, Bangalore, India
e-mail: prakash.metre@gmail.com

Gowrishankar
Department of CSE, BMSCE, VTU, Bangalore, India
e-mail: gowrishankar.cse@bmsce.ac.in

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_2

model and Sample Distribution model are designed for building the proposed vertical handoff scheme. The empirical results demonstrate the effectiveness wrt reliability and suitability of both these models in making handoff decisions.

Keywords Vertical handoff · Quality of service · Reliability · Suitability

1 Introduction

1.1 Overview on Vertical Handoff

Currently, wireless network users are being provided with resource intensive QoS. Many wireless applications demand resource intensive QoS such as video conferencing, voice transfer, messaging service etc. Due to large number of users, satisfying these QoS demands becomes difficult. To provide continuous satisfaction and meet the QoS demands of the client, cooperative communication becomes inevitable. In this framework of cooperative communication, multiple wireless networks are involved in providing communication service to a user of another wireless network. This framework ensures that, the user QoS demands are satisfied continuously, and the idle resources of other wireless networks can be utilized efficiently.

It is essential that, to build an effective cooperative communication framework, the mobile nodes should be aware of heterogeneous network structure, and seamless network switching should be possible. To connect different networks, cognitive radio provides an efficient mechanism to achieve this task.

The vertical handoff is one of the techniques employed in cooperative communication framework. If a user is not provided with the requested QoS by the existing network, then, the vertical handoff technique migrates the user from the existing network to another nearby wireless network, which can satisfy the users QoS demands. There is a rich contribution in the area of vertical handoff [1–20]. The main focus of these available vertical handoff techniques in the literature is to select suitable network to perform the handoff, which has good resources, satisfies the QoS constraints and enforces less monetary cost. Many classification schemes such as Neural Networks, Support Vector Machines etc have been utilized to build the handoff schemes. Handoff parameters such as—monetary cost, data rate, RSSI etc were used to make the handoff decision.

1.2 Motivation

Consider a scenario, wherein, a client is provided with communication service through wireless network w_i. The client has specified QoS constraints (handoff parameters) on different communication parameters such as—minimum bandwidth, RSSI and Monetary cost. Due to heavy communication load in $\boldsymbol{w_i}$, assume that,

after certain period w_i is unable to satisfy the clients QoS demands. Then, it is necessary that, a handoff decision is made to migrate the client to another network which can satisfy QoS demands of the client.

Assume that, the wireless network w_j can satisfy the clients QoS demands, and it is selected through handoff decision. Initially, the communication load in w_j might be limited, but, after some period of time, w_j might experience heavy communication load. In this scenario, w_j might not be able to satisfy clients QoS demands, which leads to the requirement of making another handoff decision. The handoff choices available during this scenario might not be monetarily attractive to the client, and the client is forced to agree for higher monetary cost. Also, frequent handoff scenarios can lead to lower effectiveness in communication service provided by wireless networks. It is critical that, the handoff decisions should also consider the reliability issue, but, until now, this issue has not been effectively addressed in the literature.

1.3 Contributions

The following contributions are made in this work:

1. A new vertical handoff scheme is proposed, which chooses suitable networks that can satisfy the clients QoS demands, and provides the expected QoS for a prolonged period.
2. The proposed vertical handoff scheme is built by using 2 probabilistic classification models namely—Bayesian Classification model and Sample Distribution model. Thus, 2 variations of the proposed vertical handoff scheme are presented.
3. Both the variations of the proposed handoff scheme are implemented in Network Simulator 3. The predicted reliability values of the proposed handoff scheme are compared against the actual reliability values obtained through simulation. The proposed handoff scheme exhibits its reliability performance by providing tight correlation between predicted reliability values and simulated reliability values.

1.4 Paper Organization

This paper is organized as follows: Sect. 2 describes the related work. The Bayesian Classification model and Sample Distribution Models are presented in Sect. 3. The proposed vertical handoff scheme is presented in Sect. 4. The empirical results are presented in Sect. 5. Finally, the work is concluded in Sect. 6.

2 Related Work

Many vertical handoff schemes based on artificial intelligence have been proposed. The handoff scheme proposed in [1] utilized handoff parameters such as-bandwidth, RSSI and monetary cost. The handoff scheme was based on fuzzy logic approach. The weight of the QoS metrics was modified if the conditions of network changed.

Handoff parameters such as-coverage area, bandwidth, power consumption and sojourn time were utilized in [2]. Different wireless networks such as-GPRS, GSM and WLAN were involved in the handoff framework. Fuzzy model was utilized in building the handoff scheme.

In [3], a vertical handoff scheme was proposed, which involved wireless networks such as-UMTS and WLAN. Handoff parameters such as mobile speed, bandwidth and other user details were utilized. Fuzzy model was adopted to build the handoff scheme.

The vertical handoff scheme proposed in [4] utilized both Neural Networks and fuzzy logic. Multi criteria decision model was proposed to make handoff decisions. Handoff parameters such as-bit error rate, signal to interference ratio and data transmission rate were utilized.

In [5], wireless networks such as-WiMAX and UMTS were involved in the handoff scheme. Handoff parameters such as-RSSI and monetary cost were utilized. A hybrid model which combined genetic algorithm and fuzzy model was utilized to build the handoff scheme.

Many Neural Network based handoff schemes have been proposed [6, 7, 20]. In [6], handoff parameters such as-RSSI and traffic intensity were utilized. In [7], handoff parameters such as-security, cost, network transmission range and network capacity were utilized. In [20], handoff parameters which reduce handoff latency were used.

3 Probabilistic Classification Models

3.1 Bayesian Classification Model

This model performs classification of the data point x, which is a d × 1 random vector by choosing a suitable class out of available K classes indicated by C_1, C_2… $C_K (1 \leq k \leq \mathrm{K})$. The data point x is assigned to that class k which has the highest class probability, and which is represented in Eq. 1. Here, $\mathrm{P}(\mathrm{C}_k|\mathrm{x})$ indicates the class probability of C_k wrt data point x. It is assumed that, class probability represented in Eq. 1 is having the distribution represented in Eq. 2. Here, a_k is a parameter which is represented in Eq. 3.

$$P(\mathrm{C_k}|\mathrm{x}) = \frac{\mathrm{P}(\mathrm{x}|\mathrm{C_k})\mathrm{P}\ (\mathrm{C_k})}{\sum_{\mathrm{j}} \mathrm{P}\ (\mathrm{x}\ |\mathrm{C_j})\mathrm{P}(\mathrm{C_j})} \tag{1}$$

$$\boxed{P(C_k|\mathrm{x}) = \frac{exp(\mathrm{a}_k)}{\sum_j exp(a_j)}} \tag{2}$$

$$\mathrm{a}_k = \log(P(\mathrm{x}|C_k)P(C_k)) \tag{3}$$

The posterior probability indicated by $P(\mathrm{x}|C_k)$ is represented in Eq. 4.

$$P(\mathrm{x}|C_k) = \frac{1}{2\pi^{\frac{\mathrm{d}}{2}}|\sum|^{\frac{1}{2}}} exp\left(\frac{-1}{2}(\mathrm{x}-\mu_k)^{\mathrm{T}}\sum^{-1}(\mathrm{x}-\mu_k)\right) \tag{4}$$

By using the result of Eq. 4, the parameter $\mathrm{a_k(x)}$ is redefined as represented in Eq. 5.

$$a_k(\mathrm{x}) = \mathrm{w}_K^T\mathrm{x} + w_{k0} \tag{5}$$

Both the parameters $\mathrm{w_k}$ and w_{k0} are represented in Eqs. 6 and 7 respectively.

$$\mathrm{w_k} = \sum^{-1}\mu_{\mathrm{k}} \tag{6}$$

$$w_{k0} = -\frac{1}{2}\mu_{\mathrm{k}}^T\sum^{-1}\mu_{\mathrm{k}} + \log P(C_k) \tag{7}$$

The parameter μ_k indicates the mean of kth class, which is approximated through sample mean as represented in Eq. 8. Here, $\mathrm{x}_1^{\mathrm{k}}, \mathrm{x}_2^{\mathrm{k}}\ldots\mathrm{x}_{nk}$ are the data points of the training set which belong to kth class.

$$\mu_{\mathrm{k}} \approx \frac{\sum_{i=1}^{nk}\mathrm{x}_i^{\mathrm{k}}}{n_k} \tag{8}$$

The covariance of kth class is indicated by $\sum$, which is represented in Eq. 9. Each class is assumed to have the same covariance matrix.

$$\sum \approx \frac{1}{n_k - 1}\sum_{i=1}^{nk}(\mathrm{x}_i^k - \mu_k)(\mathrm{x}_i^{\mathrm{k}} - \mu_k)^T \tag{9}$$

Equation 10 represents the prior probability $P(C_k)$.

$$P(C_k) = \frac{1}{K} \tag{10}$$

The data point x is assigned to that class, which has the highest class probability, which is represented in Eq. 2.

3.2 Sample Distribution Model

This classification model is also probabilistic in nature, but, it does not assume any distribution for the class probability represented in Eq. 2. Instead, the distribution is estimated from the samples. To achieve this task, the sample distribution function has to be constructed first. Then, the class probability is estimated from the constructed sample distribution function. The sample density function, which can directly approximate the class probability is not utilized, because of its poor convergence to actual density function.

Let, y be a continuous random variable, $f(y)$ and $f(y|\mathrm{x})$ represents the density and conditional density function of y respectively and x represents a data point which is a d × 1 random vector. Let, $F(y)$ and $F(y|\mathrm{x})$ represent the distribution and conditional distribution of y respectively. Let, the training set contain r samples, where each sample is an ordered pair, such that, the first element of the ordered pair contains a data point $\mathrm{x_i}(1 \leq i \leq r)$, and the second element is the value y_i associated with $\mathrm{x_i}$. Here, $\mathrm{x_i}$ is considered as the observation of x and y_i is considered as the observation of y.

The joint sample distribution function of (y, x) is represented in Eq. 11. Here, $\mathrm{F_r}(y, \mathrm{x})$ indicates the sample joint distribution function of (y, x), $\theta(\mathrm{x} - \mathrm{x_j}) = 1$, if every component of $(\mathrm{x} - \mathrm{x_j})$ is positive, otherwise $\theta(\mathrm{x} - \mathrm{x_j}) = 0$. Also, $\delta(y, \mathrm{x_j}) = 1$ if the ordered pair (y, $\mathrm{x_j}$) exists in the training set, otherwise, $\delta(y, \mathrm{x_j}) = 0$.

$$F_r(y, \mathrm{x}) = \frac{1}{r}\sum_{\mathrm{j}=1}^{\mathrm{r}} \theta\left(\mathrm{x} - \mathrm{x_j}\right) \times \delta(y, \mathrm{xj}) \tag{11}$$

The sample distribution function converges in probability to actual distribution function. This case is represented in Eq. 12.

$$\lim_{r\to\infty,P} |F(y|\mathrm{x}) - F_r(y|\mathrm{x})| \to 0 \tag{12}$$

By using Bayes theorem, result represented in Eq. 13 can be obtained. Here, $F_r(y|\mathrm{x})$ represents the sample conditional probability for y conditioned wrt x, $F_r(\mathrm{x})$ represents the sample distribution function of x, which is represented in Eq. 14.

$$F_r(y|\mathrm{x}) = \frac{F_r(y, \mathrm{x})}{F_r(\mathrm{x})} \tag{13}$$

$$F_r(\mathrm{x}) = \frac{1}{r}\sum_{\mathrm{j}=1}^{\mathrm{r}} \theta\left(\mathrm{x} - \mathrm{x_j}\right) \tag{14}$$

By using the property of density and distribution functions, the result represented in Eq. 15 is obtained. Here, $f(y|\mathrm{x})$ represents the conditional density function of y conditioned wrt x and Δy represents an extremely small real number.

$$f(y|\mathrm{x}) = \approx \frac{F_r(y+\Delta y|\mathrm{x}) - F_r(y|\mathrm{x})}{\Delta y} \tag{15}$$

To perform classification using the empirical distribution model, the classes C_1, C_2,… C_k are labeled as 1, 2,… K. The random variable y is defined to take the values in the set [1, K]. The class probability of the data point x wrt the class $C_k(1 \leq k \leq K)$ is represented in Eq. 16.

$$\boxed{P(C_k|\mathrm{x}) = f(y = k|\mathrm{x})} \tag{16}$$

4 Vertical Handoff Scheme for Reliable Communication

4.1 Training Set Generation

Each wireless network is represented through a feature vector, which is indicated in Eq. 17, and contains the information about QoS provided by the network. Here, x represents the feature vector, *br* represents the bit rate provided by the wireless network, *co* represents the monetary cost to transmit p data packets and *rssi* represents the received signal strength, which indicates the quality of received data.

$$\mathrm{X} = \begin{pmatrix} br \\ co \\ rssi \end{pmatrix} \tag{17}$$

The feature vector of each wireless network in the training set will be labeled by a reliability index, which indicates the reliability of the wireless network in providing the declared QoS. The reliability index values are indicated by (0, 1, 2,…, 9). Here, 9 indicates the lowest reliability and 0 indicates the highest reliability. Each index value uniquely represents a class $C_k(1 \leq k \leq K)$, where K = 10.

The reliability value of the training set vector $\mathrm{x_j}$ is represented in Eq. 18. Here, $case_i(\mathrm{x_j})$ is represented in Eq. 19. To obtain handoff statistics, each wireless network in the training set is observed for a series of m independent and equal length time intervals. The ith time interval is indicated by $case_i$. The function *nhandoff* ($case_i$,$\mathrm{x_j}$) represents the number of handoffs designated to the network represented by the feature vector $\mathrm{x_j}$ wrt $case_i$ and $\overline{nhandoff}\left(case_i, \mathrm{x_j}\right)$ represents the number of handoff s which were designated to the network represented by $\mathrm{x_j}$ during $case_i$, but, which had to be provided with new handoff due to traffic congestion.

$$reliability\left(\mathrm{X}_j\right) = \frac{\sum_{i=1}^{m} Case_i(\mathrm{X}_j)}{m} mod\ K \tag{18}$$

$$case_i(\mathrm{X}_j) = \frac{\overline{nhandoff}(Case_{i,}x_j)}{nhandoff(Case_{i,}x_j} \times 100 \tag{19}$$

4.2 Problem Definition

The mobile device user specifies the required QoS through the feature vector represented in Eq. 20. Here, **u** represents the feature vector for the user, br_u represents the required bit rate for the user, co_u represents the maximum cost agreeable to the user and $rssi_u$ represents the required RSSI for the user. The available wireless networks to which the user can migrate represents the test case set. Each wireless network in the test case set is represented by the feature vector shown in Eq. 17, and reliability value label is not available. The feature vectors in the test set which satisfy the user QoS parameter values br_u and $rssi_u$ and the cost parameter value $\leq co_u$ are selected for making handoff decisions, and this selected set is indicated as *handoff_set*. The task is to select suitable wireless network from the *handoff_set* which can provide reliable communication service.

$$\mathrm{u} = \begin{pmatrix} br_u \\ co_u \\ rssi_u \end{pmatrix} \tag{20}$$

4.3 Algorithm Based on Bayesian Classification Model

The vertical handoff technique using Bayesian Classification Model is presented in Algorithm 1. By using the training set which has r ordered pairs, the mean of each cluster C_k indicated by μ_{k} is calculated by using Eq. 8. The common covariance matrix of all the clusters indicated by $\sum$ is calculated by using Eq. 9.

Since, the wireless networks belonging to the *handoff_set* do not have reliability labels, the task is to estimate the reliability labels, so that, the wireless network $\in$ *handoff_set* which has the lowest reliability index value will be selected for vertical handoff. To achieve this task, the class probability of the Bayesian Classification Model represented in Eq. 2 is utilized. For each feature vector that represents a wireless network $\in$ *handoff_ set*, its class probability for each of the K classes is calculated. The class which has the highest probability will become the designated class of the feature vector. The corresponding reliability index value of the designated class will become the reliability label of the wireless network. After assigning reliability labels to every feature vector $\in$ *handoff_set*, the feature vector, which corresponds to a wireless network that has the lowest reliability index value is selected for vertical handoff.

Algorithm 1 Handoff Technique Based on Bayesian Classification Model

Let the training set contain r ordered pairs.
Estimate the mean of each cluster by using Eq. 8.
Estimate the common covariance matrix for all the clusters by using Eq. 9.
for each feature vector **x** $\in$ *handoff_set* **do**
  for each class C_k **do**
    Calculate $P(C_k|\mathrm{x})$ by using Eq. 2.
  end for
  The class which has the highest probability will become the designated class of **x**.
  Label **x** with the reliability index value of the designated class.
end for

Select the feature vector $\in$ *handoff_set* which has the lowest reliability index value.

Perform the handoff to the wireless network corresponding to the selected feature vector.

4.4 Algorithm Based on Sample Distribution Model

The vertical handoff technique based on Sample Distribution model is presented in Algorithm 2. The training set is assumed to have r ordered pairs. Each feature vector $\in$ *handoff_set* is required to be labeled by corresponding reliability index values. The class probability for each of the K classes is calculated using Eq. 16. The class which has the highest probability will become the designated class of the feature vector. The corresponding reliability index value of the designated class will become the reliability label of the wireless network. After assigning reliability labels to every feature vector $\in$ *handoff_set*, the feature vector, which corresponds to a wireless network that has the lowest reliability index value is selected for vertical handoff.

Algorithm 2 Handoff Technique Based on Sample Distribution Model

Let the training set contain r ordered pairs
for each feature vector **x** $\in$ *handoff_set* **do**
  for each class C_k do
  Calculate $P(C_k|x)$ by using Eq. 16.
  end for
The class which has the highest probability will become the designated class of **x**.
Label **x** with the reliability index value of the designated class.
end for

Select the feature vector $\in$ *handoff_set* which has the lowest reliability index value.

Perform the handoff to the wireless network corresponding to the selected feature vector.

5 Results and Discussions

5.1 Simulation Settings

The proposed vertical handoff scheme is implemented in Network Simulator 3. The simulation study involved 4 different types of wireless networks—GSM, UMTS, WiFi and WiMAX. The simulation parameter settings are presented in Table 1. For the ease of reference, the vertical handoff scheme using Bayesian Classification approach is referred as VBC, and the vertical handoff scheme using Sample Distribution model is referred as VSD.

5.2 Empirical Result Discussions

The first experiment analyzes the performance of VBC and VSD when the user specified bit rate parameter br_u is varied. The reliability index value predicted by VBC and VSD for that network to which the handoff was made is compared with the actual reliability index value for the same network observed during simulation.

Table 1 Simulation parameter setting

Parameters	Used values
Number of wireless networks in the training set	100
Number of wireless networks of each type in the training set	25
Message size	10–100 KB/s
Data rate	1–5 MB/s
Frequency band WiFi	2400 MHz
Frequency band GSM	900 MHz
Frequency band UMTS	2000 MHz
Frequency band WiMAX	3000 MHz
Monetary cost	1:0 to 5:0 cost units
RSSI	−20 to 0
Number of wireless networks available to perform handoff during test case	5 to 20

The actual reliability index of the network is calculated after making the handoff decision. The result of this analysis is illustrated in Fig. 1. Both VBC and VSD exhibit tight correlation between their predicted reliability index and actual reliability index. VSD slightly performs better, because it avoids assumptions on class distributions. The result of the same experiment wrt execution time for making the handoff decision is illustrated in Fig. 2. VSD suffers slightly in execution time, because the empirical distribution function has to be recalculated every time to make the handoff decision. Also, the execution time cost for making the final handoff decision is limited, because of the lower computational complexity involved in both the classification models.

The second experiment analyzes the performance of VBC and VSD when the user specified monetary cost parameter co_u is varied. The result of this analysis wrt reliability index value of the network to which handoff was performed is illustrated in Fig. 3. Again, VBC and VSD exhibit tight correlation between predicted reliability index value and actual reliability index value, mainly due to their effective design. The result of the same experiment wrt execution time for performing handoff is illustrated in Fig. 4. Again, VSD suffers slightly than VBC, and both VBC and VSD incur limited execution cost for the same reasons explained before.

The third experiment analyzes the performance of VBC and VSD when the user specified RSSI parameter $rssi_u$ is varied, and the result of this experiment wrt reliability index value of the network to which handoff was performed is illustrated in Fig. 5. Both VBC and VSD provide predicted reliability index values which are closer to the actual reliability index values for the same reasons explained before. The result of the same experiment wrt execution time for performing handoff is illustrated in Fig. 6, and similar results obtained before are seen.

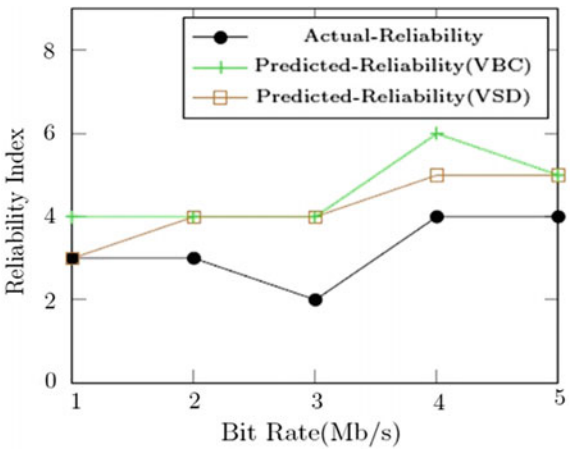


Fig. 1 Reliability versus br_u

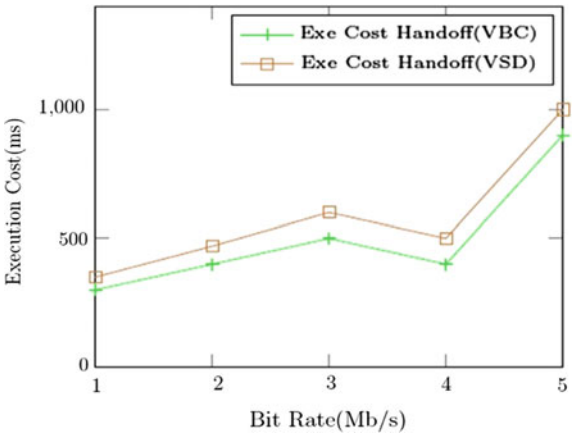


Fig. 2 Execution cost versus br_u

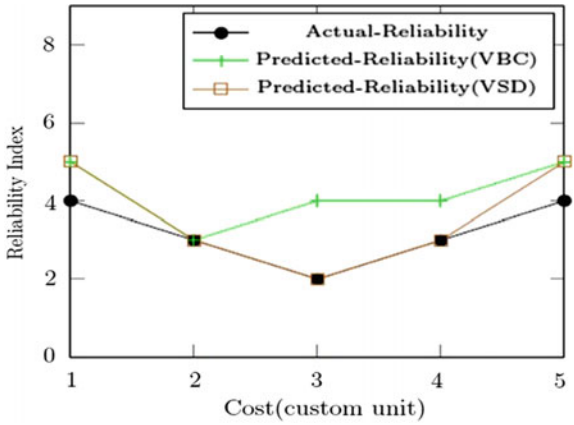


Fig. 3 Reliability versus co_u

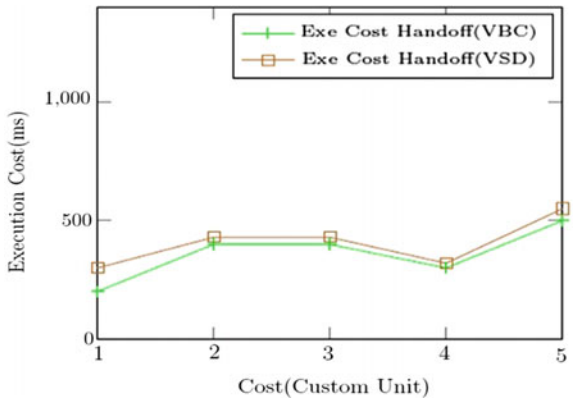


Fig. 4 Execution cost versus co_u

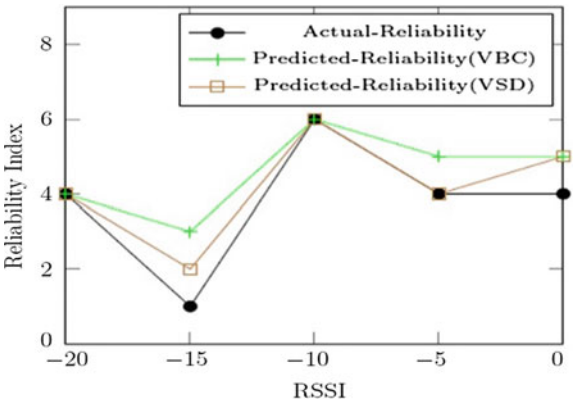


Fig. 5 Reliability versus $rssi_u$

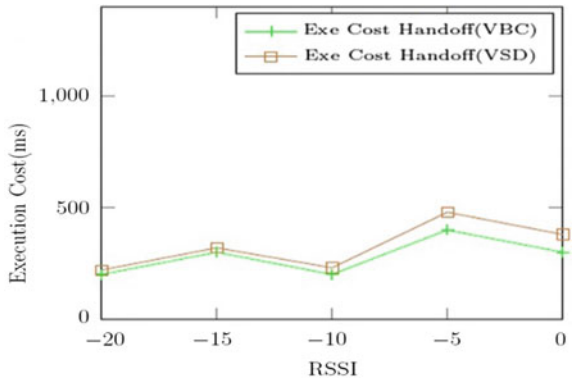


Fig. 6 Execution cost versus $rssi_u$

6 Conclusion

A new vertical handoff scheme which focused on reliability issue of handoff decisions was presented in this work. The proposed handoff scheme was built over 2 probabilistic classification models namely—Bayesian Classification model and Sample Distribution model. Empirical results demonstrated the effectiveness of the proposed vertical handoff scheme by making handoff decisions that achieve reliability. Also, the proposed handoff scheme incurs limited execution cost due to its computationally light design.

In future, there is a need to introduce flexible handoff scheme, wherein, the user is provided with the flexibility to vary the user specified parameters, and then decide the most suitable choice to make the handoff decision.

References

1. Xia L, Ling-ge J, Chen H, Hong-wei L (2008) An intelligent vertical handoff algorithm in heterogeneous wireless networks. In: Neural networks and signal processing, international conference, pp 550–555
2. Ling Y, Yi B, Zhu Q (2008) An improved vertical handoff decision algorithm for heterogeneous wireless networks. In: Wireless communications, networking and mobile computing, WiCOM 08, p 13
3. Guo Q, Zhu J, Xu X (2005) An adaptive multi-criteria vertical handoff decision algorithm for radio heterogeneous network. In: Communications, ICC 2005, IEEE international conference, pp. 2769–2773
4. Stoyanova M, Mahonen P (2007) Algorithmic approaches for vertical handoff in heterogeneous wireless environment. In: Wireless communications and networking conference, WCNC, pp 3780–3785
5. Nkansah-Gyekye Y, Agbinya JI (2008) A vertical handoff decision algorithm for next generation wireless networks. In: Third international conference on broadband communications, information technology and biomedical applications, pp 358–364
6. Bhattacharya PP (2007) Application of artificial neural network in cellular handoff management. In: Conference on computational intelligence and multimedia applications, international conference, 1, 237241. https://doi.org/10.1109/iccima.2007.252.1232414, Çalhan A, Çeken C
7. Nasser N, Guizani S, Al-Masri E (2007) Middleware vertical handoff manager: a neural network-based solution. In: Communications, ICC 07, IEEE international conference, pp 5671–5676. https://doi.org/10.1109/icc.2007.940
8. Onel T, Ersoy C, Cayrc E, Parr G (2004) A multi criteria handoff decision scheme for the next generation tactical communications systems. Comput Netw 46(5):695–708
9. Çalhan A, Çeken C (2010) An optimum vertical handoff decision algorithm based on adaptive fuzzy logic and genetic algorithm. Wirel Pers Commun. https://doi.org/10.1007/s11277-010-0210-6
10. Horrich S, Ben Jamaa S, Godlewski P (2007) Neural networks for adaptive vertical handover decision. In: Modeling and optimization in mobile, ad hoc and wire-less networks and workshops. WiOpt 2007. 5th International symposium on, p 17. https://doi.org/10.1109/wiopt.2007.4480068
11. Zayani R, Bouallegue R, Roviras D (2008) Levenberg-Marquardt learning neural network for adaptive pre-distortion for time-varying hpa with memory in ofdm systems. In: EUSIPCO 2008, 16th European signal processing conference
12. Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagation errors. Nature 323:533–536
13. Hagan MT, Menhaj MB (1994) Training feed forward network with the Marquardt algorithm. IEEE Trans Neural Netw 5(6):989–993
14. Levenberg K (1944) A method for the solution of certain nonlinear problems in least squares. Q Appl Math 2:164168
15. Marquardt DW (1963) An algorithm for least-squares estimation of nonlinear parameters. J Soc Ind Appl Math 11:431–441
16. Ceken C, Arslan H (2009) An adaptive fuzzy logic based vertical handoff decision algorithm for wireless heterogeneous networks. In: Wireless and microwave technology (WAMI) conference (WAMICON2009), p 19
17. Çalhan A, Çeken C (2012) Case study on handoff strategies for wireless overlay networks. Comput Stand Interfaces. https://doi.org/10.1016/j.csi.2012.06.002

18. Çalhan A, Çeken C (2010) An adaptive neuro-fuzzy based vertical handoff decision algorithm for wireless heterogeneous networks. In: The 21th personal, indoor and mobile radio conference, pp 2271–2276
19. Tripathi ND, Reed JH, Van Landingham HF (2001) Radio resource management in cellular systems. Kluwer, Dordrecht
20. Çalhan A, Çeken C (2013) Artificial neural network based vertical handoff algorithm for reducing handoff latency. Wirel Pers Commun. https://doi.org/10.1007/s112277-012-0944-4

Non-invasive Glucose Measurement Based on Ultrasonic Transducer and Near IR Spectrometer

Yanan Gao, Yukino Yamaoka, Yoshimitsu Nagao, Jiang Liu and Shigeru Shimamoto

Abstract This paper studies a noninvasive method to measure glucose level based on ultrasonic transducer and near infrared spectrometer. A series pair data of ultrasonic transducer from human finger, palm, wrist and arm are collected six times a day, and 16 spectral data of NIR spectrometer (reflection) from finger are collected by an OGTT experiment. The collected data are calibrated by using partial least squares regression and feed-forward back-propagation artificial neural network to predict the glucose level. In this study, error grid analysis is used to validate the prediction performance. In addition, the accuracy of the calibration models is improved.

Keywords Glucose measurement · Non-invasive · Ultrasonic transducer
Near infrared spectrometer · PLSR · BP-ANN

1 Introduction

Currently, the glucose is measured by pricking the fingertip and extracting the blood sample using a tiny disposable lancet, the extracted blood needs to be placed on one test sensor strip and inserted into the glucose meter that will show the numerical glucose value within 2 s. Although this method is the most painless way to measure glucose level, the patients are recommended to do the self-test at least 5–7 times per day depending on the diabetic type. So the main issue of the invasive finger pricking for glucose test are inconvenient, a little painful and high cost. Additionally, the lancet prick has a high risk of infection.

During the last decade, there have been many non-invasive methods to improve the glucose level measurement and the optical method is regarded as the main approach in most researches. The target of this article is to explore the relationship

Y. Gao (✉) · Y. Yamaoka · Y. Nagao · J. Liu · S. Shimamoto
Waseda University, Tokyo 1690072, Japan
e-mail: gaoyanan0926@fuji.waseda.jp

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_3

between glucose concentration and ultrasonic and near infrared light, try to predict the glucose level within acceptable range by doing the calibration scheme and validation.

2 Theory

Light attenuates due to the absorption and scattering of human tissue. The attenuation of light can be expressed as:

$$I = I_0 e^{-\mu_{eff} L} \tag{1}$$

which is the light transport theory. This equation describes light propagation in human tissue through a set of spectroscopic properties. And in this equation I is the reflected light intensity, I_0 is incident light intensity.

$$\mu_{eff} = \sqrt{3\mu_a(\mu_a + \mu_s')} \tag{2}$$

μ_{eff} is shown as Eq. (2) interacts with the absorption coefficient μ_a and reduced scattering coefficient μ_a', where $\mu_a' = \mu_s[1 - g]$. And the absorption coefficient can be described as the absorbance per unit path length equation that contains the molar absorption coefficient $\in$ and the molar concentration $C : 2.303 \in C\ \mathrm{cm}^{-1}$ [1]. The changes in glucose concentration can affect the light scattered intensity from human tissue. The higher glucose concentration, the more glucose molecule in the blood vessel. So it leads less scattering, less optical path and less absorption. It means the absorbance decreases with the increase of glucose concentration [2].

3 Experimental Method

3.1 Measurement Method

In this paper, both Ultrasonic transducer and Near infrared spectrometer are applied to blood glucose measurement.

3.1.1 Ultrasonic Transducer

Currently, the commonly used frequency range for ultrasonic diagnostic apparatus is 2–10 MHz, where 3–5 MHz has been widely used. In this experiment, a pair of 5 MHz Ultrasonic transducers with transmitter and receiver are selected to connect with the signal generator and signal analyzer.

Figure 1 shows the block diagram of the ultrasonic transducer experiment. There are two volunteers in this experiment, one male subject and one female subject. Those two ultrasonic transducers apply to four measurement targets: fingertip, palm, wrist and arm. Those targets are measured six times a day, 90 min before and after breakfast, lunch and dinner. Meanwhile the real glucose level is measured by traditional glucose meter as reference value. In this experiment, the room temperature is also considered to explore the influence for the relationship between glucose and ultrasonic.

3.1.2 Near Infrared Spectrometer

In this part, Near Infrared Spectrometer (AOFT: Acousto Optic Tunable Filter method) is applied to middle fingertip of left hand by reflection method. The device and experimental schematic are shown in Fig. 2. Wavelength range is 1200–2500 nm. This device includes the spectrometer and laptop where corresponding software is installed. The resolution is 1 nm, scan number is 4 times and integration number is 5 times.

There are two data pre-process methods, one is SNV (standard normal variate transformation) [3] processing that can eliminate solid particle size, surface scattering and the influence of NIR diffuse reflection spectrum from optical path difference. And another one is differential [4] calculation that can erase baseline and

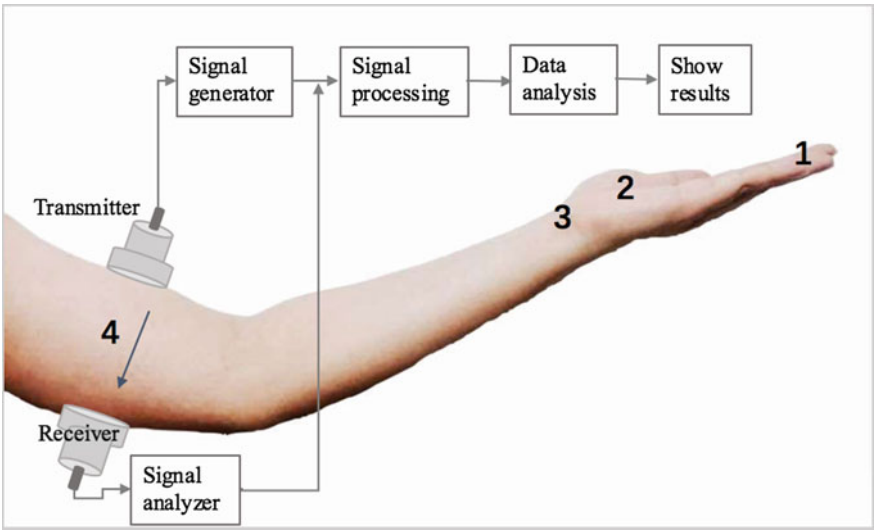


Fig. 1 The block diagram of the ultrasonic transducer experiment

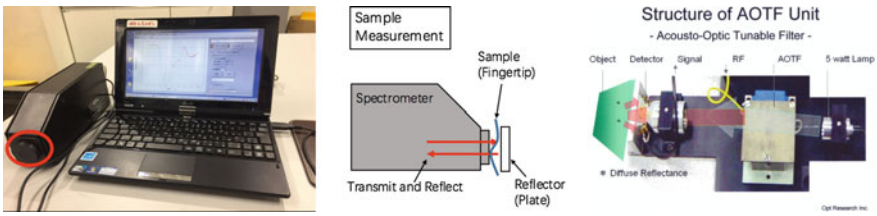


Fig. 2 The spectrometer device and experimental schematic representation (left) and the structure of AOFT unit in the spectrometer device (right)

other background interference, recognize overlapping peak, improve resolution and sensitivity effectively. And We also apply squalane oil to fingertip to erase some noise since the fingertip can absorb infrared easily.

In this experiment oral glucose tolerance test is used to help achieve the wide blood glucose level. Oral Glucose Tolerance Test: the subjects are fasting for more than 10 h and no breakfast before the experiment day. (1) Prepare the sugar water that 250 ml solution with 75 g of sugar. (2) Measure the baseline glucose level first and then smear Squalane oil on test fingertip. (3) After 5 min drink the 250 ml solution with 75 g of sugar. (4) And then start the near infrared experiment, save the absorbance spectral data and measure glucose level by glucose meter every 10 min for two hours, totally 16 times. And each time we collect 30 times spectral data to do the average calculation in order to reduce noise and improve signal-to-noise ratio.

3.2 Analysis Method

Partial Least Square regression method and back-propagation artificial neural network method are utilized in this article to analysis data. PLSR can find a linear regression model by projecting the predicted variables and the observable variables to a new space [5]. BP-ANN is another calibration model that applies to linear relations as well as nonlinear [6]. It through the output node feedback to hidden layer and input layer is used to adjust the weights value and then repeats that process until get the most ideal result. In this research those two methods can explore the codependent relationship among the volunteers' glucose from glucose meter and the corresponding experiment data to construct calibration model, and then we use this model to predict other observable variables.

In addition, Clarke's error grid analysis (EGA) [7] is used for the comparison tool. It can accurately represent the actual situation of the blood glucose measurement by showing results in five regions. So after the prediction, we need to use this method to compare the glucose level from glucose meter and the predicted result from the ultrasonic transducer and near infrared.

4 Result and Discussion

4.1 Ultrasonic Transducer

In ultrasonic transducer experiment, 122 data sets are collected in the ultrasonic experiment, this article takes the first 102 data sets to established prediction calibration model and the left 20 data sets to do validation that in order to test the performance of prediction scheme.

Figure 3 shows EGA result of two subjects in this experiment. The finger shows most of the predicted results fall in the region A, which means the established scheme is capable of making prediction. And the correlation coefficient R of subjects are 0.5419 and 0.4558, the RMSEP (root mean square error of prediction) of the glucose level are 12.1444 and 14.3531 mg/dl in PLSR method. While in BP-ANN method, R of subjects are 0.7332 and 0.6995, RMSEP are 9.3541 and 10.9674 mg/dl. In this situation, the accuracy of ANN method is better than PLSR method.

$$Y_1 = 2.1958\,x_1 + 0.2363\,x_2 - 0.4361\,x_3 - 0.6979\,x_4 - 2.9557\,x_5 + 85.1532 \quad (3)$$

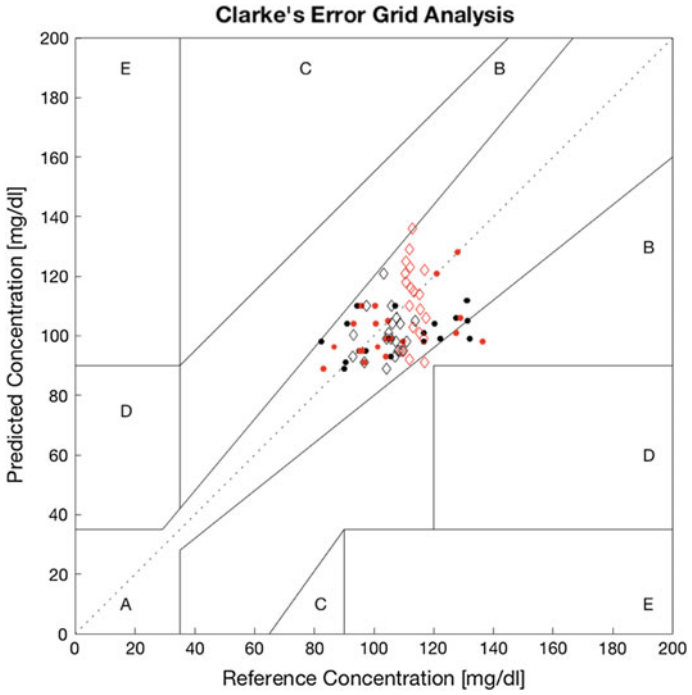


Fig. 3 The EGA result of two subjects from ultrasonic experiment by the PLSR method: circle and ANN diamond: star; female: red, male: black

$$Y_2 = 1.7565\,x_1 - 0.2551\,x_2 - 0.0910\,x_3 - 0.0688\,x_4 + 4.5107\,x_5 - 9.376 \quad (4)$$

In the equation, Y_1 is the predicted glucose value, x_1, x_2, x_3, x_4, x_5 are represent the output value of fingertip, palm, wrist, arm and room temperature respectively. By considering the metabolism rate with the fluctuation of the blood glucose, everyone should have their unique prediction equation to describe the glucose level. In this prediction, fingertip and temperature are more important than other three parameters, while the temperature makes negative influence on subject1 and positive influence on subject2 are opposite relation for two subjects.

4.2 Near Infrared Spectrometer

Figure 4 shows the five subjects' glucose range between 70 and 210 mg/dl are measured by glucose meter every 10 min from the OGTT and the absorbance spectrum of one subject's fingertip for 16 times from infrared spectrometer. In order to achieve the prediction of blood glucose concentration by analysis the characteristic absorption spectrum. There are some absorption wavelength areas of glucose, first area is 1450–1490 nm, second area is 1530–1570 nm and third area is 2130–2170 nm. They attribute to combination bands of –CO stretching vibration and –OH stretching and bending vibration of blood glucose.

This research takes the first eleven times spectral data sets and corresponding glucose concentrations as the predictor to establish the calibration model, takes left five data sets as the tester to validate the model performance in this experiment.

The PLSR and BP-ANN validation results of five subjects are showed in Fig. 5. The points with different colors and different shapes represent the results of five subjects from three wavelength areas in those two figures. All the results fall in the region A, which means they are acceptable scheme for the blood glucose prediction. And Table 1 shows the correlation coefficient R and RMSEP of five subjects in two different methods. By comparing the values, the prediction accuracy of ANN method is better than PLSR method, which is same as the ultrasonic transducer experiment.

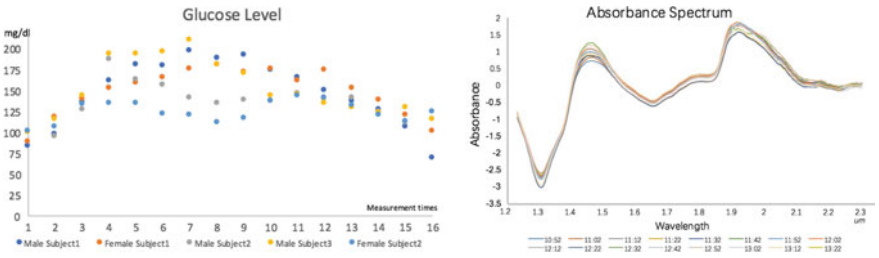


Fig. 4 The five subjects' glucose range from oral glucose tolerance test and the absorbance spectrum of subject's fingertip for 16 times

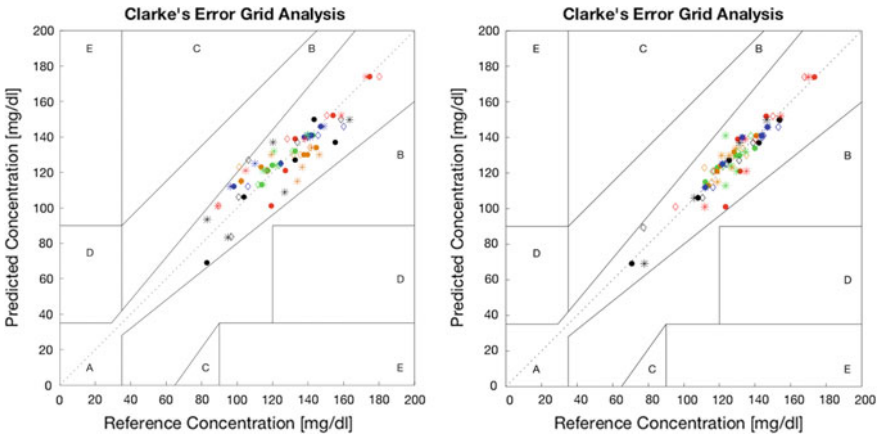


Fig. 5 The EGA result of five subjects from three characteristic spectrums by PLSR method (left) and ANN method (right), at first area: diamond; second area: circle; third area: star

Table 1 The results of calibration curve from the PLS regression and ANN method

Tester \ Area / Value (Mg/dl)		PLSR			ANN		
		First	Second	Third	First	Second	Third
Subject1	R	0.9929	0.9100	0.9836	0.9944	0.9589	0.9700
	RMSEP	8.7726	10.244	10.644	4.7942	9.500	8.4711
Subject2	R	0.8937	0.9435	0.7919	0.9805	0.9974	0.9976
	RMSEP	9.5874	10.370	10.533	7.909	3.4857	5.6635
Subject3	R	0.8299	0.9756	0.8233	0.9879	0.9763	0.8868
	RMSEP	6.3895	4.9964	11.251	2.5833	3.0257	6.6323
Subject4	R	0.9139	0.9807	0.7106	0.9858	0.9844	0.8626
	RMSEP	11.889	10.384	9.7377	3.2696	2.8307	6.0879
Subject5	R	0.9793	0.9647	0.9376	0.9853	0.9861	0.9923
	RMSEP	7.9751	7.1070	10.972	4.9674	4.2217	4.2217

5 Conclusion

In the paper a non-invasive glucose measurement scheme using ultrasonic transducer and Near IR spectrometer are explored. The predicted results from ultrasonic transducer experiment are less than the international standard ISO 151197 (International standard of SMGB the error is less than or equal to 15 mg/dl to comply with ISO 151197). And in the near infrared experiment, the results show

the feasibility of the development of non-invasive glucose measurement scheme based on reflectance through acousto-optic tunable filter. The prediction ability of ANN method is better than PLS regression method, especially in the nonlinear correlation between independent variable and dependent variable.

References

1. Yadav J, Rani A (2014) Near-infrared led based non-invasive blood glucose sensor. In Signal Processing and Integrated Networks (SPIN), pp 591–594
2. Amir O, Weinstein D (2007) Continuous noninvasive glucose monitoring technology based on occlusion spectroscopy
3. Rinnan A, van den Berg F (2009) Review of the most common pre-processing techniques for near-infrared spectra. Trends Anal Chem 28(10):1201–1222
4. Smilde A (2005) Multi-way analysis: applications in the chemical sciences. Wiley, Hoboken
5. Cheng J-H (2017) PLSR applied to NIR and HSI spectral data modeling to predict chemical properties of fish muscle". Food Eng Rev 9(1):36–49
6. Malik BA, Naqash A (2016) Backpropagation artificial neural network for determination of glucose concentration from near-infrared spectra, ICACCI, IEEE, pp 2688–2691
7. Clarke WL (2005) The original clarke error grid analysis. Diab Technol Ther 7(5):776–779

On Trust and Security Risk: Mobile Commerce Acceptance and Readiness in Sudan

Kennedy Njenga and Sara Salih

Abstract This paper looks at the mobile commerce readiness and presents an insightful look at Sudan and the rise of new opportunities for mobile technology use in the wake of emerging technology risks. Data elicited from Sudanese users of mobile applications was analysed with the objective of understanding and explaining the extent to which constructs such as trust and security risks has influenced Sudan's readiness for mobile commerce. The work applied the Unified Theory of Acceptance and Use of Technology (UTAUT) model towards this end. A sample size of 145 Sudanese respondents working in the public sector provided useful data that was subjected to the UTAUT approach. Findings suggest that Sudanese users' perceived technology risk and trust of m-commerce services in mobile platforms tended to inhibit the desired rate of m-commerce adoption and therefore readiness. These findings are seen to be in contrast of m-commerce readiness in Sudan's neighbouring eastern and central African countries such as Rwanda, Kenya and Ethiopia.

Keywords M-commerce · Information security · Sudan · Trust
Security risk · Sudanese public sector

1 Introduction

1.1 M-Commerce: Opportunities and Security Risk

The rise of mobile computing devices especially in Africa is growing and with the introduction of devices such as smartphones and tablets, endless commercial opportunities for mobile commerce (m-commerce) transactions has correspondingly grown [4, 11]. The growth of mobile technologies has seen a significant upsurge in m-commerce, with most Sub-Saharan African countries such as Rwanda, Kenya

K. Njenga (✉) · S. Salih
University of Johannesburg, Johannesburg, South Africa
e-mail: knjenga@uj.ac.za

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_4

and Ethiopia benefiting greatly from these developments [2]. However, Internet penetration, the primary driver for m-commerce success, remains relatively low in many other Sub-Saharan African countries with the exclusion of South Africa [6]. This has especially been the case in Sudan, where the significantly advanced American and European styles of Internet infrastructure models remain relatively unknown [11]. Sudan is seen as a unique country in Sub-Sahara Africa because it lacks advanced wireless mobile commerce infrastructure (such as *Amazon*™ *VISA*™ *Mastercard*™, and *PayPal*™) and commerce protocols such as SMS/USSD, WAP/WAP 2.0/Mobile Web and SIM Toolkit (STK) [13]. The use of m-commerce devices has been subject to many emerging information security vulnerabilities and security threats (such as malware, viruses, trojans, man-in-the-middle-attacks, denial-of-service-attacks and phishing to name but a few) that would influence Sudan m-commerce readiness. According to many studies, mobile devices are inherently vulnerable to platform and software application security risks [14] due to the use of WML Script that has been specifically developed for capability-limited Wireless Application Protocol (WAP) enabled devices. The purpose of this paper is therefore to understand Sudan's m-commerce readiness in regards to m-commerce in contest to how trusting the users would be in the face of technology risk.

2 Literature Review

M-commerce is still in its primal stage in Sudan since the country has poor information technology (IT) infrastructure. The ICT infrastructure includes 11,000 km of optic fibre national backbone, compressed by a digital microwave network, and domestic satellite system that covers more than 80% of the country [10]. This has made it difficult for Sudan to leverage the full benefits of m-commerce. Unlike other Sub-Saharan African countries such as Kenya, m-commerce and specifically *M-Pesa*™ and *M-shwari*™ initiatives are not only trusted services, but have made substantial inroads while m-commerce in Sudan lags behind [3, 5]. The main mobile network operators (MNOs) that currently dominate the Sudanese mobile market include: *Sudani*™; *Zain*™; and *MTN*™. These three operators use the global system for mobile communication (GSM) technologies [8].

2.1 M-Payment Services in Sudan

Sudan's use of technology is different from other Sub-Saharan countries [9]. Many commonly known technologies that other countries take for granted such as *Amazon*™, *GooglePlay*™ and other technologies that facilitate e-commerce/m-commerce were for a long period unavailable in Sudan [17]. Sudan's

m-commerce model follows a typical model with emergent new mobile payment services, such as *Gorooshi*™, a system supported by the *Sudani*™ MNO. *Gorooshi*™ allows *Sudani*™'s clients to transact via mobile phones at point of sales centres in banks and at other authorised agents across the country, similar to *M-Pesa*™ in Kenya. *Gorooshi*™, includes sub-services, such as cash-in and cash-out (e-wallet, to purchase electricity and water and pay all governmental fees). *E-wallet* allows customers to deposit and withdraw from their accounts through the Fisal Islamic Bank, sales points, and authorised agents. *Gorooshi*™, also provides money transfer services that allow users to transfer money to other subscribed and non-subscribed clients, while also enabling *Sudani*™ users to purchase electricity and water and pay all governmental fees [11]. The Sudan m-commerce model works like a typical model illustrated by Fig. 1 where a banking infrastructure and a mobile-telephony infrastructure is present.

2.2 *Information Security and Risk Concerns and Technology Acceptance*

The issue regarding m-commerce information security risk and online mobile payments has started becoming a concern to the Sudanese government. The reported security breaches and increasing threats of malware and viruses in mobile devices, the loss of privacy has had the effect of leaving Sudanese m-commerce users feeling vulnerable to these risks and no longer fully trust these systems [11]. Globally, the 2014 Kaspersky lab's Global IT Risk Survey findings show that with the advent of

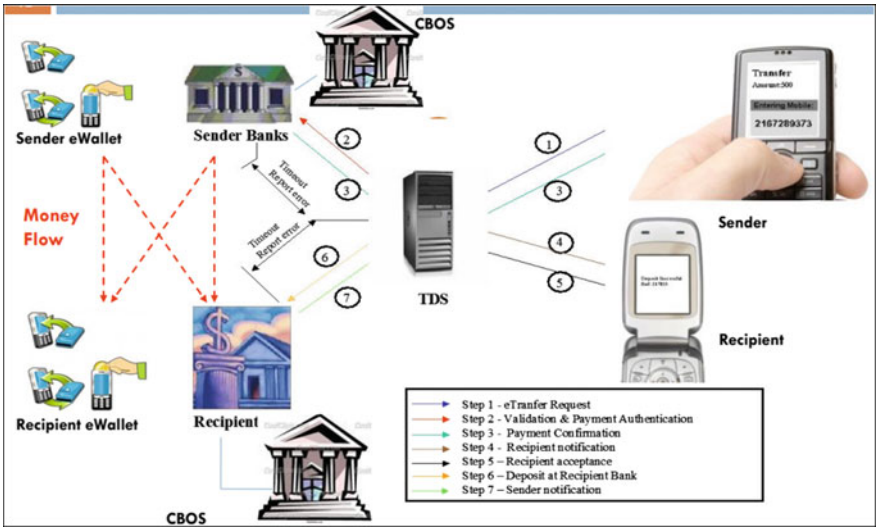


Fig. 1 Example of a representative mobile commerce model

mobile working, mobile threats are now being realised. According to this report; "*An unprotected mobile device provides access to sensitive data and gives cybercriminals an easy point of entry to an otherwise secure system*" [7]. These technology risks are generated by hackers, malicious software, disgruntled employees, competitors all of which pose a risk to the use of technology [7].

Studies have shown that many factors influence the acceptance of new technologies to users. We apply the Unified Theory of Acceptance and Use of Technology (UTAUT) to help explain Sudan's context and use four primary factors and four moderators that predict behaviour and intention to use technology [15]. The four factors are listed as performance expectancy, effort expectancy, social influence, and facilitating conditions, while the four moderators are listed as gender, age, experience, and voluntariness of use [15]. UTAUT postulates that performance expectancy, effort expectancy and social influence would influence behaviour intention to use technology while behaviour intentions and facilitating conditions determine technology use, as illustrated in Fig. 2 [15, 16]. We introduce the element of technology risk in our adopted and extended model because as Kaspersky Lab [7] point out, that while many business continue to integrate business processes with mobile devices, this has created security risk and threats to both businesses and to users. It is for this reason that this study decided to adopt UTAUT and extend its use as illustrated in Fig. 3.

Our adopted model therefore postulates that performance expectancy, effort expectancy, trust and perceived technology risk would influence behaviour intention to use m-commerce technology while behaviour intentions determine how m-commerce is actually used in Sudan. As such we developed the following hypothesis as follows (Table 1).

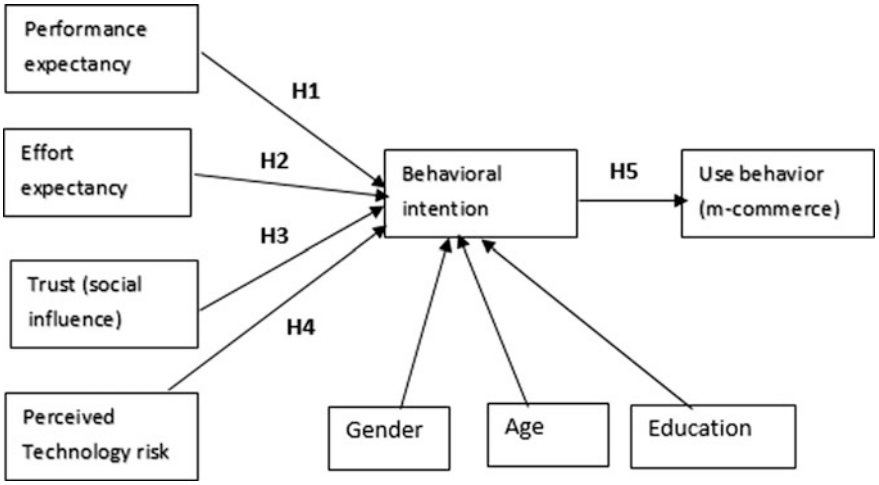


Fig. 2 Unified theory of acceptance and use of technology (UTAUT) [15]

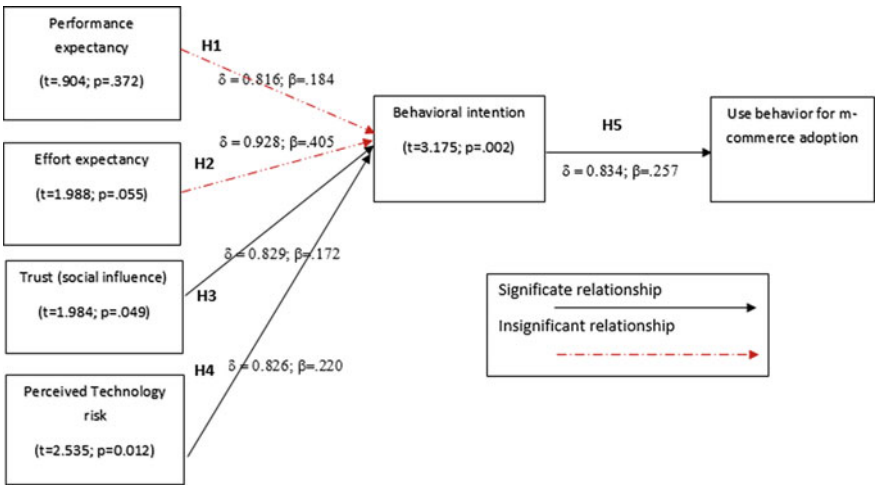


Fig. 3 Revised model

Table 1 Hypothesis development

Variable	Hypothesis
Performance expectancy	**H1**: Performance expectancy will influence the behavioural intention for m-commerce adoption by a user in the Sudanese government
Effort expectancy	**H2**: Effort expectancy will influence the behavioural intention for m-commerce adoption by a user in the Sudanese government
Trust	**H3**: Trust will influence the behavioural intention for m-commerce adoption by a user in the Sudanese government
Perceived technology risk	**H4**: Perceived technology risk will influence the behavioural intention for m-commerce adoption by a user in the Sudanese government
Behavioural intention	**H5**: Behavioural intention will influence use for m-commerce in the Sudanese government

3 Methodology

This study follows a quantitative design that adopts a positivist deductive approach to understanding and analysing data. A survey was carried out with the help of the questionnaire and a data elicitation instrument. The targeted population were Sudanese mobile users working with various Sundanese government departments since these were established as being most likely early adopters to m-commerce. The survey was distributed through online social media platforms (Facebook Sudanese). A total of 150 responses were collected, with five being rejected, and thus 145 responses were used for this study.

4 Data Analysis

To ensure data is good and reliable for further analysis, Cronbach's alpha must exceed 0.70. Using Cronbach's alpha analysis, the instrument's reliability was tested and results provided by Table 2 [12]. Cronbach alpha results ranged between 0.816 and 0.834, exceeding the 0.70 threshold.

4.1 Analysis Regression

Regression analysis is one of the useful methods of analysis when you have a dependent variable that is measured on a continuous scale, or an interval ratio when you have two or more independent variables or predictive variables to predict the dependent variable [12]. To predict whether or not the independent variables influence the m-commerce adoption in the Sudanese governmental sector, based on a series of other independent variables, a multiple regression analysis was applied.

4.1.1 Effort Expectancy and Performance Expectancy as Predictors of Behavioural Intention

Multiple regression tests employed to test the Effort Expectancy and Performance Expectancy, show the significance of the Effort Expectancy and Performance Expectancy in the model, with F = 7.402 and $p < 0.05$, which is significant as shown by Table 3.

Furthermore, Table 3 shows that both Effort Expectancy and Performance Expectancy ($P < 0.05$) have no significant effects on the behavioural intention to adopt mobile service for governmental transactions in Sudan. Neither of these variables played a significant role in the model while controlling for the effect of each other, this could be as a result of the violation of underlying assumptions. However, in a simple regression model using only Effort Expectancy as the predictor, it was highly significant. So the more the Effort Expectancy, the higher the Behavioural Intention (t = 3.750, $p = 0.001$).

Table 2 Test of reliability

Construct name	Cronbach's alpha	Mean	N of items
Performance expectancy	0.816	7.68	3
Effort expectancy	0.928	12.76	2
Trust	0.829	8.28	3
Perceived technology risk	0.826	8.63	4
Behavioural intention	0.834	10.59	4

Table 3 Effort expectancy and performance expectancy regression analysis

Model	Unstandardised coefficients		Standardised coefficients	t	Sig.	95.0% confidence interval for B	
	B	Std. error	Beta			Lower bound	Upper bound
Constant	1.533	0.722		2.122	0.041	0.065	3.000
Effort expectancy	0.356	0.179	0.405	1.988	0.055	−0.008	0.719
Performance expectancy	0.213	0.236	0.184	0.904	0.372	−0.266	0.693

4.1.2 Trust and Perceived Technology Risk as Predictors of behavioural intention

The R square result shows how much variance is described by the model in the independent variables [12]. In this study, the R square results indicate that trust and Perceived Technology Risk represent 0.109% of the factors that influence the Sudanese's intention to adopt m-commerce for governmental transactions.

In addition, Table 4 shows that trust and Perceived Technology Risk ($P < 0.05$) significantly influence the mobile users' intention to adopt m-commerce for governmental services in Sudan.

4.1.3 Behavioural Intention as a Predictor Performance Expectancy

Table 5 shows that Performance Expectancy ($P < 0.05$) (t = 3.175) significant at 0.002 has no strong effect on the behavioural intention to adopt mobile service for governmental transactions in Sudan.

Table 4 Trust and perceived risk regression analysis

Model	Unstandardised coefficients		Standardised coefficients	t	Sig.	95.0% confidence interval for B	
	B	Std. error	Beta			Lower bound	Upper bound
Constant	2.828	0.250		11.297	0.000	2.333	3.323
Trust	0.161	0.081	0.172	1.984	0.049	0.001	0.321
Perceived technology risk	0.202	0.080	0.220	2.535	0.012	0.044	0.359

Table 5 Behavioural intention and performance expectancy regression analysis

Model	Unstandardised coefficients		Standardised coefficients	t	Sig.
	B	Std. error	Beta		
Constant	3.658	0.097		37.776	0.000
behavioural intention	0.645	0.203	0.257	3.175	0.002

4.2 Testing Results

A multiple regression summary of constructs tested suggest that UTAUT was able to predict the factors that influence Sudanese users' potential to adopt m-commerce for governmental services. The summary of test results is presented in Table 6.

Figure 3 illustrates the revised model, which consists of five constructs, performance expectancy, effort expectancy, trust, perceived technology risk, and behavioural intention. The independent factors in this model indicate a 41.2% likelihood of m-commerce adoption in the Sudanese governmental sector.

5 Discussion

From our data analysis, we noted that study participants exhibited mistrust over new commercial m-commerce technologies. Trust was seen as an overriding reason for the lack of proper adoption of m-commerce in Sudan. Indeed, thirty-eight percent of study participants mentioned that that they did not trust the available services on offer via these portals. Another possible reason why m-commerce was

Table 6 Regression results

	Regression	Result
H1	t value = 0.904, $p < 0.001$	Performance expectancy will influence the behavioural intention for m-commerce adoption by a user in the Sudanese government **is rejected**
H2	t value = 1.988, $p < 0.001$	Effort expectancy will influence the behavioural intention for m-commerce adoption by a user in the Sudanese government **is rejected**
H3	t value = 1.984, $p < 0.001$	Trust will influence the behavioural intention for m-commerce adoption by a user in the Sudanese government **is supported**
H4	t value = 2.535, $p < 0.001$	Perceived technology risk will influence the behavioural intention for m-commerce adoption by a user in the Sudanese government **is supported**
H5	t value = 3.175, $p < 0.001$	Behavioural intention will influence use for m-commerce in the Sudanese government **is supported**

not readily accepted as envisaged was perceived technology risk. Perceived technology risk was considered an important factor influencing the adoption of m-commerce in Sudan. This could be explained as follows; firstly, m-commerce is relatively new to the Sudanese, and the Sudanese are still fearful of new technologies, particularly those that would involve money transfer. Sixteen percent of the study participants stated that they did not use m-commerce because they were not aware of the existence of such services in Sudan. Twenty percent of the respondents stated that dealing with these services was difficult, and that the interface was not efficient as they would have wanted. Some study participants indicated that they were used to carrying out financial transactions using computers and mistrusted the mobile interface. Taking cognisance of the above highlighted concerns it therefore followed that both effort expectancy performance expectancy were insignificant predictors of the Sudanese intention to use m-commerce. This observation contradicts prior studies done in Malaysia and Saudi Arabia [1] which showed effort expectancy and performance expectancy as important predictors to behavioural intention and use of technology. Our study therefore supports the notion that Sudanese users are not as confident in making financial transactions for governmental services using m-commerce technologies as initially envisaged because they do not trust these models and that these models are perceived to inherently risky. We attribute this to the historical conditions that have prevailed in the country.

6 Conclusion

We systematically used the classical UTAUT model to help us understand the prevailing use of and adoption of m-commerce technology in Sudan and of contributing to knowledge on this regard. We organised our work by firstly presenting the unique case of Sudan and its readiness to adopt to m-commerce use. We have presented from a literature review, the present state of Sudan as one which has in recent times started using m-commerce services such as *Gorooshi*™. Our quantitative approach to the study shows that Sudanese users' perceived technology risk and trust has had a greater bearing on adoption of m-commerce than other technology factors such as effort expectancy and performance expectancy. The Sudanese experience is in contrast to the more stable neighbouring eastern and central African countries such as Rwanda, Kenya and Ethiopia. This has in effect differentiated the m-commerce adoption experience compared to theses other African countries. We hope this work highlights a promising future of research on technology adoption, acceptance and use into more African countries where this kind of research is needed.

References

1. Algethmi MA (2014) Mobile commerce innovation in the airline sector: an investigation of mobile services acceptance in Saudi Arabia. Doctoral dissertation, Brunel University School of Engineering and Design Ph.D. Thesis
2. Ashraf AR, Thongpapanl NT, Menguc B, Northey G (2016, Oct) The role of m-commerce readiness in emerging and developed markets. American Marketing Association
3. Cook T, McKay C (2015) How M-Shwari works: the story so far. Consultative group to assist the poor (CGAP) and financial sector deepening (FSD)
4. Heinze J, Thomann M, Fischer P (2017) Ladders to m-commerce resistance: a qualitative means-end approach. Comput Hum Behav 73:362–374
5. Jack W, Suri T (2011) Mobile money: the economics of M-PESA (No. w16721). National Bureau of Economic Research
6. Joubert J, Van Belle J (2013) The role of trust and risk in mobile commerce adoption within South Africa. Int J Bus Humanit Technol 3(2):27–38
7. Kaspersky Lab. Global IT security risks: 2014 [Online]. Available at https://media.kaspersky.com/en/business-security/Global-IT-Risks-Report-2014-Threat-Security-Data-Breaches.pdf. Accessed 20 Feb 2018
8. Khattab I, Balola Y, Eldabi TA (2012) Factors influencing branchless banking for microfinance in Sudan: theoretical perspectives and future directions
9. Malik M, Malik M (2015) The efficacy of United States sanctions on the Republic of Sudan. J Georgetown Univ Qatar Middle East Stud Student Assoc 2015:7. https://doi.org/10.5339/messa.2015.7
10. Mohamed EH, Ping WANG, Elhadi OA (2014, June) E-commerce in Sudan: the reality and prospects. In 2014 international conference on management science and management innovation (MSMI 2014). Atlantis Press
11. Mohammed AMA (2015) A comparative study for mobile payment in Sudan case study: M-PESA in Kenya. Doctoral dissertation, Sudan University of Science and Technology
12. Pallant J (2016) SPSS survival manual: a step by step guide to data analysis using IBM SPSS, 6th edn. Allen and Unwin, Sydney
13. Prendergast J, Brad Brooks R (2016) Modernized sanctions for Sudan: unfinished business for the Obama Administration. The Enough Project, Washington, pp 8–10
14. Vanathi B, Shanmugam K, Uthairaj VR (2016) A secure m-commerce architecture for service provider to improvize quantity and quality of the products using fingerprint authentication and gender classification. Asian J Inf Technol 15(2):232–242
15. Venkatesh V, Morris MG, Davis GB, Davis FD (2003) User acceptance of information technology: toward a unified view. MIS Q 27(3):425–478
16. Venkatesh V, Thong JY, Xu X (2016) Unified theory of acceptance and use of technology: a synthesis and the road ahead. J Assoc Inf Syst 17(5):328–376
17. York J, Carlson K (2014) Sudan tech sanctions harm innovation and development: US government and corporations must act. Electronic Frontier Foundation. Retrieved 7 Aug 2017, from https://www.eff.org/deeplinks/2014/06/sudan-tech-sanctions-harm-innovation-development-us-government-and-corporations-must-act

An Adaptive Security Approach According to the Reliability Level of Drones Using Blockchain

Hyun-Koo Moon, Wooyeob Lee, Seungmin Kim, Jaewan Park, Jeongmin Lee and Inwhee Joe

Abstract In drone environments using batteries, it is important to reduce unnecessary overhead. Also, for security of data, encryption is also an important factor. However, in the case of encryption, overhead is generated, but encryption and overhead are always in a trade-off relationship in this way. In this paper, in order to reduce overhead of drone, the concept of reliability was applied. We applied low level security to drone with high reliability to reduce overhead and conversely applied high level security to drone with low reliability. In order to adjust the security level, adjust the size of Key used for encryption and reduce the weight of the security algorithm itself. The reliability of the drone was determined based on the number of times of receiving data and by applying group reliability through drone grouping, it was made possible to determine hierarchical reliability. Moreover, by storing the communication records of all drones in the blockchain, the integrity of the reliability is guaranteed.

Keywords Adaptive security · Reliability · Blockchain · Security level

H.-K. Moon · S. Kim · J. Park · J. Lee · I. Joe (✉)
Department of Computer Software, Hanyang University, Seoul, South Korea
e-mail: iwjoe@hanyang.ac.kr

H.-K. Moon
e-mail: moonhyunkoo@hanyang.ac.kr

S. Kim
e-mail: superhero@hanyang.ac.kr

J. Park
e-mail: xp0207@hanyang.ac.kr

J. Lee
e-mail: jmlee3096@hanyang.ac.kr

W. Lee
Department of Computer Science and Engineering, Hanyang University,
Seoul, South Korea
e-mail: matias12@hanyang.ac.kr

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_5

1 Introduction

As interest in drone increases recently, security of drone data is important. However, it is difficult for the drone environment to use the security technology applied as the wired network as it is because the characteristic of the drone is that it is difficult to operate without replacing the battery for a long period of time, the energy efficiency is more than anything. So we consider reliability of drone and blockchain in terms of energy efficiency. It is exactly the security level that greatly affects these energy efficiencies [1, 2].

Blockchain is the core technology used in virtual currencies such as bitcoin and are security technologies that sharing and contrasting the entire transaction ledgers of all traders and securing transactions [3].

In this paper, we propose the security scheme that determines the reliability of the drone using blockchain on an individual or group basis, adjusts the security level at the time of data transfer according to these reliability, and uses resources efficiently.

2 Reliability Determination Technique

All drones are managed in groups, and each group consists of a head drones and general drones. In addition, the group was not separated by GRS, and as a result, blockchain can be operated on a group basis.

As shown in Fig. 1, blockchain for managing the reliability of individual drones and blockchain for managing the reliability of each group are hierarchically operated.

Rules for measuring reliability are defined as Table 1.

In the blockchain, the communication record and the reliability value of the recipient are stored. Upon data transfer, the sender confirms the transaction in the blockchain, and when the reliability of the recipient is determined by all the drones, the sender transmits the data and generates a new transaction. Figure 2. shows a blockchain transaction involved in data transfer.

To transfer data in the same group drones use reliability in internal blockchain. In case of different group of drones, reliability is calculated follow equation.

$$X(\text{Reliability of Recipient}) = \frac{\text{Reliability of Recipient Group} * \text{Recipient's Current Reliability}}{\text{Reliability of SenderGroup}} \tag{1}$$

Drone usually runs within a group, but there are circumstances where you need to move to another group depending on the situation. At the time, the moving procedure of the group is as follows.

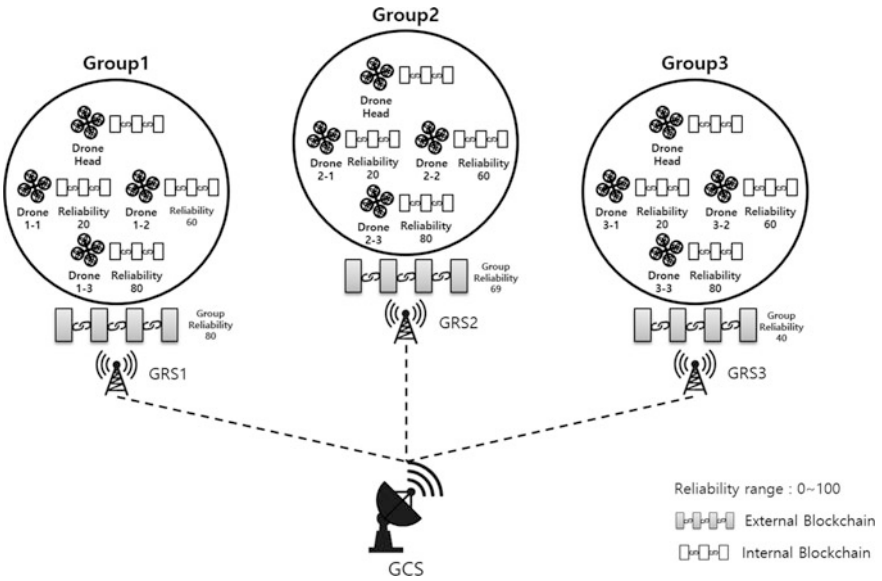


Fig. 1 Overall system configuration to determine the reliability of drones

Table 1 A table of rules for determining the reliability of drones

Rules	Contents	Remarks
1	The reliability of the drones registered in the initial group is set to 10	
2	The maximum reliability of the drones is set to 100 and the minimum reliability is set at 0	
3	The reliability of the drones (or groups) receiving the data increases by one (Per mission)	Prevent increasing attacker drones' reliability
4	The reliability of the drones (or groups) receiving the data decreases by one (Per mission)	Prevent increasing attacker drones' reliability
5	Drone heads do not have reliability	

1. Change the ID of new drone and delete block
2. Update the latest block from the drone head
3. Calculate new reliability values in groups

$$X(\text{New Reliability}) = \frac{\text{Reliability of Group before moving} * \text{Own Current Reliability}}{\text{Reliability of Group after moving}} \tag{2}$$

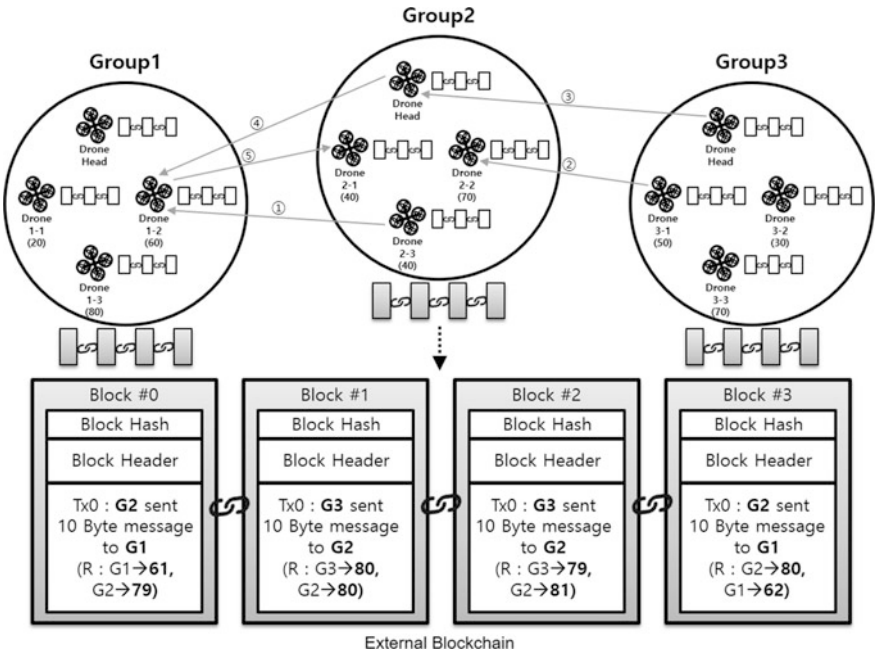


Fig. 2 Drone communication between groups through blockchain

3 Adaptive Security According to Drone Reliability

Based on reliability, the sender determines the security level. Security is applied to all data transfers and the highest security level applies to drones with reliability between 0 and 10. At the time, data authentication, integrity, confidentiality was applied. The lowest security level applies to drone whose reliability is between 41 and 100.

Based on the reliability determined in this manner, the security strength is adjusted based on the following Table 2.

4 Performance

We measured the time overhead of the security algorithm according to the reliability presented in this paper. The test environment was a Java 9 in Window 10 and Intel® Core i7-6700 CPU.

As Fig. 3, the more the key size is increased, the more the required time increases. Therefore, if we change the encryption key size and algorithm using reliability, we can see that the total required time when viewing the entire network can be shortened.

Table 2 Security adjustment based on reliability level

Reliability level	Security service	Algorithm	Key size	Group
0–10	Authentication/Signature/ Encryption	SHA-512/Public key (RSA)	2048	External
		SHA-256/Public key (RSA)	1024	Internal
11–20	Signature/Encryption	SHA-224/Public key (RSA)	1024	External
		SHA-1/Public Key (RSA)	512	Internal
21–30	Encryption	Block Cipher (AES)	128	–
31–40	Encryption	Block Cipher(DES)	64	–
41–50	Mess authentication	Keyed hash (HMAC)	128	–
51–100	User authentication	Hash (SHA-1)		Internal
		Keyed hash (HMAC)	128	External

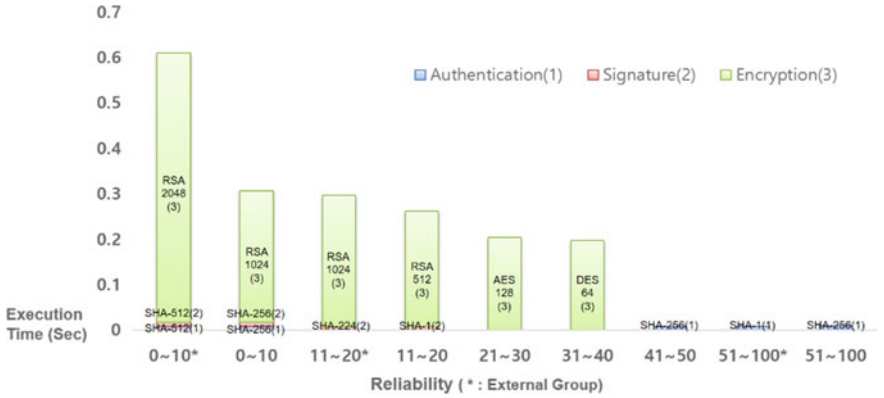


Fig. 3 Time overhead based on reliability evaluation

Based on the first graph in Fig. 4, it can be seen that as the drone with lower reliability evaluation is lower, the overhead change rate accompanying the message increase is larger.

The second graph in Fig. 4 compares the overheads of the drones to which the weight reduction mechanism presented in this paper is applied and the overhead of the drones not so. A drone that does not apply the weight saving mechanism always applies the highest level of security and transmits data, so that much overhead is generated correspondingly. On the other hand, the drone to which the weight reduction mechanism is applied is based on the evaluation of reliability, the security level is determined, and the overhead corresponding to this is applied to apply the weighted security technology, which can reduce the overhead.

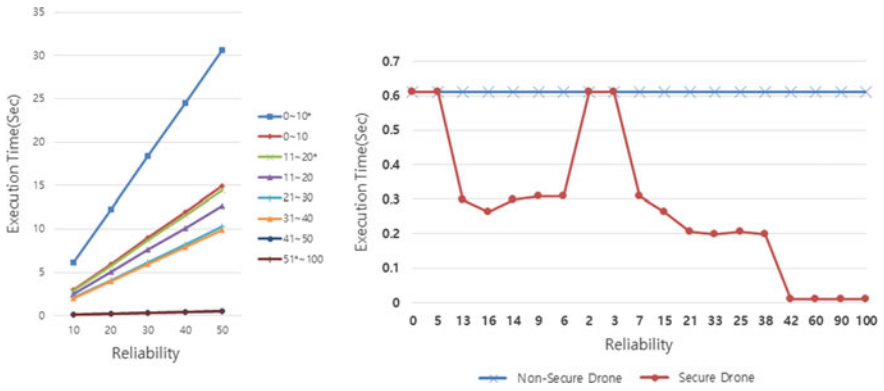


Fig. 4 Time overhead according to the number of messages and comparison between secure drones and general drones with reliability changes

5 Conclusion

In this paper, we aimed to reduce unnecessary overhead by assigning reliability value of drones, efficient key size adjustment considering the overall security level, and algorithm weight reduction. In addition, blockchain was used to ensure consistency of reliability values and to ensure hierarchical reliability management on a group basis.

By measuring the overhead of time according to the number of sent messages via the encryption algorithm and key size adjustment during reliability and comparing this with the drone with general security, the difference in time overhead Were compared.

In the future, depending on the network configuration and drone characteristics, add parameters that influence the reliability value, improve the accuracy of reliability determination, improve security techniques through various security algorithms and fine key size adjustment, we can use resources of the drone environment as more efficient.

Acknowledgements This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science, ICT & Future Planning) (NO. 2016R1A2B4013118).

References

1. Rault TF, Bouabdallah AS, Challal YT (2014) Energy efficiency in wireless sensor networks 67, 104–122
2. Cook KLBF (2014) The silent force multiplier: the history and role of UAVs in warfare, vol 4128, pp. 104–1158. Springer, Heidelberg

3. Nakamoto SF (2008) Bitcoin: a peer-to-peer electronic cash system
4. Aitzhan F, Zhumabekuly NS, Svetinovic DT (2016) Security and privacy in decentralized energy trading through multi-signatures, blockchain and anonymous messaging streams, pp 181–184. In: IEEE Transactions on Dependable and Secure Computing

Dynamic Key Management in Heterogeneous Wireless Sensor Networks Based on Residual Energy

S. Lavanya and M. Usha

Abstract Wireless sensor networks (WSNs) are mostly deployed in military, hospitals and other hostile environment and also used in building the Internet of Things applications where a huge number of devices is interconnected through the Internet. As these devices are heterogeneous in nature and are highly resource constrained devices, security issues constitute a major problem for their deployment. Among these issues, the key distribution plays a vital role and requires more research contributions. The proposed work takes the advantage of high-end sensors (H-sensors) which are equipped with tamper-resistant hardware in heterogeneous sensor networks (HSNs). The authenticity, confidentiality, and freshness of the message are guaranteed by using nonces and third-party key authorities. The comparison and theoretical analysis show that the proposed scheme requires less amount of storage space to store the keys. The proposed scheme is simulated using the widely accepted AVISPA (Automated Validation of Internet Security Protocols and Applications) tool and it is proved to be safe and secure. Also, this scheme provides better performance against node capture and replay attack.

Keywords WSNs · Heterogeneous · IoT · Clustering · Residual energy
Key management

1 Introduction

In the Internet of Things, the physical devices or things embedded with electronics, sensors, actuators, micro-electro mechanical systems, transportations, animals, etc., are ubiquitously integrated in order to provide intelligent applications for human beings. A wireless sensor network (WSNs) is one of the enabling technologies of

S. Lavanya (✉) · M. Usha
Sona College of Technology, Salem, India
e-mail: lavanya@sonatech.ac.in

M. Usha
e-mail: usha@sonatech.ac.in

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_6

IoT. In IoT, the large scale of WSNs is used to build the real-time applications. Basically, the wireless sensor networks are very large in size and they consist of sensor devices with severe energy constraint and limited memory. Due to this resource-constrained nature of WSNs, the security issues constitute a major problem in deploying those devices. Among these security issues, the key management is challenging because of the discrepancies in the existing cryptographic primitives [1].

Many security primitives have been designed for homogeneous wireless sensor networks which suffer from high energy constraints and memory overhead. In recent years, the heterogeneous sensor networks (HSNs) have been considered as an alternative topology which provides energy efficiency and security benefits. The heterogeneous sensor networks consist of two types of sensor nodes (SNs): a small number of powerful H-sensors and a large number of low power L-sensors.

The proposed work uses three levels of nodes such as common node (base station), cluster head and cluster node. It creates a dynamic key using random key generation method for secure communication between SNs. Any SN can acquire a pair-wise key when it wants to communicate with another SN. The routing information will be given by the routing protocol for enabling communication between SNs. One of the authors of this paper has already provided the research contribution towards the transport protocols [2] of WSNs and she is extending the work on routing protocols [3] of WSNs. The proposed scheme reduces the storage space required by the SNs. As a result, the communication and computation overhead will be reduced; in turn, the energy consumption [4] can be reduced. Also, it provides better security.

2 Related Work

The authors of [5] firstly published the probabilistic key pre-distribution scheme which assigns a random subset of keys from the key pool to each node before deployment and it uses a pair-wise common key for secured communication. The authors in [6] analyzed that pairing-based cryptographic techniques can be suitable for resource-constrained devices in WSNs. The key management scheme proposed in [7] used the cluster based WSN architecture. The authors used both private and public key cryptography to establish the secure communication link between sensor nodes and gateways. It improved the storage space and transmission overhead. Many approaches in the literature followed the pairing scheme. An asymmetric key pre-distribution technique for heterogeneous sensor networks was proposed [8] for pre-loading large number of keys in each powerful high-end sensors (H-sensors) and a small number of keys in each low-end sensors(L-sensors). The sensor nodes need to store number of messages to establish the shared keys. The work proposed in [9], uses bi-variate symmetric polynomial key pre-distribution method to preload the sensor nodes in three-tier hierarchical network architecture. It presented an improved key distribution method for large-scale wireless sensor networks.

The authors in [10] presented a unified framework for distributed key management mechanisms in HSNs. It developed the analytical model based on two key distribution schemes: (i) key-pool based key distribution scheme (ii) polynomial-pool based scheme. The proposed key management scheme proved that it provides connectivity, reliability, and resilience through the analytical model. The authors in [11] demonstrated that a probabilistic unbalanced key distribution mechanism not only provides the security but also reduces the consequences of compromising nodes in the network.

A key agreement protocol was proposed in [12] is based on pairing-based Elliptic Curve Cryptography (ECC). It uses pairing and identity-based encryption properties to compute the shared secret between any two nodes that need to communicate. The novel secure key management module was developed in [13] provides an efficient post-distribution key establishment in hierarchical clustering topology. The authors proved that it gives better resistance against replay and node capture attack. Another identity based key management (IBKM) scheme was proposed [14] to establish the pair-wise key [15] between any two nodes want to communicate in the network and cluster key to enable in-network processing in each cluster of nodes.

The authors of [16] proposed a dynamic and secure key management (DSKM) infrastructure for HSNs based on ECC and signcryption method (combination of forward security and public verifiability characteristics). This scheme prevents SN compromise using periodic authentication. The authors of [17] rely on the theory of random graph to build a secured connectivity in large HSNs. It uses heterogeneous structure and hash chains to broadcast messages in a secure channel. An efficient key management mechanism with implicit certificates approach used Elliptic Curve Cryptography [18], Diffie-Hellman key exchange and Qu-Vanstone implicit certificates. The authors of [19] proposed an efficient and hybrid key management (EHKM) scheme for HSNs, which uses the advantages of public key cryptography and symmetric key cryptography and provides better security, connectivity, and scalability.

3 Network Architecture

A large number of wireless sensor nodes (includes both H-sensor and L-sensor) are randomly distributed in an area. A sink/base station is located in a specified, well-protected place. As in Fig. 1, the whole large network is segregated into clusters using well-developed clustering algorithms. Each cluster consists of a few H-sensor nodes and several L-sensor nodes. One of the H-sensor nodes is elected as cluster head (CH), based on its energy as described in Sect. 3.1. Two layers are maintained in this architecture. The first layer is between the common node and cluster head. The second layer is between cluster head and cluster nodes (CNs). In every cluster, one H-sensor acts as a CH and the remaining H-sensors and L-sensors act as cluster nodes.

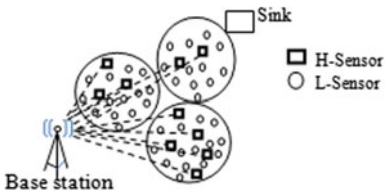


Fig. 1 Network architecture

3.1 Cluster Formation and Cluster Head Selection

H-sensor nodes in HSNs carry different initial energy. Initially, a node with the highest energy is elected as CH by the base station. If a particular node is fixed as a cluster head for a long time then it will die soon. This reduces the lifetime of the network. To overcome this shortcoming, a cluster head of dynamic adaptive scheduling scheme [20] can be used. After a period, another node from the list of H-sensor nodes available in the same cluster will be elected as CH, based on the value of T, given in Eq. (1).

$$T = \frac{E_{oi}}{\sum_{j=1}^{n-1} E_{oi}/n - 1} \tag{1}$$

E_{oi}—is the residual energy of ith H-sensor, n- is the total number of H-sensor nodes in one cluster $\sum E_{oj}$—indicates the sum of remaining H-sensor nodes' energy. If T > 1, then the node 'i' can be selected as the cluster head. At a regular interval, the base station calculates the threshold value using Eq. (2) and changes the CH accordingly. It also broadcasts the message about the new CH selection to the cluster nodes and assigns the corresponding secret key with them.

$$E_{threshold} = (1/n) * \sum_{i=1}^{n} E_{res}(i) \tag{2}$$

3.2 Assumptions

The heterogeneous wireless sensor networks can be grouped into clusters based on the following assumptions:

1. L-sensors are not equipped with tamper-resistant hardware [21]. There should be no trust assumptions on L-sensor and capabilities of the adversaries.
2. H-sensors are equipped with tamper-resistant hardware [21, 19]. These are capable of broadcasting the messages in a secure manner.

Table 1 Notations used in this paper

Symbol	Description
H	H-sensor
L1	L-sensor1
L2	L-sensor2
ID_{L1}	ID of L-sensor1
ID_{L2}	ID of L-sensor2
K_{L1H}	Shared secret key between L-sensor1 and H-sensor
K_{L2H}	Shared secret key between L-sensor2 and H-sensor
K_{L1L2}	Pair-wise key between L-sensor1 and L-sensor2
N_{L1}	Nonce generated by L-sensor1
N_{L2}	Nonce generated by L-sensor2

3. It is easy to form the clusters among the HSNs, and the secure communications can be established in every cluster.
4. Base stations are known as command nodes. These nodes are trusted by all of the nodes in the network [8] and are operated with intrusion detection system. The notations used in this paper are given in Table 1.

4 Proposed Key Management Scheme

Without the loss of generality, it is suspected that the IDs (32-bit) of all the nodes in the cluster are preloaded in the corresponding CH. Cluster nodes are assigned with its own ID. Each CH of a particular cluster is preloaded with IDs and secret keys of all the nodes in the same cluster [22, 23]. Also, it is preloaded with the key shared between the base station. Whenever there is a need for establishing a secure communication between any two nodes in the particular cluster, they require a pair-wise key. This key is dynamically generated by the third-party CH using random key generation method [24].

The proposed scheme involves five different message transactions between the nodes and pursues the following steps as in Fig. 2.

1. L-sensor1 sends a request message consists of ID of itself, ID_{L1}, ID of L-sensor2, $ID_{L2,}$ and a nonce, N_{L1}, encrypted by the shared secret key, $K_{L1H,}$ between H-sensor, to H-sensor which is being a CH.
2. H-sensor decrypts the message by using K_{L1H} and verifies the identity of L-sensor1. It dynamically generates the pair-wise key, K_{L1L2}, using random key generation method. Also, it encrypts the message which consists of the ID of L-sensor1, ID_{L1}, and a key K_{L1L2} using a shared secret key, K_{L2H}, between L-sensor2 and itself. Finally, it combines the above messages with nonce, N_{L1}, for authentication purpose and preventing the replay attack.

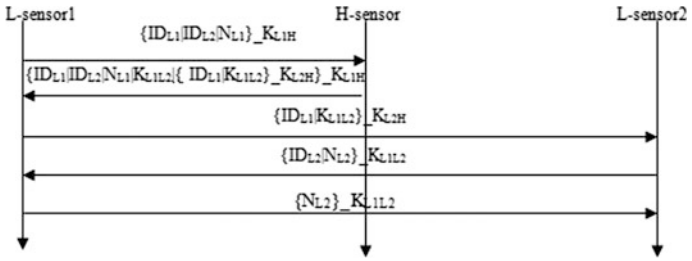


Fig. 2 Key distribution phase

3. L-sensor1 decrypts the message sent from H-sensor and sends the message, $\{ID_{L1}|K_{L1L2}\}\_K_{L2H}$ to L-sensor2.
4. L-sensor2 decrypts the message and ensures the presence of L-sensor1 and generates the nonce, N_{L2}. Then, it sends the ID of itself, ID_{L2}, and the nonce encrypted using a key, K_{L1L2}.
5. L-sensor1 sends the nonce, $N_{L2,}$ encrypted using a key, K_{L1L2}.

Now, both L-sensor1 and L-sensor2 can securely communicate with one another using the shared secret key, K_{L1L2}.

5 Storage Space Requirement and Comparative Analysis

Suppose the number of CH and CNs in the HSNs are M and N respectively. The number of preloaded keys in CH node is N + M. Each CH is preloaded with N/M secret keys shared with CNs and 1 secret key shared with base station. The number of keys stored in each L-sensor is 1. The total number of preloaded keys in HSN is 2 N + M. Figure 3 shows the comparison of storage space requirement between proposed scheme and other schemes where number of clusters = 10.

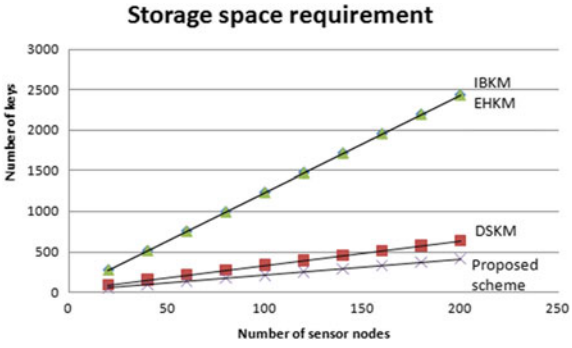


Fig. 3 Comparison of storage space requirement

Table 2 Comparative analysis

Performance evaluation metrics	IBKM [14]	DSKM [16]	EHKM [19]	Proposed scheme
Key establishment method	Identity-based key generation	Random key generation	Hashing+ identity-based key generation	Random key generation
Key management method	Asymmetric cryptography	ECC scheme+ signcryption cryptography	ECC scheme+ symmetric cryptography	Symmetric cryptography
# preloaded keys	M (N + 3) + 2 N	3 M + 3 N	M(N + 3) + 2 N	2 N + M
# keys in every L-sensor	2	3	2	1

The proposed scheme is qualitatively compared with the similar other schemes in the literature as shown in Table 2. It clearly shows that the proposed scheme reduces the storage space required by the nodes in the network and the number of keys to be stored in the L-sensor.

6 Simulation Results

The role specifications of L-sensor1, H-sensor, and L-sensor2 have been simulated in AVISPA tool under CL-AtSe model using HLPSL [4, 12] as shown in Figs. 4, 5 and 6 respectively. Also, these role specifications simulate the message transaction between those nodes. The result shown in Fig. 7 proves that the proposed scheme is safe and secure.

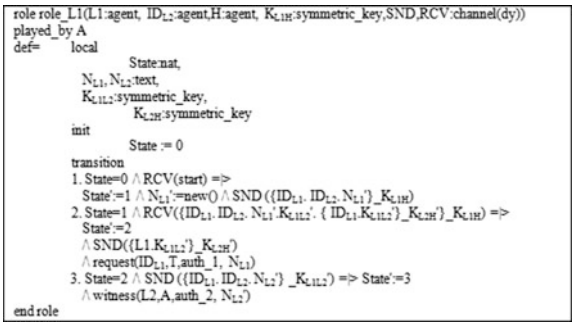


Fig. 4 Role specification of L-sensor1 in HLPSL

Fig. 5 Role specification of H-sensor in HLPSL

```
Role role_H(H:agent, ID_L1:agent, ID_L2:agent, K_L1H, K_L2H:symmetric_key, SND,
RCV:channel(dy))
played_by H
def=    local
                State:nat,N_L1:text,K_L1L2:symmetric_key
        init
                State := 0
        transition
                1. State=0 ∧ RCV({ID_L1. ID_L2. N_L1'}_K_L1H) =|>

                 State':=1 ∧ K_L1L2':=new() ∧ SND ({ID_L1.ID_L2.K_L1L2'.N_L1'.
                                                  {A.K_L1L2'}_K_L2H}_K_L1H)
                   ∧ secret(K_L1L2',sec_1,{ID_L1, ID_L2,H})
                   ∧ witness(ID_L1,T,auth_1,N_L1')
end role
```

Fig. 6 Role specification of L-sensor2 in HLPSL

```
role role_L2(ID_L2:agent, ID_L1:agent, H:agent, K_L2H:symmetric_key, SND, RCV: channel(dy))
played_by L2
def=    local
                State:nat, N_L1, N_L2:text,K_L1L2:symmetric_key
        init
                State := 0
        transition
                1.  State=0 ∧ RCV({A.K_L1L2'}_K_L2H) =|>
                    State':=1 ∧ N_L2':= new() ∧ SND({ID_L1. ID_L2. N_L2'}_K_L1L2')
                2. State=1 ∧ RCV ({ID_L1. ID_L2. N_L2'}_K_L1L2') =|> State':=2
                          ∧ request(ID_L1, ID_L2,auth_2, N_L2')
end role
```

Fig. 7 Simulation results for the CL-AtSe back-end

```
SUMMARY
  SAFE
DETAILS
  BOUNDED_NUMBER_OF_SESSIONS
  TYPED_MODEL
PROTOCOL
  /home/span/span/testsuite/results/New.if
GOAL
  As Specified
BACKEND
  CL-AtSe
STATISTICS
  Analysed    : 13 states
  Reachable   : 3 states
  Translation: 0.06 seconds
  Computation: 0.00 seconds
```

7 Conclusion

The proposed dynamic key management scheme takes the advantage of H-sensor in HSNs. The cluster head can be changed at a regular interval based on the residual energy of the H-sensors present in the same cluster. The proposed scheme used shared secret key and pair-wise key for enabling secure communication in HSNs. The comparative analysis shows that the proposed scheme uses the small number of keys to support the large size of the network. Moreover, the proposed scheme has been simulated using AVISPA tool and proved that it is safe under CL-AtSe model. Also, it provides the security against node capture and replay attacks. In future, the research contributions will be made on scalability and key revocation schemes.

References

1. Lavanya S (2017) A survey on key management in internet of things. In: Proceedings of international conference on intelligent computer systems, pp 572–580, SSRN
2. Sridevi S (2011) Taxonomy of transport protocols for wireless sensor networks. In: International conference on recent trends in information technology, pp 467–472, IEEE
3. Archana KT (2017) Amplified trickle algorithm in routing protocol for low power. In: Proceedings of international conference on intelligent computer systems, pp 401–413, SSRN
4. Sridevi S (2014) Energy-aware QoS based routing protocols for heterogeneous WSNs—a survey. Int J Comput Sci Bus Inform 11(1)
5. Eschenauer L (2002) A key-management scheme for distributed sensor networks. In: Proceedings of 9th ACM conference on computer communications security, pp 41–47, ACM
6. Oliveira LB (2007) TinyTate: computing the tate pairing in resource-constrained sensor nodes. In: Proceedings of 6th IEEE international symposium on network computer applications, NCA 2007, pp 318–323, IEEE
7. Azarderakhsh R (2008) A key management scheme for cluster based wireless sensor networks. In: International conference on embedded ubiquitous computing, pp 222–227, IEEE
8. Du X (2007) An effective key management scheme for heterogeneous sensor networks. Ad Hoc Netw 5(1):24–34
9. Cheng Y (2007) An improved key distribution mechanism for large-scale hierarchical wireless sensor networks. Ad Hoc Netw 5(1):35–48
10. Guizani M (2008) A framework for a distributed key management scheme in heterogeneous wireless sensor networks. IEEE Trans Wirel Commun 7(2):639–647
11. Traynor P (2006) Establishing pair-wise keys in heterogeneous sensor networks. In: Proceedings of 25th IEEE international conference on computer communications, pp 1–12, IEEE
12. Rahman SM (2010) Private key agreement and secure communication for heterogeneous sensor networks. J Parallel Distrib Comput 70(8):858–870
13. Gawdan IS (2011) A novel secure key management module for hierarchical clustering wireless sensor networks. In: 3rd international conference on computational intelligence, modelling & simulation, pp 312–316, IEEE
14. Boujelben M (2011) IKM—an identity based key management scheme for heterogeneous sensor networks. J Commun 6(2):185–197
15. Boujelben M (2009) Establishing pairwise keys in heterogeneous two-tiered wireless sensor networks. Proceedings—2009 3rd international conference on sensor technologies and applications, pp 442–448, IEEE
16. Alagheband MR (2012) Dynamic and secure key management model for hierarchical heterogeneous sensor networks. J IET Inform Secur 6(4), 271–280
17. Zhao H (2012) Hashed random key pre-distribution scheme for large heterogeneous sensor networks. In: IEEE 11th international conference on trust, security and privacy in computing and communications, pp 706–713, IEEE
18. Sciancalepore S (2015) Key management protocol with implicit certificates for IoT systems. In: Proceedings of the 2015 workshop on IoT challenges mobile and industrial systems—IoT-Sys'15, pp 37–42, ACM
19. Ying Z (2014) An efficient and hybrid key management for heterogeneous wireless sensor networks. In: 26th Chinese control and decision conference, pp 1881–1885, IEEE
20. Xuegong Q (2010) A control algorithm based on double cluster-head for heterogeneous wireless sensor network. In: 2nd international conference on industrial and information systems, IEEE, vol 4, no 1, pp 3–6
21. Banihashemian S (2010) A new key management scheme in heterogeneous wireless sensor networks. In: International conference on advanced communication technology, pp 141–146, IEEE

22. Akhbarifar S (2014) A survey on key pre-distribution schemes for security in wireless sensor networks. Int J Comput Netw Commun Sec 2(12):423–442
23. Xiao Y (2007) A survey of key management schemes in wireless sensor networks. Comput Commun 30(11):2314–2341
24. Zhang J (2010) Wireless sensor network key management survey and taxonomy. J Netw Comput Appl 33(2):63–75

Memory Management Strategy for PCM-Based IoT Cloud Server

Tae Hoon Noh and Se Jin Kwon

Abstract Most large-scale data server systems are having difficulties applying modern data usage patterns to such systems because recent data request patterns of users are sequential, and users tend to request up-to-date data. In this regard, customized systems are necessary for handling such requests efficiently. This paper deals with issues related to how conventional large-scale data server systems utilize memory, and how data are stored in storage devices. In addition, the paper analyzes data usage patterns of users, utilizing a cold storage system, and proposes a main memory system based on the analysis. This paper proposes a hybrid main memory system that utilizes DRAM and phase change memory (PCM). PCM is regarded as the next generation of non-volatile memory. Using a main memory that utilizes PCM, which operates similar to DRAM, and non-volatile storage, the proposed system improves the data processing efficiency. The paper also proposes an algorithm for processing data with the use of DRAM as a buffer. In addition, the paper proposes a system architecture with a tree-type block data and hash-type data block link. Moreover, this study compares the performance of an existing system with that of the proposed system using sequential and random data workloads. The results of the comparison show that performance improves by 10% when using a sequential data load, and remains almost at the same level when using a random data workload.

Keywords Cache storage · Control engineering computing · Non-volatile memory

T. H. Noh
UxFactory, Seongnam 13474, Republic of Korea

S. J. Kwon (✉)
Department of Computer Engineering, Kangwon National University,
Samcheok 25806, Republic of Korea
e-mail: sjkwon@kangwon.ac.kr

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_7

1 Introduction

Large and small data severs have emerged in order to provide personalized IoT services. Most services have been changed from offline to online, and many companies have concentrated on storing and analyzing the data patterns [1]. As the value of IoT depends upon the data quality, the personalized IoT services require a large-scale system model that deals with computer network, embedded devices, servers, and data analysis. Among various areas in the IoT, the researches related to storing and managing the data in a large-scale server have become very important issue, as the real-time online services involve fast access, efficient space management, and server power management.

Unfortunately, current IoT based data large-scale servers are not much different from the previous conventional servers. Although IoT based large-scale servers need to provide fast access, space management, and power management, the conventional servers use a hard disk drive (HDD) and dynamic random access memory (DRAM) for the main memory [2]. However, despite its long life span, this type of system has many drawbacks, including a low speed of the read-write operations, a large volume, heat, and noise. To solve these shortcomings, flash memory was developed, which has replaced HDD as a storage device. Flash memory can carry out read, write, and erase operations quickly in an electronic manner, resolving the problems of heat and noise [3]. Based on the merits of DRAM and flash memory, researches into new data server systems have been carried out. A data server consisting solely of flash memory provides a ten-fold increase in performance compared with HDD. In addition, flash memory has great scalability, and incurs no difficulty in cooling the system or in the power management. It also maximizes the efficiency of the performance and energy effect [4].

All flash-based IoT cloud servers provide ten times faster speed than HDD-based servers. Accordingly, large-scale storage server systems that depend on flash memory are widely used; nonetheless, there are issues regarding the life span of the storage owing to the characteristics of flash memory, and many researches into such issues have been conducted. Most recent server systems follow the Cold Storage System, in which the amount of unused data is four- to five-times greater than the data actually used. Furthermore, the occurrence of sequential read and write operations tends to take place owing to the large usage of image and video data [5, 6].

When many read and write operations are requested, they are conducted by accessing storage from the main memory in existing large-scale server systems. However, this method may cause the performance degradation depending upon the data request pattern. Most large-scale server system offer all data services equally. However, current systems are inefficient compared to an actual environment, which has patterns of user requests that are focused on the operations of recent data [7, 8]. Consequently, in this paper, we propose architecture for an efficient large-scale server system that processes the data mainly using data request patterns.

In this paper, we deal with two issues: (1) a system architecture for enlarging the storage life span and (2) efficient utilization of memory to improve the response

time of the server. To deal with these issues, we propose a Hybrid Main Memory (DRAM + PRAM) system using phase change memory (PCM). This system can improve the performance of the server response and enlarge the life span of the storage. In Sect. 2, we describe the utilization of the main memory, DRAM and flash memory, which are currently in use. In addition, we provide an explanation regarding the existing hybrid main memory and next-generation memory, which will eventually be commercialized. In Sect. 3, we provide an explanation of the proposed hybrid main memory, thereby suggesting a system that enlarges the life span of the storage and efficiently manages the server. In Sect. 4, we evaluate the performance of the proposed system, and in Sect. 5, we provide some concluding remarks and a description of further research.

2 Background

2.1 Phase Change Memory

PCM is a next-generation non-volatile memory that is able to access data in a byte-addressable manner. Table 1 shows the performance of DRAM, flash memory, and PCM. A PCM exhibit lower read, write, and erase operational speeds than flash memory and has limited endurance (10^7 to 10^8 in one memory cell on average). This makes it hard to compose the main memory using only PCM. Flash memory has a limited endurance, but includes a controller inside the device and utilizes a Flash Translation Layer (FTL). Wear-leveling, which depends on the operating system, is not suitable for PCM, because it was designed based on DRAM.

To overcome these drawbacks, PCM requires the wear-leveling technique itself. Owing to these characteristics, researches into hybrid main memory, which uses PCM and DRAM as the main memory, are being conducted. In recent researches, hardware techniques such as Start-gap, which leverages the endurance from the WL-Reviver Model [9] and wear-leveling, which stores the data log page in PCM [10] and software techniques [11] that assure the durability and consistency of PCM have been undertaken. However, we propose new applications for PCM because it depends on the particular model, which uses a cold storage system for a large-scale storage server.

Table 1 DRAM, PCM, and flash memory

Attribute	DRAM	PCM	Flash memory
Non-volatility	No	Yes	Yes
Idle power	~100 mW/GB	~1 mW/GB	~10 mW/GB
Bandwidth	~GB/s	50–100 MB/s	5–40 MB/s
Write latency	20–50 ns	~1 μs	~500 μs
Erase cycles	∞	10^7–10^8	10^4–10^5

2.2 DRAM + Flash Memory System

A large-scale server system has a structure similar to that shown in Fig. 1. The server system receives read, write, and erase operation requests through an I/O controller or network adapter, and these instructions are initially sent to the CPU. Next, the executed instructions are sent to the main memory, and the data operations are conducted from the main memory to the storage [12].

A representative solution to managing the main memory in a page unit is the employment of a page table. A page table can be realized using hierarchical paging, a hashed page table, an inverted page table, and so forth. Among these, a hashed page table has been applied to many server systems for the management of the main memory [13].

The design of a main memory hash table with DRAM uses standard chained hashing, which utilizes a set of keys, values, or linear chaining. However, these hashing solutions bring about a large overhead and easy incur hash collisions, and are therefore not suitable for application to memory caching systems such as Memcached and the recently used fundamental NoSQL system [14]. To do so, we made a main memory system using advanced cuckoo hashing and hopscotch hashing. However, using this system can be a problem owing to the endurance and asymmetric speed of the read and write operations of PCM. Hence, we simply supplemented the system architecture for PCM, and utilized it in the present study.

3 Hybrid DRAM + PCM Architecture

Before looking at the proposed hybrid DRAM PCM architecture, we should define some relevant problems. (1) We constructed an efficient system by using the system environment of the cold storage system of a large-scale storage server (which can be

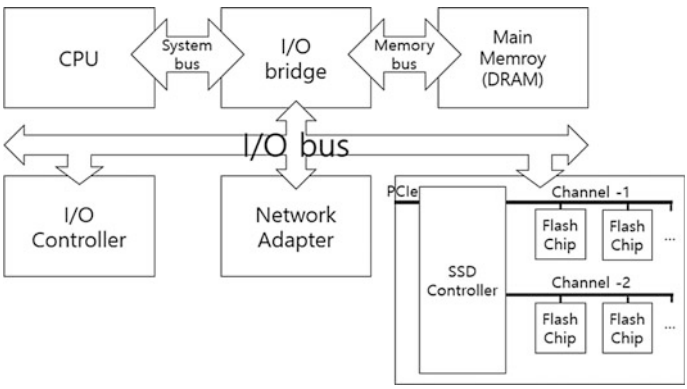


Fig. 1 DRAM based data storage

divided into hot-cold data, and accounts for sequential data). (2) We maximized the use of storage comprising flash memory. The architecture proposed in this paper also does not regard the existing file system, memory caching, or the FTL of flash memory.

3.1 Overall Architecture

Figure 2a shows the proposed large-scale server system, and Fig. 2b shows the internal architecture of the hybrid DRAM PCM. The main memory can be divided into DRAM and both DRAM and PCM. The former has the basic role of main memory. Our suggested architecture is a hybrid that uses both DRAM and PCM, and uses a strategy with two types of buffers: one stores and uses data, and the other processes the former one and sort the data. PCM stores and handles the data from the DRAM buffer. Similar to the DRAM buffer, data are handled in a block unit, and the data are managed in a block unit using a Hot Data Queue. The other components are operated on an existing system basis.

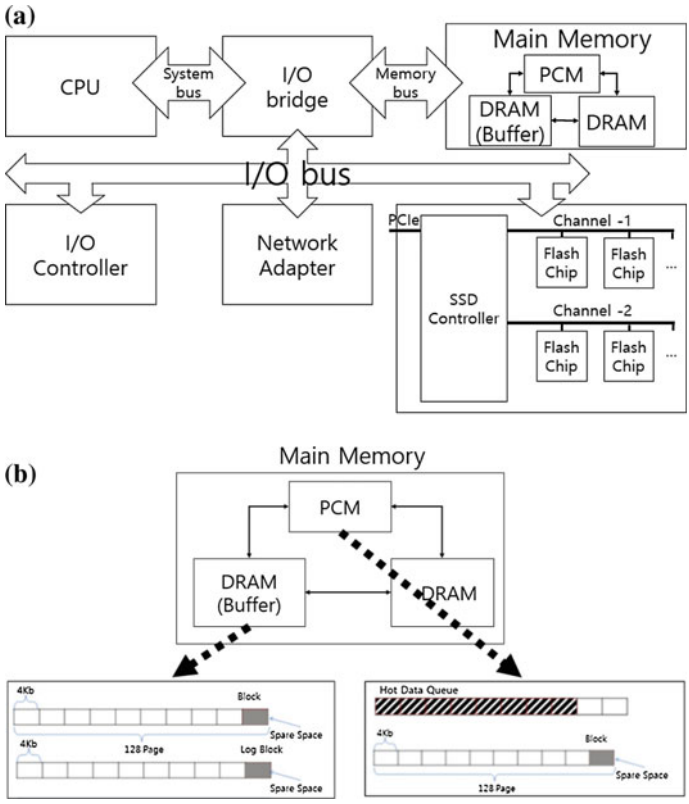


Fig. 2 PCM based data storage

3.2 Main Memory Operation

- Main Memory Operation: This system conducts the memory operation in a page unit, processes it, and composes a block with 128 pages. The reason why the data are composed using a block unit is to prolong the lifespan of the device and minimize the endurance of the PCM by reducing the write operations by processing the data in a page unit through the use of DRAM. Figure 2b shows that each page of a DRAM buffer is 4 KB, and each block has 127 pages and one spare space page. A total of 127 pages manage the data, and a spare space page is used for storing the information of the current block to improve the read and write operations, which are in a block unit.
- Main Memory PCM: PCM can be divided into two components: one is in a block unit that stores data, and the other is a data queue that manages the hot data. The hot data queue composes the data queue in the order of blocks inserted from the main memory buffer, and is an important index for a merge operation for existing block data and writing the data to flash memory.
- DRAM Buffer Write Operation: The DRAM buffer is divided into a block buffer and log block buffer. The block buffer manages the data requested write operations, and the log block buffer sorts and stores the data. The data buffer sequentially stores the write operations requested by a user, at the same time stores them in a binary tree based buffer. The reason for storing the data in a binary-tree based buffer is to be able to access the data more quickly than when using page-based access. Through the process of binary-tree based access, the address is recorded into a table that interconnects the requested logical page number (LPN) and physical page number (PPN). The first and last LPNs are recorded in a spare area of the log block, and whether the state of the block is full or not is recorded. A buff mapping table is stored here as well.
- PCM Write Operation: PCM comprises a hot data queue, mapping table, and data block sent to DRAM. The mapping table is constructed using a chain hash, and a PBN is made based on the last value of the LPN recorded in the spare space of each data block. The process of storing the data block sent from DRAM to PCM is as follows.

(a) A data block is received from DRAM.
(b) Whether a PBN referring to the mapping table exists is confirmed.
(c) If there is an empty PBN in the mapping table, the data block stores the PBN and records it to the mapping table.
(d) If a PBN is in the mapping table, it is chained to the preliminary data block and recorded in the mapping table.
(e) The data block from DRAM is compared with the existing PBN, the same LPN in a spare space is checked, and the update is confirmed.
(f) If the hot data queue is full, the data block from DRAM is stored in flash storage.

4 Performance Evaluation

This paper suggests the efficient utilization of memory to prolong the lifespan of the storage and improve the response time of the server in a large-scale server system. We evaluated the performance in a large-scale server system with hybrid main memory. The evaluation can be divided into two parts. First, we compared the endurance between the flash memory in DRAM + flash storage and in our suggested environment. Next, we analyzed and compared the utilization between the two systems. Finally, we analyzed the additional effects based on the above statistics.

The environment of the target evaluation is as follows. We used five types of workload for a performance evaluation. To apply the system theory of cold storage, which is the basis of the present paper, we evaluated the performance using sequential data, images, web data, and other elements. Then, for a comparison, we used the workloads, which are mainly composed of the random data.

Figure 3 shows a comparison with a write-operation applied to a conventional large-scale server system that uses a trace of sequential and random data, and to the hybrid PCM + DRAM system. Figure 3a shows the sequential data of a real image. It indicates that the data are very sequential and have an overlapped data pattern of 45.5%. As a result of applying each system, we can see that the write operations are similar to each other because the majority of data can be used by loading entirely into memory owing to the smaller data than the other workload. In this way, we can see that, in the case of the hybrid system, the write operations occur less frequently than those of a conventional system for image data.

Figure 3b and c show that the hybrid system has a better performance capability when a sequential data pattern is used and a hot-cold system is applied. This is because the data have a sequential pattern, and an overlapped data pattern of more than 90%. This shows the strength of the hybrid system, which processes the overlapped data by storing them in PCM.

Figure 3d and e show the results of applying a random data pattern to each system. For little available data, we process a general system such as a mail or authentication system, and conventional systems show a level of performance twice that of the suggested hybrid system. However, as the available memory increases, the gap between them decreases because, as the amount of available data that can be stored in PCM increases, the PCM can store data. In this way, we assume that if the PCM can store more data, it can perform as well as a conventional system for a random data pattern.

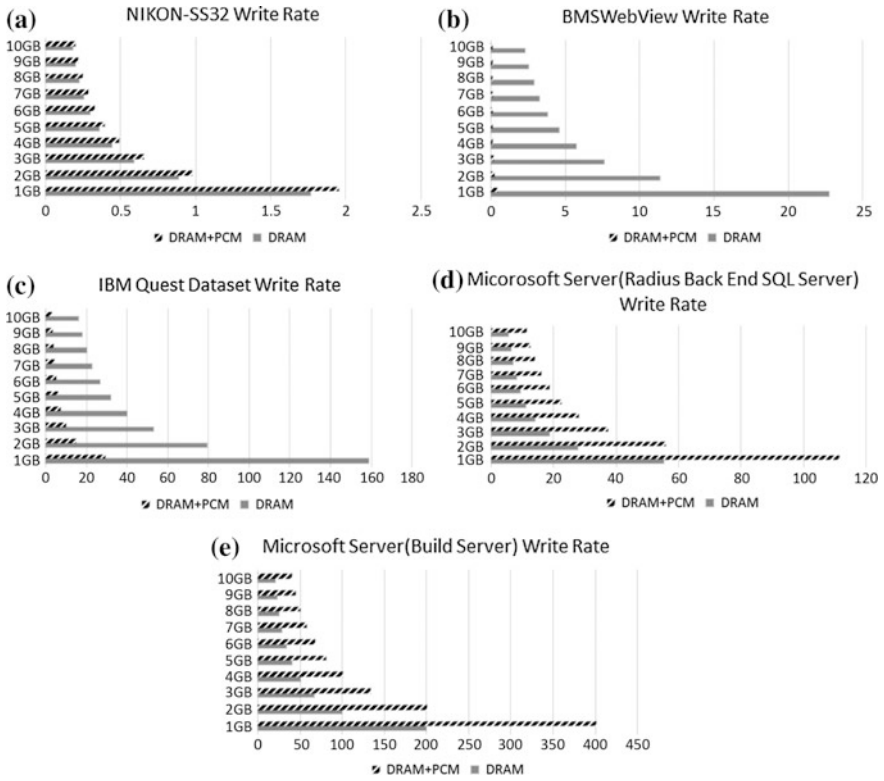


Fig. 3 Experiments with workloads

5 Conclusion

This paper proposed a hybrid PCM + DRAM main memory system for improving the inefficient memory structure in a large-scale server system based on DRAM. We obtained 10–200% more efficient memory utilization than in conventional systems. However, for a random data pattern, memory utilization of 90% is achieved. In further researches, we are planning to construct a detailed system environment using the proposed hybrid main memory. Researches into memory I/O that uses the actual main memory and file system as a hybrid main memory for storing data from the main memory to flash storage are needed.

Acknowledgements This work was supported by Basic Science Research through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2017R1D1A3B04031440).

This study was also supported by 2017 Research Grant from Kangwon National University.

References

1. Bashir M, Gill A (2016) Towards an IoT big data analytics framework: smart buildings systems. In: 2016 IEEE 18th international conference on high performance computing and communications, pp 1326–1332
2. Cai H, Xu B, Jiang L, Vasilakos AV (2017) IoT-based big data storage system in cloud computing: perspectives and challenges. IEEE Internet Things J 4(1):75–87
3. Kang W, Lee S, Moon B, Kee Y, Oh M (2014) Durable write cache in flash memory SSD for relational and NoSQL databases. In: Proceedings of the 2014 ACM SIGMOD international conference on management of data, pp 529–540
4. Malladi KT, Lee BC, Nothaft FA, Kozyrakis C, Periyathambi K, Horowitz M (2012) Towards energy-proportional datacenter memory with mobile DRAM. In: Proceedings of the 39th annual international symposium on computer architecture, pp 37–48
5. Grabner H, Nater F, Druey M, Gool LV (2013) Visual interestingness in image sequences. In: Proceedings of the 21th ACM international conference on multimedia, pp 1017–1026
6. Pakbaznia E, Pedram M (2009) Minimizing data center cooling and server power coasts. In: ISLPED'09 proceedings of the 2009 ACM/IEEE international symposium on low power electronics and design, pp 145–150
7. Gajic RB, Appuswamy R, Ailamaki A (2016) Cheap data analytics using cold storage devices. Proc VLDB Endowment 9(12):1029–1040
8. Zhong K, Zhu X, Wang T, Zhang D, Lue X, Liu D, Liu W, Sha EHH (2014) DR. Swap: energy-efficient paging for smartphones. In: Proceedings of the 2014 international symposium on low power electronics and design, pp 81–86
9. Fan J, Jiang S, Shu J, Sun L, Hu Q (2014) WL-reviver: a framework for reviving any wear-leveling techniques in the face of failures on phase change memory. In: 44th annual IEEE/IFIP international conference on dependable systems and networks, pp 228–239
10. Qureshi MK, Karidis J, Franceschini M, Srinivasan V, Lastras L, Abali B (2009) Enhancing lifetime and security of PCM-based main memory with start-gap wear leveling. In: Proceedings of the 42nd annual IEEE/ACM international symposium on microarchitecture, pp 14–23
11. Coburn J, Caulfield AM, Akel A, Grupp LM, Cupta RK, Jhala R, Swanson S (2011) NV-heaps: making persistent objects fast and safe with next-generation, non-volatile memories. In: Proceedings of the sixteenth international conference on Architectural support for programming languages and operating systems, pp 105–108
12. Hennessy JL, Patterson DA (2011) Computer architecture: a quantitative approach, 5th edn. Morgan Kaufmann, Burlington, pp 72–78
13. Yu M, Rexford J (2009) Hash, don't cache: fast packet forwarding for enterprise edge routers. In: Proceedings of the 1st ACM workshop on research on enterprise networking, pp 37–44
14. Lu Y, Sun H, Wang X, Liu X (2014) R-Memcached: a consistent cache replication scheme with Memcached. In: Proceedings of the Posters & Demos Session, pp 29–30

Sleep Posture Recognition for Bedridden Patient

Nitikorn Srisrisawang and Lalita Narupiyakul

Abstract One type of patients that needs to live on the bed for a certain time or worst, for the rest of their life is called bedridden. This type of patients need special attention from caretaker to regularly change the posture of the patient in order to prevent symptom named bed sore or pressure sore which will happen when the weight of the patient is applied to some points of the body too long which leads to injury to that certain points of the body. This research will carried out to design a system to relieve the work for the caretaker of a bedridden patient. This system consists of three parts; Sleep data collection where computer that connected to Kinect will continuously monitor the patient and send the data to the next part, Sleep posture analysis which will determine the postures of the patient from the input data, and Sleep notification part which will notify user with the current state of the patient. There are 3 machine learning algorithms that were chosen to compare their performance; Decision Tree (DT), Neural Network (NN), and Support Vector Machine (SVM). In the case of using the data from the same subjects as in the training set, DT shows lower accuracy at 93.33% than NN and SVM which achieve 100%. Similarly, in the case of using dataset that is not in the training set, DT still performs at 90% while both NN and SVM achieve 100%, the data are tested from both the subjects within the training set and new subjects but without any error exclusion which illustrates that NN which achieves 63.33% accuracy is more prone to the data with error than SVM which is 57.78%. Hence, NN is implemented with the system.

Keywords Kinect · Bed sore · Bedridden · Sleep posture · Recognition

N. Srisrisawang · L. Narupiyakul
Integrative Computational Bioscience Center, Mahidol University,
Salaya, Thailand
e-mail: nitikorn.sssw@gmail.com

L. Narupiyakul (✉)
Department of Computer Engineering, Faculty of Engineering,
Mahidol University, 999 Phuttamonthon 4 Road, Salaya
Nakhon Pathom 73170, Thailand
e-mail: lalita.nar@mahidol.ac.th

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes
in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_8

1 Introduction

Bedridden is a term to describe patients that have to stay on the bed and cannot leave their bed. When bedridden patients do not change their posture regularly, it can cause medical complications, such as, bed sore and urinary tract infection [1]. However, those complications can be prevented by reminding care giver to change patient's posture frequently.

The sleep posture monitoring system is proposed in order to prevent medical complications. There are many kinds of sensors that can detect the patient movement on the bed. In this paper, Kinect is chosen to capture the patient movement on the bed. The input data represents the body joints of the patient on the bed. The data then sent to the web server to classify the postures. The system then informs the user via mobile notification regarding the posture information. However, this work will focus mainly on the algorithm aspect of the system and the process of selecting the classifiers to be implemented in the system.

This work contains 5 sections: Sect. 1 Introduction, Sect. 2 Literature review which will conclude the other works that focus on sleeping posture Sect. 3 Sleep posture recognition system which will describes how the structure of the proposed system is and how to implement and assess different machine learning algorithm, Sect. 4 Experimental result which will report the result of model assessment from Methodology part, and lastly, Sect. 5 Conclusion.

2 Literature Review

There are many works that attempt to differentiate the sleeping postures of a person. The goals are mainly to prevent any medical complex caused by staying in the same position for too long. For example, pressure ulcer in the case of patients in the postoperative period, patients with Gastroesophageal Reflux Disease (GRD) which some works stated that laying on their side can reduce its effect [2] or patients with obstructive sleep apnea where their throat muscles are relaxed and result in narrowed airway. The works are separated by the sources of the data used in the implementation of algorithms.

The following works proposed recognition systems with a single data source in which the data used to extract the recognition features come only from one type of sensor. The sensor can be either an array of a force sensor or single depth camera.

Zachary et al. [3] implemented a system using four load cells which are placed beneath the bed. Center of Pressure (CoP) in both x-axis and y-axis are calculated from the signals from four load cells then CoPs are used to determine Angle of displacement (ANG) of each sample. Finally, K-mean clustering is used to classify the data into four groups and the generalized accuracies are 0.68, 0.57, 0.69, and 0.33 in the cases of Back, Right, Left, and Stomach respectively. When the number

of groups is reduced to, the generalized accuracies are 0.92, 0.75, and 0.86 in the cases of Back/Stomach, Right, and Left.

Similarly, Yousefi et al. [4] focused on using Force Sensor Array (FSA) which comprises of 32 × 64 force sensors that can measure the pressure up to 100 mmHg per sensor. The data from FSA system can be treated as image data. After normalization, the data were projected into an Eigenspace then two different methods which are Principal Component Analysis (PCA) and Independent Component Analysis (ICA) are applied separately. K Nearest Neighbor (KNN) classifier is used to classify each sleep posture. Both methods illustrate satisfactory accuracies which are 97.7 and 94.3% in the cases of PCA and ICA respectively. Moreover, recall rate is also in the range of 90–99% in each posture for both ICA and PCA.

On the other hand, Timo et al. [5] proposed sleep posture recognition system using depth camera which was hung from the ceiling. A 3D graph of depth called Bed Aligned Map (BAM) is extracted from each depth image. Three layers of Convolutional Neural Network (CNN) are applied to BAMs. The outputs from CNN are then fed as inputs for Multilayer Perceptron (MLP) that classifies four classes of sleep positions which is Empty, Right, Supine, and Left. However, they also try the others two methods which are MLP without CNN and Histograms of Oriented Gradients (HoG) with Support Vector Machines (SVM). Their results conclude that their first proposed method (CNN + MLP) is the best method among the rest with accuracies ranged from 94.00–98.40% in each class. The second-best method is HoG + SVM with accuracies from 85.20–96.00% while the worst method is MLP that shows accuracies in ranged from 78.80–84.80%.

Some works rely on a system that has more than one sensor (or source of raw data) in an attempt to reduce disadvantages of each sensor.

Weimin et al. [6] proposed a multimodal system that combines video data and force sensors together. This requires spatial-temporal registration to consider both data at once. Color map and edge map are extracted from each frame of the video data while the data from Force Sensing Resistor (FSR) are treated as 2D data. Moreover, joint feature, which is the leg count, is also extracted. PCA is applied to the features before a training session of multi-class SVM. Their result can be interpreted that with each sensor alone, the accuracies are lower than multimodal approach and the accuracies are improved with a present of the joint feature. They achieve the best accuracies at 99.01 and 94.05% in case of Person dependent and Person independent, respectively.

Similarly, Torres et al. [7] proposed multimodal sleep posture recognition system which comprises of three sensors namely, pressure sensor, depth sensor, RGB sensor. There are 10 sleep postures: Soldier U, Soldier D, Faller R, Faller L, Log R, Log L, Yearner R, Yearner L, Fetal R, and Fetal L plus 1 case of background; 3 light conditions: bright, medium, and dark; and 4 occlusions: clear, blanket, pillow, and blanket and pillow. The raw data are subtracted with the background then converted to gray scale data and normalized. Then, HoG is used to extract the feature from RGB data while Image Geometric Moments (gMOM) is used to extract the feature from both depth data and pressure data. After that, data from each sensor are trained with Support Vector Classifier (SVC) and Linear Discriminant

Analysis (LDA). The outputs are compared with the ground truth label to estimate the trust of each sensor. Each trust then becomes the weight of the corresponding classifier. Finally, the output of each model is used to find the maximum which then becomes the output of the multimodal classifier. In the case of the best situation (Bright light condition and clear occlusion) both SVC and LDA can achieve 100% accuracies while in the case of the worst situation (Dark light condition and Blanket and pillow occlusion), the accuracies are 17.7 and 18.6% (SVC and LDA, respectively).

In the next section, our implemented system used to recognize the sleep posture will be explained.

3 Sleep Posture Recognition System

In this section, the first part of the bedridden monitoring system is an explanation of proposed system design which will cover the overview of the system. The second part is about how the data are acquired and how the data are processed and used to train the classification model.

3.1 System Design

Our proposed system consists of 3 main parts namely; Sleep data collection part, Sleep posture analysis part, and Sleep notification part. The first part, sleep data collection part, comprises of the computer connecting to Kinect that continuously monitors a bedridden patient, the information from this part is sent to Cloud service in sleep posture analysis part via the Internet. The algorithm then applied to the data from Cloud service via Microsoft Azure Machine Learning Studio to identify the sleep posture. The last part, Sleep notification part, communicates with a mobile application to notify the user regarding patient's sleeping posture. Figure 1 shows an overview of proposed system.

Data from Kinect are received at 30 frames per second. In every 5 s, the system determines whether the patient is in the frame of the depth sensor. If the patient is not present in the frame, the system will notify the user that the patient is missing from the frame. If the patient is in the frame, the system considers 10 sequential frames of skeleton joints data from Kinect. The data from Kinect contains confidence of each joint which ranged from 0 to 1. The data is more reliable when the value is closer to 1. However, only one frame which has the highest skeleton joint confidence value will be chosen to find the posture. The process of finding the posture will be discussed in the next section. The posture of the current frame is then compared with the previous candidate. If the patient changes the posture, the system will apprise the user about the new posture of the patient. On the other hand,

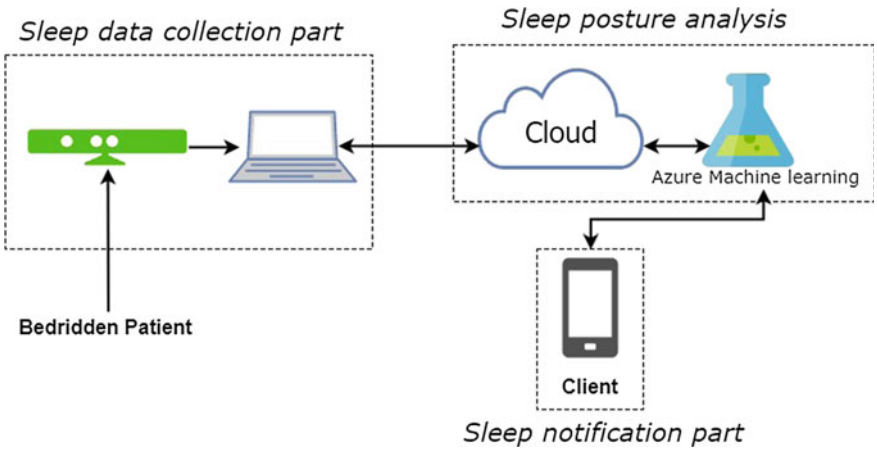


Fig. 1 An overview of system architecture

if the patient stays in the same posture for more than a certain period of time (for example, 60 s), the system will inform the user that the patient is staying in the same posture for too long. This feature could be applied in order to prevent bed sore.

3.2 Kinect Data Analysis

In this section, the details about data collection and data analysis will be given.

Experimental Data Acquisition. A mattress is placed in the same position and Kinect sensor is placed with camera tripod just below the mattress and oriented so that it points downward to the mattress. Data are collected from six normal people; the data from three of them are used to train and test the models and data from another three people are used only to test the model. There are three sleep postures that will be considered, namely: Normal (face upward), LeftFlip (face leftward), and RightFlip (face rightward). Each people have to repeat each posture for 30 times. During data collection, the mattress and the tripod are positioned in the same position and orientation.

Data Processing. In order to analyze the data from Kinect, Microsoft Azure Machine Learning Studio (MAMLS) is chosen as a tool due to its ability to provide a result as a web service in the context of a system architecture [8]. Furthermore, the performance of MAMLS can analyze and return the result instantly. The data collected from the previous step are uploaded into dataset of MAMLS. The data from three people are merged together into a single dataset. After that, the dataset is normalized with Normalize Module. Then, the dataset is split into a training dataset and a testing dataset with a ratio of 7:3, respectively.

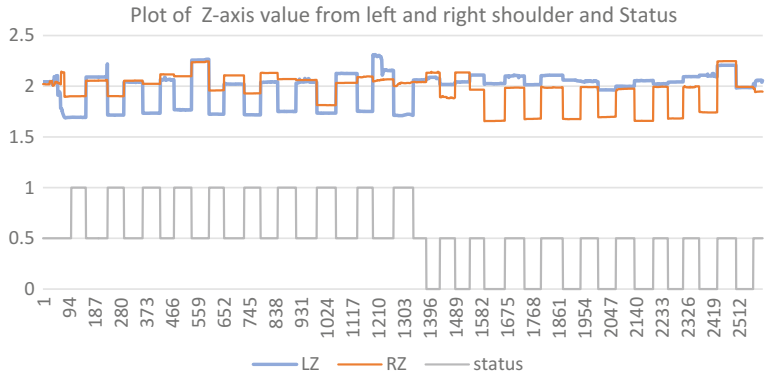


Fig. 2 The plot of Z-value from left and right shoulder and their status

The data concerned is skeleton joint position derived from Kinect sensor and after observations, we found that the value of the position of Z-axis in the left shoulder and right shoulder are good candidates for classification since the value in the Z-axis of both shoulders vary the most when the patient changes their posture. Figure 2 illustrates the plots of values from Z-axis of both shoulders with LZ being Left Z-value and RZ being Right Z-value while the status indicates the status of data belong to. Status at 0.5 means Normal, 0 means LeftFlip, and 1 means RightFlip.

However, the model is separated into two models, one to classify LeftFlip and Normal, the other one to classify RightFlip and Normal. Both models are merged into a single model that will predict three classes. There are three classifying techniques that will be used and compared: Decision Tree (DT), Neural Network (NN), and SVM. After deploying web service, the API key will be used to communicate with web service.

Models Performance Assessment. After the models are trained, the models will be evaluated to find the most suitable technique to be implemented in the system. The result of evaluation will be elaborated in Sect. 4.

4 Experimental Results

The experiment can be divided into 3 sections; Validation with the testing dataset, Validation with new dataset absent in the model, and Validation with both testing dataset and new dataset without error exclusion. Note that every section is performed with error data exclusion except the last section with does not exclude error data (Error in this work means the data with low joint data confidence level occurring by Kinect camera).

4.1 Validation with Testing Dataset

We then test each algorithm with testing dataset while the data with error are excluded from the testing process. The results from the same person within the training dataset will be discussed.

Table 1 shows the accuracies of each subject that were averaged over every class and the average accuracies of each classifier technique over every subject. It evidentially reports that NN and SVM which achieve 100% accuracies can perform better than DT which achieves at 93.33% accuracy. The data collected for this testing is derived from the same subjects that are used to train the model.

4.2 Validation with New Dataset Absent in the Model

The validation in this section is done by using a dataset which is not present in the training dataset. In the case of using data from three subjects that are new to the model, DT performs at 90% accuracy while both NN and SVM perform at 100% accuracies (Table 2).

4.3 Validation with Both Testing and New Datasets Without Error Exclusion

There is no way to compare the performance of NN and SVM with the current results since they both achieve 100%. We then set up another experiment where, unlike the first experiment, every data point is included no matter if it contains error or not, in order to compare the performance of NN and SVM.

Table 1 The results from first three subjects for each classifier

Subject	DT	NN	SVM
S1	86.67	100	100
S2	100	100	100
S3	93.33	100	100
Average	93.33	100	100

Table 2 The results from new subject

Subject	DT	NN	SVM
N1	100	100	100
N2	73.33	100	100
N3	96.67	100	100
Average	90	100	100

Table 3 The results from the data without any exclusion

Subject	NN	SVM
S3	76.67	70
N1	53.33	43.33
N3	60	60
Average	63.33	57.78

From Table 3, NN shows better performance at 63.33% than SVM which is 57.78%. According to the aforementioned result, it can be seen that NN is the best classifier among the 3 chosen algorithms. Finally, NN will be the classifier to be implemented with the proposed system.

5 Conclusion

According to the result, NN is the best model to be used when errors are present in the data. However, SVM also performs well in the case where errors are absent in the dataset. Nonetheless, DT performs the worst among 3 models but not in the aspect of visualization of the model since DT can be visualized easily with a simple tree diagram. The proposed system with NN can help notify users when the patient is absent within the frame or when the user stays too long in the same position.

Acknowledgements This research is financially supported by Crown Property Bureau Funding, Thailand.

References

1. Bains P, Minhas AS (2011) Profile of Home-based Caregivers of Bedridden Patients in North India. Indian J Community Med 36:114–119. https://doi.org/10.4103/0970-0218.84129
2. Khoury RM, Camacho-Lobato L, Katz PO et al (1999) Influence of spontaneous sleep positions on nighttime recumbent reflux in patients with gastroesophageal reflux disease. Am J Gastroenterol 94:2069–2073. https://doi.org/10.1111/j.1572-0241.1999.01279.x
3. Beattie ZT, Hagen CC, Hayes TL (2011) Classification of lying position using load cells under the bed. In: 2011 annual international conference on IEEE engineering in medicine and biology society IEEE, pp 474–477
4. Yousefi R, Ostadabbas S, Faezipour M et al (2011) Bed posture classification for pressure ulcer prevention. In: annual international conference on IEEE engineering in medicine and biology society IEEE, pp 7175–7178
5. Grimm T, Martinez M, Benz A, Stiefelhagen R (2016) Sleep position classification from a depth camera using Bed Aligned Maps. In: 2016 23rd international conference on pattern recognition IEEE, pp 319–324

6. Huang W, Wai AAP, Foo SF et al (2010) Multimodal sleeping posture classification. 2010 20th International conference on pattern recognition 4336–4339. https://doi.org/10.1109/icpr.2010.1054
7. Torres C, Hammond SD, Fried JC, Manjunath BS (2015) Sleep pose recognition in an icu using multimodal data and environmental feedback. Springer, Cham, pp 56–66
8. Team A (2016) AzureML: Anatomy of a machine learning service. In: Dorard L, Reid MD, Martin FJ (eds) Proc. 2nd international conference on prediction APIs Apps, PMLR, Sydney, Australia, pp 1–13

Wearable Devices Downlink and Uplink Transmission in Multi-hop HetNet with Full-Duplex Relay

Rong Ye, Kang Kang, Zhenni Pan and Shigeru Shimamoto

Abstract This paper we provide a tractable system for 2-tier users, wearable device terminals (WTDs) are in tier 2 and under the control of tier 1 users. Tier 1 user works as full-duplex relay to charge the WTDs on the downlink (DL) based on wireless power transfer (WPT), then help WTDs to transmit the information on the uplink (UL). Both of the users and picocell base stations (BS) are randomly distributed based on homogeneous poison point process (HPPP). We only considered the user association for the tier 1 users that tier 1 users always connect to the closest BS on UL since WTDs tier 2 users are close enough to tier 1 users. In this paper, we analyzed the processes of wireless power transfer on DL and information transmission on UL. We derived analysis of the average harvested energy on DL for both 2-tier users, the average ergodic rate and outage probability on UL for WTDs tier 2 users. According to the analytical results, it is observed that with the density of the picocell base stations increased, the average harvested energy increased, but the average ergodic rate and UL outage probability degraded significantly because of the UL performance decreased.

Keywords Wireless power transfer · Full-duplex relay · Wearable device terminals Stochastic geometry

1 Introduction

As the WTDs developed rapidly, the WTDs played the important role in our day life, wearable communication networks will be the next frontier type of networks for wireless communications. A main challenge for wearable networks is to find green methods to support them in dense environments with low-latency, high energy efficiency [1–3]. Traditional energy harvesting sources such as solar, wind, hydroelectric depend on the locations and environments. Wireless power transfer

R. Ye (✉) · K. Kang · Z. Pan · S. Shimamoto
Waseda University, Tokyo, Japan
e-mail: 2014yerong@gmail.com

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_9

(WPT) is an alternative approach to prolong the lifetime of mobile devices since the radio signals carry the information and the energy at the same time [4]. With the deployment of heterogeneous network (HetNet), lower transmit power BSs such as picocells, fetmocells and millimeter-wave BSs are used to extend the edge coverage in macro cellular networks with high power efficiency [5, 9]. An analytical model for k-tier MIMO applied HetNet has been proposed, the different user association scheme can make sure the users harvest more energy on DL and minimized the effect of path loss on UL [6]. In addition, the potentially harmful interference received by the users will become a precious new energy source. Recently, the potential of harvesting the ambient energy in fifth-generation (5G) networks has been investigated [7, 8]. Moreover, full duplex communication is capable to double the capacity through transmitting and receiving on the same channel at the same time by suffering from the self-interference that the receiving side will be effected by the transmitting side. Along with the advances in signal designs and mitigation schemes showed in self-interference cancellation methods [10], full-duplex can potentially be utilized in heterogeneous cellular network [11].

The contributions of this paper are summarized as below:

- We present a tractable system model that all of the WTDs are central under controlled by high level priority relay terminal devices such as smart phones, pad. Especially some low-power devices such like smart watches, smart glasses, wearable fitness trackers, accessories, etc. Imagine that many people may bring more than 3 kinds of WTDs, it is inconvenient to charge them in the subway station, railway station or airport that such crowded places. Unlike [12], we proposed that the picocell BS and 2-tier users randomly located with particular transmit power, path loss exponent, density.
- We derived analysis of the average harvested energy DL for both 2-tier users, that shows how much energy the users can harvest from the DL and ambient RF at the same time. We analyzed the outage probability and the average ergodic rate on UL for WTDs tier 2 users. The outage probability related to the SINR threshold equals to the CDF of the SINR. The average ergodic rate integrated by the throughput threshold.

2 System Description

The entire communication consists of two different phases, namely, energy harvesting phase and wireless information transfer phase. Assuming each one duration of a communication block is T. The first phase of duration is τT, where $\tau \in (0, 1)$ is the time allocation factor, each tier user harvests energy from direct DL and the ambient RF. The rest time of duration $(1 - \tau)$T is capable of transmitting the information on the UL [13]. Figure 1 depicts 2-tier users deployed in network such as full-duplex tier 1 relay and WTD tier 2 users. Tier 1 users as a full-duplex relay can harvest the energy from the picocell BSs and transfer the energy to tier 2 users

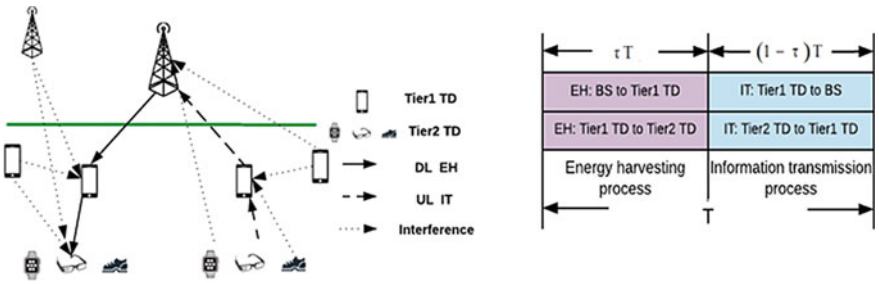


Fig. 1 The system network model and protocol show that on the DL Tier1/2 users harvest the energy in time slot τT, then on UL Tier1/2 users upload the information in time slot $(1-\tau)$T

at the same time slot τT on DL, then support the tier 2 users use the harvested energy as transmit power to transmit data on UL at the same time slot $(1-\tau)$T.

We assume that each picocell BS is equipped with single antenna. The locations of picocell BSs and 2-tier users are modeled following an independent HPPP Φ_{Bs} with density $\lambda_{Bs}, \lambda_{Um}, \lambda_{Us}$. In the high density HetNet, the density of the WTDs tier 2 users is much larger than picocell BSs and the tier 1 users, considering that there exists only one active WTDs tier 2 user and tier 1 user at each time slot in each picocell.

In this system model, we did not consider direct energy transfer and information transfer with WTDs tier 2 user and picocell base stations, because picocell is limited resources and only server for tier 1 users because tier 1 users have high requirement in information transfer and energy.

As we know the full-duplex relay will suffer from the self-interference, but along with the development of the self-interference cancellation (SIC) to further reduce the interference by using analog/digital domain SIC, although this part of study is still challengeable. We considered the self-interference can be perfectly cancelled in this paper.

2.1 DL Wireless Power Transfer

In the picocell, only one user of each tier can be allowed to communicate with its serving picocell BS at a divided time slot, the allocated power transfer time for each tier user is τT, where $\tau \in (0,1)$. On the phase of energy harvesting, the total harvested energy at a typical tier 1 user o that is associated with the closest picocell base station and the total harvested energy at a typical WTD tier 2 user o are given by

$$\begin{cases} E_{Um} = \eta\tau T P_{Bs} g_o L_0 \left(\max\left\{\left|X_{o,m}\right|, d\right\}\right) + \eta\tau T (I_{Bs} + I_{mu}) \\ E_{Us} = \eta\tau T P_{Um} h_o L_0 \left(\max\left\{\left|X_{o,u}\right|, d\right\}\right) + \eta\tau T \left(I'_{Bs} + I'_{mu}\right) \end{cases}. \quad (1)$$

Where E_{Um},E_{Us} are the ideal total energy that 2-tier users can harvest on the DL from the serving picocell BS directly, also can be charged from the interfered users,

η $(0 < \eta < 1)$ is the RF-to-DC conversion efficiency, P_{Bs}, P_{Um} are the picocell transmit power and tier 1 user transmit power, respectively, $L(max\{|\mathrm{X}|, d\}) = \beta|\mathrm{X}|^{-\alpha}$ is the path loss function, β is the frequency dependent constant value, $\alpha(2 < \alpha < 4)$ is the path loss exponent, d is the reference distance, $g_o, h_o \sim \exp(1)$ and $\left|\mathrm{X}_{\mathrm{o,m}}\right|$, $\left|\mathrm{X}_{\mathrm{o,u}}\right|$ are the small-scale fading channel power gain, by mean 1, and the distance between the serving picocell BS and the typical tier 1 user, the distance between the tier 1 user and the active WTD tier 2 user, respectively based on the Rayleigh fading. The energy harvested from the interference can be given by

$$\begin{cases} \mathrm{I_{Bs}} = \sum_{\mathrm{i}\in\Phi_{\mathrm{Bs}}\backslash\{\mathrm{o}\}} \mathrm{P_{Bs}} g_{\mathrm{b,o}} \mathrm{L}_0(\max\{|\mathrm{X_i}|, \mathrm{d}\}) \\ \mathrm{I_{mu}} = \sum \mathrm{P_{Um}} g_{\mathrm{m,o}} \mathrm{L}_0\left(\max\left\{\left|\mathrm{X_j}\right|, \mathrm{d}\right\}\right) \end{cases}, \tag{2}$$

$$\begin{cases} \mathrm{I'_{Bs}} = \sum \mathrm{P_{Bs}} h_{\mathrm{b,o}} \mathrm{L}_0(\max\{|\mathrm{X_k}|, \mathrm{d}\}) \\ \mathrm{I'_{mu}} = \sum_{\mathrm{s}\in\Phi_{\mathrm{Um}}\backslash\{\mathrm{o}\}} \mathrm{P_{Um}} h_{\mathrm{m,o}} \mathrm{L}_0(\max\{|\mathrm{X_s}|, \mathrm{d}\}) \end{cases}, \tag{3}$$

where $\mathrm{I_{Bs}}$, $\mathrm{I_{mu}}$, $\mathrm{I'_{Bs}}$, $\mathrm{I'_{mu}}$ are the sum of the interference harvested from the DL under the condition of full-duplex relay network, including the interference from the interfering picocell BSs and interfering tier 1 users, $g, h \sim \exp(1)$, $|\mathrm{X}|$ are the small-scale fading interfering channel power gain by mean 1, the distance between interfering picocell BSs $\mathrm{i} \in \Phi_{\mathrm{Bs}}\backslash\{\mathrm{o}\}$ (except the serving picocell BS) or tier 1 users $\mathrm{s} \in \Phi_{\mathrm{Um}}\backslash\{\mathrm{o}\}$ (except the serving tier 1 user) and the typical charged users, moreover based on the Rayleigh fading.

2.2 UL Information Transmission

In UL information transmission phase, the 2-tier users will make use of the harvested energy as transmit power to transmit the data on UL at the same time slot $(1-\tau)\mathrm{T}$, where $\tau \in (0, 1)$. SINR received on the tier 1 user from typical WTD tier 2 user can be given as

$$\mathrm{SINR_{Um}} = \frac{\mathrm{P_{Us}^{DL}} h'_{\mathrm{o}} \mathrm{L}_0\left(\max\left\{\left|\mathrm{X'_{o,u}}\right|, \mathrm{d}\right\}\right)}{\mathrm{I'_{mu}} + \mathrm{I'_{us}} + \sigma^2} \tag{4}$$

where $\mathrm{P_{Us}^{DL}} = \frac{\mathrm{E_{Us}}}{(1-\tau)\mathrm{T}}$, $h'_{\mathrm{o}} \sim \exp(1)$, $\left|\mathrm{X'_{o,u}}\right|$ are the small-scale fading interfering channel power gain by mean 1, the distance between typical WTD tier 2 user and serving tier 1 user, respectively. In addition

$$\begin{cases} \mathrm{I'_{mu}} = \sum \mathrm{P_{Um}^{DL}} g'_{\mathrm{m,o}} \mathrm{L}_0\left(\max\left\{\left|\mathrm{X'_j}\right|, \mathrm{d}\right\}\right) \\ \mathrm{I'_{us}} = \sum\limits_{\mathrm{s'}\in\Phi_{\mathrm{Us}}\backslash\{\mathrm{o}\}} \mathrm{P_{Us}^{DL}} h'_{\mathrm{o,b}} \mathrm{L}_0\left(\max\left\{\left|\mathrm{X'_k}\right|, \mathrm{d}\right\}\right) \end{cases} \tag{5}$$

where I'_{mu}, I'_{us} are the sum of the interference including the interference from the interfering tier 1 users and interfering WTD tier 2 users, g, $h \sim \exp(1)$, $|X|$ are the small-scale fading interfering channel power gain by mean 1, the distance between interfering tier 1 users or tier 2 users $s' \in \Phi_{Us} \backslash \{o\}$ (except the typical tier 2 user) and the typical tier 1 user based on the Rayleigh fading.

3 Exact Analysis of Downlink and Uplink Transmission

Suggesting x is the distance between the tier 1 user and the picocell BS. Since the typical tier 1 user connect with the closest picocell BS, no other BS can be closer than X. The probability density function (pdf) of x can be derived as

$$\Pr(x > X) = e^{-\lambda\pi X^2}. \tag{6}$$

Therefore, the CDF is $\Pr(x \leq X) = F_x(X) = 1 - e^{-\lambda\pi X^2}$, the pdf can be denoted as $f_x(x) = \frac{dF_x(X)}{dx} = e^{-\lambda\pi x^2} 2\pi\lambda x$.

3.1 Phase-1 DL Wireless Power Transfer

On the DL full-duplex cooperative HetNet, 2-tier user harvest the directed energy and the RF at divided time slot τT as follows.

Theorem 1 *The average harvested energy at 2-tier typical users are given by*

$$\begin{aligned}\tilde{E}_{Um} = \eta\tau T\beta(P_{Bs}(1(x \leq d)d^{-\alpha_B} + 1(x > d)x^{-\alpha_B}) \\ + 2\pi\lambda_{Bs}P_{Bs}\left(1(x<d)\left(\frac{1}{2}d^{-\alpha_B}\left(d^2 - x^2\right) - \frac{d^{2-\alpha_B}}{2-\alpha_B}\right)\right. \\ \left. -1(x > d)\frac{X^{2-\alpha_B}}{2-\alpha_B}\right) + 2\pi\lambda_{Bs}P_{Mu}\left(\tfrac{1}{2}d^{2-\alpha_B} + \frac{d^{2-\alpha_B}}{\alpha_B - 2}\right)\Big)\end{aligned} \tag{7}$$

and

$$\begin{aligned}\tilde{E}_{Us} = \eta\tau T\beta(P_{Um}(1(y \leq d)d^{-\alpha_U} + 1(y > d)y^{-\alpha_U}) \\ + 2\pi\lambda_{Bs}P_{Um}\left(1(y \leq d)\left(\frac{1}{2}d^{-\alpha_U}\left(d^2 - y^2\right) - \frac{d^{2-\alpha_U}}{2-\alpha_U}\right)\right. \\ \left. -1(y > d)\frac{y^{2-\alpha_U}}{2-\alpha_U}\right) + 2\pi\lambda_{Bs}P_{Bs}\left(\tfrac{1}{2}d^{2-\alpha_B} + \frac{d^{2-\alpha_B}}{\alpha_B - 2}\right)\Big)\end{aligned} \tag{8}$$

where $\tilde{E}_{Um}$, $\tilde{E}_{Us}$ *are the total energy can be harvested on the DL on tier 2 user, given*

by (1), (2), (3), *follows the Campbell's theorem* $\mathbf{E}\left\{\sum_{x_i \in \Phi} f(x_i)\right\} = \lambda \int \mathbf{E}\{f(x)\}dx$.

The average harvested energy for a typical tier 1 user that is associated with a closest picocell BS. From (6), the average harvested energy can be denoted as $\bar{\mathrm{E}}(\mathrm{x}) = \int_0^\infty \tilde{\mathrm{E}}(\mathrm{x})\mathrm{f}(\mathrm{x})\mathrm{dx}$. We set the distance between 2-tier users as a constant value. Imagining that tier 2 wearable devices just in the pockets or bags to connect the smart phone in your hand, so the distance is close enough to be assumed as a constant value d_0.

3.2 Phase 2-Wireless Information Transmission

On the UL full-duplex cooperative HetNet, the WTD tier 2 users transmit the information signal with a special transmit power that harvested from the DL at the divided same time slot $(1-\tau)\mathrm{T}$ [14].

Theorem 2 *The average ergodic rate received at tier 1 user is given by*

$$\begin{aligned} \mathrm{R}_{\mathrm{Um}}^{\mathrm{Exact}} &= \mathrm{E}_{\mathrm{SINR_{Um}}}\{(1-\tau)\ln(1+\mathrm{SINR_{Um}})\} \\ &= \int_{\mathrm{R_1}>0}^{\infty} \Pr(\mathrm{SINR_{Um}} > \mathrm{t_1})\mathrm{dR_1} \end{aligned} \tag{9}$$

where $\mathrm{t_1} = \mathrm{e}^{\mathrm{R_1}/(1-\tau)} - 1$, $\mathrm{R_1}$ *is the throughput threshold, then from* (4), (5) *we can derive the coverage probability of* $SINR_{Um}$ as

$$\begin{aligned} \Pr(\mathrm{SINR_{Um}} > \mathrm{t_1}) &= \Pr\left(\frac{\mathrm{P_{Us}^{DL}} \mathrm{h_o'} \mathrm{L_0}\left(\max\left\{\left|\mathrm{X_{o,u}'}\right|, \mathrm{d}\right\}\right)}{\mathrm{I_{mu}'} + \mathrm{I_{us}'} + \sigma^2} > \mathrm{t_1}\right) \\ &= \Pr\left(\mathrm{h_o'} > \mathrm{t_1}\Delta_{\mathrm{Us}}^{-1}\left(\mathrm{I_{mu}'} + \mathrm{I_{us}'} + \sigma^2\right)\mathrm{d}^{\alpha_{\mathrm{U}}}\right) \\ &= \mathcal{L}_{\mathrm{I_{mu}'}}\left(\mathrm{t_1}\Delta_{\mathrm{Us}}^{-1}\mathrm{d}^{\alpha_{\mathrm{U}}}\right) \cdot \mathcal{L}_{\mathrm{I_{us}'}}\left(\mathrm{t_1}\Delta_{\mathrm{Us}}^{-1}\mathrm{d}^{\alpha_{\mathrm{U}}}\right) \cdot \exp\left(-\mathrm{t_1}\sigma^2\Delta_{\mathrm{Us}}^{-1}\mathrm{d}^{\alpha_{\mathrm{U}}}\right) \end{aligned} \tag{10}$$

where, $\Delta_{\mathrm{Us}} = \mathrm{P_{Us}^{DL}}\beta = \frac{\mathrm{E_{Us}}}{(1-\tau)\mathrm{T}}\beta$, *the tier 2 WTDs is close enough with tier 1 user, so there is no user association for typical tier 2 user, we set as reference distance* $d = 1$. *Therefore, the UL outage probability* [15] *is given by*

$$\mathrm{P_{out_{Um}}^{UL}} = 1 - \Pr(\mathrm{SINR_{Um}(x)} > \mathrm{t_1}) \tag{11}$$

In addtion

$$\begin{aligned}\mathcal{L}_{\mathrm{I}'_{\mathrm{mu}}}(\mathrm{s}) &= exp\left(-2\pi\lambda_{\mathrm{Bs}}\int_0^{\infty}\frac{s\Delta_{\mathrm{Us}}\max\{\nu,\mathrm{d}\}^{-\alpha_{\mathrm{U}}}}{1+s\Delta_{\mathrm{Us}}\max\{\nu,\mathrm{d}\}^{-\alpha_{\mathrm{U}}}}\,vdv\right)\\ &= exp\left(-2\pi\lambda_{\mathrm{Bs}}\left(\int_0^{\mathrm{d}}\frac{s\Delta_{\mathrm{Us}}\mathrm{d}^{-\alpha_{\mathrm{U}}}}{1+s\Delta_{\mathrm{Us}}\mathrm{d}^{-\alpha_{\mathrm{U}}}}\,vdv\right.\right.\\ &\quad\left.\left.+\int_{\mathrm{d}}^{\infty}\frac{s\Delta_{\mathrm{Us}}v^{-\alpha_{\mathrm{U}}}}{1+s\Delta_{\mathrm{Us}}v^{-\alpha_{\mathrm{U}}}}\,vdv\right)\right)\\ &= exp\left(-2\pi\lambda_{\mathrm{Bs}}\left(\frac{\mathrm{t}_1\mathrm{d}^{\alpha_{\mathrm{U}}}}{2(1+\mathrm{t}_1)}\right.\right.\\ &\quad\left.\left.+\frac{\mathrm{t}_1\mathrm{d}^2}{\alpha_{\mathrm{U}}-2}2F1\left[1,\frac{\alpha_{\mathrm{U}}-2}{\alpha_{\mathrm{U}}},2-\frac{2}{\alpha_{\mathrm{U}}},-\mathrm{t}_1\right]\right)\right)\end{aligned} \tag{12}$$

and

$$\begin{aligned}\mathcal{L}_{\mathrm{I}'_{\mathrm{us}}}(\mathrm{s}) &= exp\left(-2\pi\lambda_{\mathrm{Bs}}\int_{d_0}^{\infty}\frac{s\Delta_{\mathrm{Us}}\max\{\mu,\mathrm{d}\}^{-\alpha_{\mathrm{U}}}}{1+s\Delta_{\mathrm{Us}}\max\{\mu,\mathrm{d}\}^{-\alpha_{\mathrm{U}}}}\,\mu\,d\,\mu\right)\\ &= exp\left(-2\pi\lambda_{\mathrm{Bs}}\left(\int_{d_0}^{\mathrm{d}}\frac{s\Delta_{\mathrm{Us}}\mathrm{d}^{-\alpha_{\mathrm{U}}}}{1+s\Delta_{\mathrm{Us}}\mathrm{d}^{-\alpha_{\mathrm{U}}}}\,\mu\,d\,\mu\right.\right.\\ &\quad\left.\left.+\int_{\mathrm{d}}^{\infty}\frac{s\Delta_{\mathrm{Us}}\mu^{-\alpha_{\mathrm{U}}}}{1+s\Delta_{\mathrm{Us}}\mu^{-\alpha_{\mathrm{U}}}}\,\mu\,d\,\mu\right)\right)\\ &= exp\left(-2\pi\lambda_{\mathrm{Bs}}\left(\frac{(\mathrm{d}^2-d_0^2)\mathrm{t}_1}{2(1+\mathrm{t}_1)}\right.\right.\\ &\quad\left.\left.+\frac{\mathrm{t}_1\mathrm{d}^2}{\alpha_{\mathrm{U}}-2}2F1\left[1,\frac{\alpha_{\mathrm{U}}-2}{\alpha_{\mathrm{U}}},2-\frac{2}{\alpha_{\mathrm{U}}},-\mathrm{t}_1\right]\right)\right)\end{aligned} \tag{13}$$

where $\mathcal{L}_{\mathrm{I}'_{\mathrm{mu}}}(.)$, $\mathcal{L}_{\mathrm{I}'_{\mathrm{us}}}(.)$ based on the Laplace transform are the PDF of $\mathrm{I}'_{\mathrm{mu}}$, $\mathrm{I}'_{\mathrm{us}}$, and are derived by making use of the generating function of PPP [16].

4 Numerical Results

In this section, we use the analytical expressions of the average received energy, the UL average ergodic rate, and UL outage probability to evaluate the system network. The exact results are validated by the method of Monte Carlo, where random the

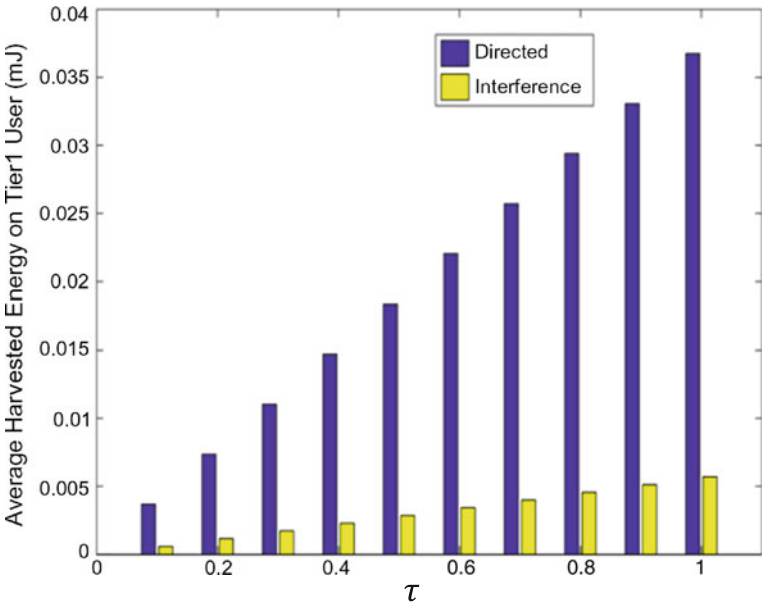


Fig. 2 The average harvested energy on tier 1 user with parameters $\alpha_B = 3.5$, $\alpha_U = 3$

distance of each tier users based on the density $\lambda_{\mathrm{Bs}} = 10^{-3}$, $\alpha_B = 3.5$, $\alpha_U = 3$, $d = 1$. For all the numerical results, $\mathrm{P_{Bs}} = 33\,\mathrm{dBm}$, $\mathrm{P_{Bs}} = 25\,\mathrm{dBm}$.

In Figs. 2 and 3 plot the effect of allocated time slot factor for energy harvesting phrase to the average harvested energy. We can observe that the harvested energy divided into the direct WPT and from the ambient RF. The average harvested energy increased with the allocated time slot for energy harvesting, and the energy harvested directly from the signal larger than from the ambient RF.

Figure 4 plots the effect of transmit power and the density λ_{Bs} to the average harvested energy of 2-tier users. With the increasing density of picocell BS and transmit power, the harvested energy will increase. When the transmit power of picocell BS are little bigger than the tier 1 user, the average harvested energy are not significantly improved.

Figure 5 plots the effect of density of picocell BS and allocated time slot factor for energy harvesting to the average ergodic rate received on tier 1users. With the increasing of the λ_{Bs}, the average ergodic rate decreased. Due to the tradeoff between the benefits from the transmit power increased, the more allocated time slot for energy harvesting and the less time left for information transmission, the average ergodic rate of WTD decreased significantly.

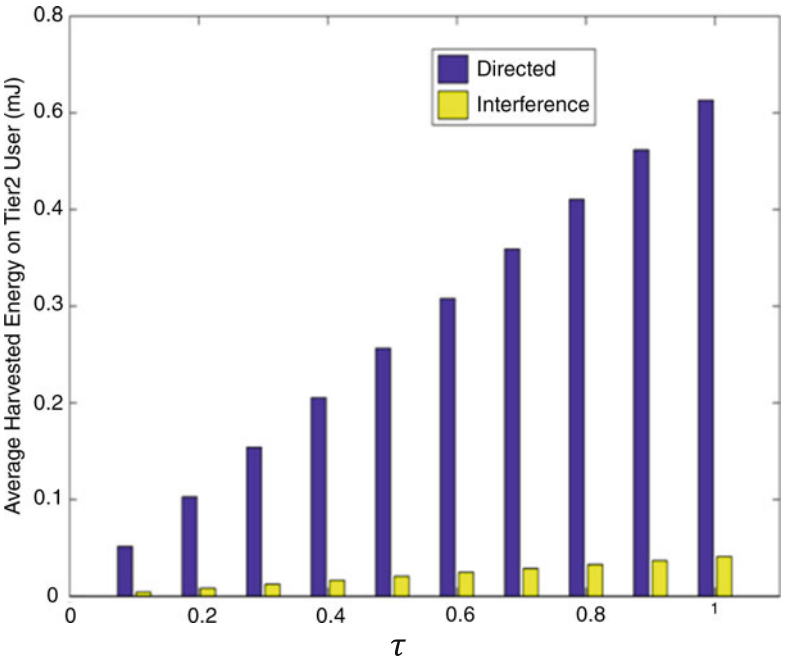


Fig. 3 The average harvested energy on tier 2 WTD with parameters $\alpha_B = 3.5$, $\alpha_U = 3$

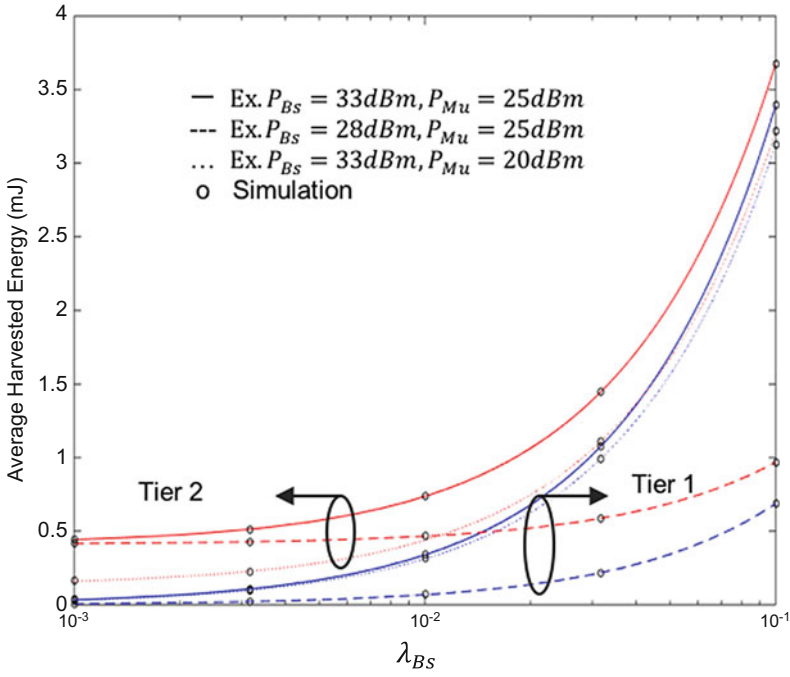


Fig. 4 The comparison of average harvested energy in 2-tier users with $\alpha_B = 3.5$, $\alpha_U = 3$

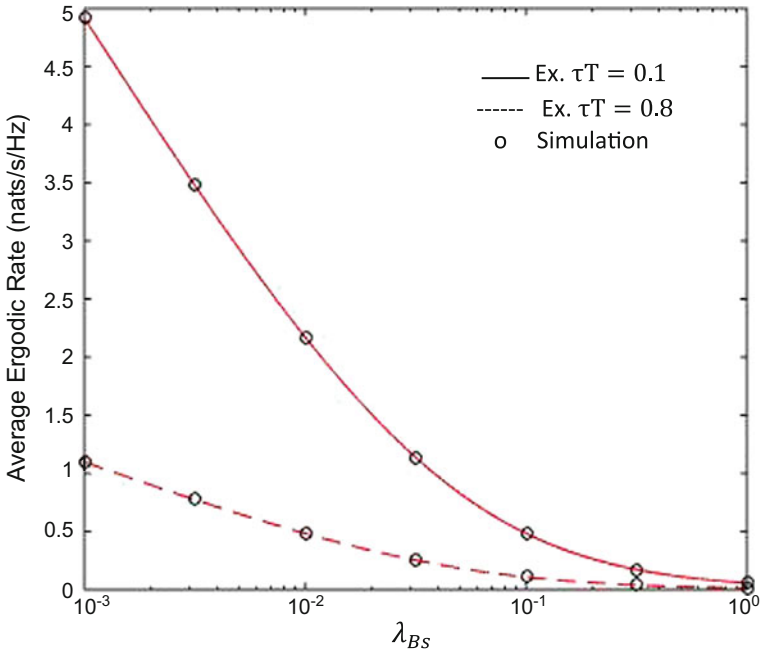


Fig. 5 The average ergodic rate of tier 2 WTD on UL with $\tau T = 0.1/0.8$

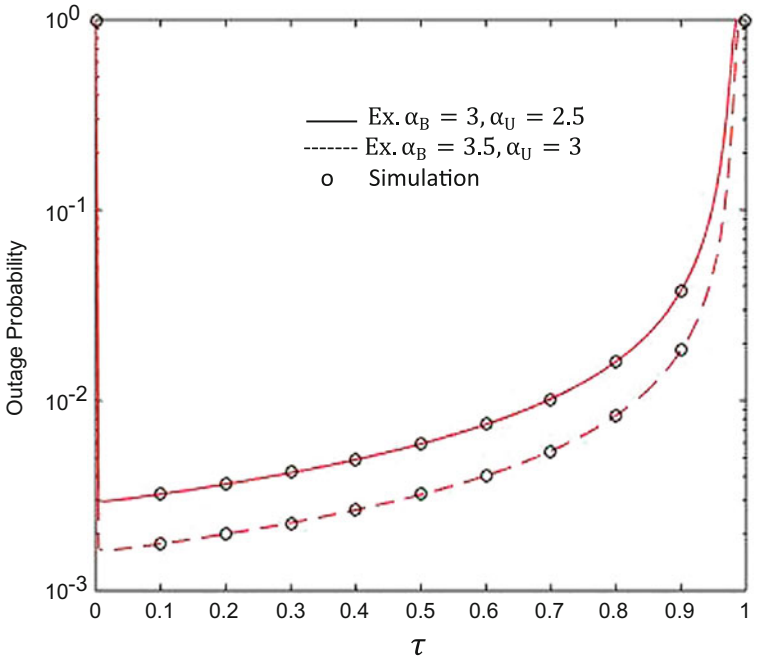


Fig. 6 The UL outage probability of tier 2 WTD with $\alpha_B = 3.5/3$, $\alpha_U = 3/2.5$

Figure 6 plots the effect of allocated time slot for energy harvesting and path loss exponent to outage probability. With the increasing of the τT, UL outage probability improved then degraded due to the tradeoff benefit from the increased transmit power of typical WTD and detrimental increased transmit power of interfering users. With path loss exponent increased, outage probability improved.

5 Conclusions

In this paper, we examined a tractable model for multi-hop HetNet system with full-duplex cooperative relay. The tier 1 user and the WTDs tier 2 users can be charged on DL at the same allocated time slot τT, then tier 1 user will support tier 2 user transmit the information on UL at the same allocated time slot $(1-\tau)$T. By the distribution of HPPP with particular density, we derived the equations of the average harvested energy, the average ergodic rate and UL outage probability. It is observed that the average harvested energy increased with the allocated time slot for energy harvesting. Due to the transmit power increased, transmit power of interfering users also increased. The less time left for information transmission, the average ergodic rate of WTD and the outage probability degraded.

References

1. Pyattaev A, Johnsson K, Andreev S et al (2015) Communication challenges in high-density deployments of wearable wireless devices. Wirel Commun IEEE 22(1):12–18
2. Wu G, Talwar S, Johnsson K et al (2011) M2M: from mobile to embedded internet. IEEE Commun Mag 49(4):36–43
3. Zhang X, Jiang H, Zhang L et al (2010) An energy-efficient ASIC for wireless body sensor networks in medical applications. IEEE Trans Biomed Circ Syst 4(1):11–18
4. Lu X, Niyato D, Wang P et al (2015) Wireless charger networking for mobile devices: fundamentals, standards, and applications. IEEE Wirel Commun 22(2):126–135
5. Dhillon HS, Ganti RK, Baccelli F et al (2012) Modeling and analysis of K-Tier downlink heterogeneous cellular networks. IEEE J Sel Areas Commun 30(3):550–560
6. Zhu Y, Wang L, Wong KK et al (2016) Wireless power transfer in massive MIMO-aided HetNets with user association. IEEE Trans Commun 64(10):4181–4195
7. Hossain E, Rasti M, Tabassum H et al (2014) Evolution toward 5G multi-tier cellular wireless networks: An interference management perspective. IEEE Wirel Commun 21(3):118–127
8. Liu D, Wang L, Chen Y et al (2015) User association in 5G networks: a survey and an outlook. IEEE Commun Surv Tutorials 18(2):1018–1044
9. Zhu Y, Zheng G, Wang L et al (2018) Content placement in cache-enabled sub-6 GHz and millimeter-wave multi-antenna dense small cell networks. arXiv preprint arXiv:1801.05756
10. Tong Z, Haenggi M (2015) Throughput analysis for full-duplex wireless networks with imperfect self-interference cancellation. IEEE Trans Commun 63(11):4490–4500
11. Lee J, Quek TQS (2015) Hybrid full-/half-duplex system analysis in heterogeneous wireless networks. IEEE Trans Wireless Commun 14(5):2883–2895

12. Chandrasekhar V, Andrews JG, Muharemovic T et al (2009) Power control in two-tier femtocell networks. IEEE Trans Wireless Commun 8(8):4316–4328
13. Luo S, Rui Z, Teng JL (2013) Optimal save-then-transmit protocol for energy harvesting wireless transmitters. IEEE Trans Wireless Commun 12(3):1196–1207
14. Novlan TD, Dhillon HS, Andrews JG (2013) Analytical modeling of uplink cellular networks. IEEE Trans Wireless Commun 12(6):2669–2679
15. Andrews JG, Baccelli F, Ganti RK (2011) A tractable approach to coverage and rate in cellular networks. IEEE Trans Commun 59(11):3122–3134
16. Baccelli F, Błaszczyszyn B (2009) Stochastic geometry and wireless networks: volume I theory. Found Trends® Network 3(3–4):249–449

An Adaptive Contention Window Mechanism for SMAC Protocol in Wireless Sensor Network

Jieying Zhou, Yinglin Liu, Rongfa Qiu and Shaopeng Huang

Abstract This paper proposes an adaptive contention window mechanism (ACWM) to improve the performance of SMAC protocol in wireless sensor network. This mechanism adjusts contention window by introducing two parameters to measure the state and the fairness of the network. In order to reflect the congestion of the network, ACWM introduces a parameter which represents the occupancy of the channel. This parameter is the ratio between conflict time and free time of each node. We use another parameter to measure the fairness among nodes and balance the time when nodes occupy channel. According to the dynamic changes of the network, the competition window is adaptively adjusted to improve the throughput, reduce the network delay and minimize the additional energy consumption caused by conflicts. ACWM protocol was compared with SMAC via the simulation software NS2 in a scenario of dense node topology. Simulation results show that ACWM protocol achieved better performance with higher throughput, lower average delay and higher energy utilization efficiency.

Keywords SMAC protocol · Higher throughput · Wireless sensor network Contention window

1 Introduction

Sensor Media Access Control (SMAC) protocol is a Media Access Control (MAC) protocol designed for Wireless Sensor Network (WSN) [1–3]. Traditional MAC protocol is based on minimum delay and maximum throughput. It is not applicable for WSN [4, 5]. WSN focuses on reducing energy consumption when a

J. Zhou · Y. Liu (✉) · R. Qiu · S. Huang
School of Electronics and Information Technology,
Sun Yat-sen University, Guangzhou 510006, Guangdong, China
e-mail: liuylin6@mail2.sysu.edu.cn

J. Zhou
e-mail: isszjy@mail.sysu.edu.cn

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_10

certain throughput is guaranteed. WSN has limited energy and unique channel access technology. Due to these characteristics, SMAC with dormancy mechanism and based on IEEE 802.11 is proposed [6, 7]. In order to improve the connectivity of the network, the protocol mentioned above adjusts and exchanges the synchronous information to maintain synchronization between nodes. Besides, the protocol reduces the energy consumption by periodic monitoring and sleep. Generally speaking, reducing energy is the main purpose of SMAC protocol. It has better energy-saving performance [8–10] than IEEE 802.11. However, the node sometimes is not able to transmit data in time because of the periodic sleep mechanism in this protocol. Consequently, the energy utilization and throughput of the network will decrease. In order to solve the problem mentioned above, an improved protocol–ACWM is introduced in this paper. The rest of this paper is organized as follows: the improved SMAC protocol with adaptive contention window mechanism (ACWM) is proposed in Sect. 2. Section 3 will introduce the simulation of ACWM and SMAC via the software NS2 in a scenario of dense node topology. Moreover, the simulation results of the two mechanisms are always compared and analyzed in this section. Section 4 will draw a conclusion of this paper.

2 Adaptive Contention Window Mechanism for SMAC

To overcome the shortcomings of SMAC, this paper proposes an adaptive contention window mechanism for SMAC. ACWM measures the network channel occupancy, fairness with conflict free ratio and node channel occupancy which is used to adjust contention window adaptively.

2.1 Introduction of ACWM

Random backoff mechanism with fixed contention window is widely used in SMAC protocol. Therefore, according to the deficiency of SMAC backoff mechanism, this paper proposes a mechanism called Adaptive Contention Window Mechanism (ACWM). This mechanism adjusts contention window by introducing two parameters to measure the state and the fairness of the network. In order to reflect the congestion of the network, ACWM introduces a parameter called *cf* which represents the occupancy of the channel. *cf* is the ratio between the conflict time and the free time of each node. By measuring the occupancy of channel to determine the current network state, then we can select a reasonable competition window, so as to improve the overall network throughput. Meanwhile if fairness of the network is considered in practical application, we can improve the overall performance. And then, we use C_0 to measure the fairness among nodes and balance the time when nodes occupy channel. Finally, two values are used to measure the current network load and fairness. According to the dynamic changes

of the network, the competition window is adaptively adjusted to improve the throughput, reduce the network delay and minimize the additional energy consumption caused by conflicts.

(1) **Calculation of cf and C_0**

In order to obtain a better network performance, this paper introduces a parameter that can be used to reflect the current state of the network—conflict free ratio cf. The definition is as follows: cf refers to the ratio of the average conflict time and the average idle time in a synchronous cycle.

$$cf = \frac{t_{col}}{t_{free}} \tag{1}$$

where t_{col} is the average conflict time in the data transmission process, t_{free} is the average idle time length.

From formula (1) we can see that, when cf has a small value, t_{col} is much smaller than t_{free}. It means that the nodes need to wait for a long time to access the channel, which causes the waste of channel resources. Then the contention window should appropriately reduce the waiting time of the node. And when cf take larger value, t_{col} is much greater than t_{free}, which means the probability of the occurrence of conflict is very large and the channel is very congested. There are many nodes competing the channel at the same time. So we can adaptively increase the contention window to reduce the occurrence of conflicts according to the network state.

Conflict free ratio can reflect the current network state, and we can use it to adaptively adjust the contention window. But in practical application, it may lead to certain unwanted results that some nodes with heavy load occupy the channel for a long time. When the number of routing network is small, these nodes become the bottlenecks of the network. Because other nodes cannot send and receive data and this will reduce throughput of network. Therefore, considering the current network situation, if fairness is considered in this situation, the overall network performance will be better. Therefore, this paper introduces a parameter to measure fairness among nodes—node channel occupancy ratio, which is defined as follows:

Node channel occupies ratio—C_0, which means in a synchronous period, the ratio between the total time that the nodes occupy channel and the total time that the channel occupied by the other neighbor nodes.

$$C_0 = \frac{t_{ocp}}{t_{other}} \tag{2}$$

where t_{ocp} is the time length that channel occupied by the node, t_{other} is the time length that other nodes occupy the channel. Before a node begins to transmit data, it will start competing the channel first. If successful, it will occupy the channel for a period of time. If not, it is likely to receive the control frames sent by other node. Therefore, the node acquires that the channel is occupied by other node in the network.

In ACWM, according to the current network state and certain fairness, the conflict free ratio and the node channel occupancy ratio are used to adaptively adjust the contention window. This mechanism will calculate different time parameters in a full transmission cycle. Situations that may happen in a complete transmission process are shown in Fig. 1.

Figure 1 shows that in a complete transmission process, the node will back off for some time before competing channel. t_{free} is the backoff time of a node before the competition. t_{ocp} represents the time that the node occupies the channel in the transmission process if there is no conflict or if there is a conflict in a period of time after the transmission. t_{col} represents the conflict time in the transmission process. If there is no conflict in the transmission process, t_{col} is zero. If the node monitors that the channel is occupied by another node before the transmission, then t_{other} is used to represent the time that the channel is occupied by the other node.

In order to prevent the network from changing the contention window frequently, the adaptive contention window is adjusted at the end of each cycle. So the nodes will calculate t_{free} and t_{col} in the transmission process of the network. When the node receives the ACK of last segment or maximum number of retransmission has been reached in the transfer process of RTS or DATA, it will calculate the intermediate data of conflict free ratio (cf_{meta}) based on t_{free} and t_{col}. Because when the above 3 situations appear, it means that the node has completed data transmission or the data transmission has ended. Choosing the last ACK is because that SMAC protocol divides data into segments, and each segment contains an ACK reply. Data transmission will not end until the last ACK is received. The calculations of t_{ocp} and t_{other} are different. They are calculated in each listen cycle and the value of C_0 is calculated at the end of a synchronous period.

In the calculation of the conflict free ratio, the parameter θ is introduced in a weighted approach. Then the relationship of the current transmission and the previous transmission is considered in the calculation. Then the value of cf is more meaningful.

Specific calculation:

$$cf_{meta}^{(n)} = \theta \times cf_{meta}^{(n-1)} + (1-\theta) \times \frac{t_{col}^{(n)}}{t_{free}^{(n)}} \tag{3}$$

$$fc = E[cf_{meta}^{(n)}] = \frac{cf_{meta}^{(n)}}{syncperiod} \tag{4}$$

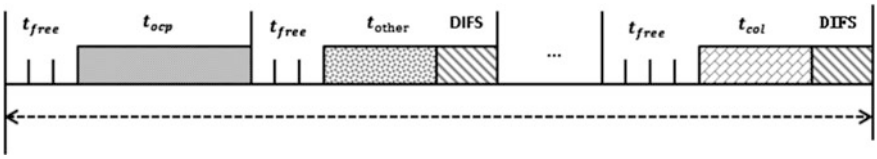


Fig. 1 Status of the complete transmission process

Node channel occupancy ratio C_0:

$$C_0 = \frac{t_{col}^1 + t_{col}^2 + t_{col}^3 + \cdots + t_{col}^{syncperiod}}{t_{free}^1 + t_{free}^2 + t_{free}^3 + \cdots + t_{free}^{syncperiod}} \tag{5}$$

(2) **Adaptive adjustment of contention window**

Thus, at the end of a synchronous period, nodes will calculate C_0 and cf according to the statistical parameters and then adaptively adjust the contention window according to the following mechanism. Specific network conditions are considered in the calculation of cf and C_0, such as the number of network nodes and routing choices. The border nodes are important for synchronization in the network. The nodes that occupy the channel for a long time may become the bottlenecks of the overall network. So fairness factor should be taken into consideration. When the node number of network is large, nodes with heavy load should transmit data first, that means the network state should be considered first. Algorithm 1 is the process of ACWM.

Algorithm 1 ACWM

```
inline double adjust_cw(double fc)
{
if (cf ≥ B + ω) then
   CW_data = int(min(CW_data * (α + θ), CWmax));
else if (cf < B + ω && cf ≥ C + ω) then
        CW_data = int(min(CW_data + 1), CWmax);
else if (cf < C + ω && cf ≥ C − ω ) then
        CW_data = CW_data;
else
        CW_data = int(max(CW_data * (α − θ), CWmin));
end if
if (t_ocp == 0 && t_other ≠ 0) then
   CW_data = int(max(CW_data/α, CWmin));
else if (t_ocp ≠ 0 && t_other == 0) then
        CW_data = int(min(CW_data*α, CWmax));
else if (t_ocp ≠ 0 && t_other ≠ 0) then
        if (t_ocp/t_other > β)
           CW_data = int(min(CW_data + φ, CWmax));
        else if (t_ocp/t_other ≤ β&&(t_ocp/t_other > 1/β))
                   CW_data = CW_data;
        else CW_data = int(max(CW_data − φ, CWmin));
else
     CW_data = CW_data;
end if
```

As shown in Algorithm 1, in order to balance the fluctuation of the network, ω and other regulatory factors are introduced in the process. When $cf \geq \mathrm{B} + \omega$, the conflict time of network is longer than free time. Network congestion occurs and the probability of conflicts is increasing. At this time the contention window should be increased to reduce the probability of the node accesses channel, thereby the occurrence of conflicts is reduced and the throughput is increased. When $\mathrm{C} + \omega \leq cf < B + \omega$, it is considered that the network conflicts exist but not very serious, so the increasing of contention window should not be too great. When $\mathrm{C} - \omega \leq cf < \mathrm{B} - \omega$, it is considered as that the *cf* value is close to the optimal *cf*, the network is in very good state, the node is using channel reasonably, therefore the contention window should not be changed. When $cf < \mathrm{C} - \omega$, it is considered that *cf* is smaller than the optimal *cf* value. The free time of the node is too long. The channel is not used effectively. So that we should reduce the contention window to increase the probability for the nodes access channel, then throughput is finally improved.

When t_ocp == 0&&t_other ≠ 0, it is considered that the channel is occupied by other nodes. So the contention window should be reduced to increase the probability for accessing the channel. And when t_ocp ≠ 0&&t_other == 0, the time that current node occupies the channel is too long. In order to create opportunities for other nodes to transmit data, we should increase the contention window to reduce the probability for current node accesses channel. Then other nodes also have the opportunity to use the channel, therefore the throughput is improved. And when both values are not zero, then the channel is normal. When $C_0 > \beta$, the time that current node occupies the channel is too long, then we should increase the contention window slightly to reduce the probability of the channel occupied by the node. When $1/\beta < C_0 < \beta$, the node occupies channel reasonably, the network state is fairly good, so the contention window does not need any adjustment. When $C_0 < \beta$, the time that other node occupies the channel is too long, so it is necessary to reduce the contention window slightly to increase the probability for the node accesses channel.

3 Simulation Results and Performance Analysis

In this section, ACWM protocol is compared with SMAC. Dense node topology is used to simulate the two schemes in this paper. This paper simulates dense node topology by using the topology in Fig. 2. It consists of 9 nodes. Distance between nodes is 150 m. Simulation time is 1000 s. After 50 s from beginning, node 0 sends a data packet to node 8 and ends at 800 s. And after 100 s from beginning, node 6 sends a data packet to node 2 and ends at 850 s.

Figure 3 shows that the throughput varies when sending interval increases. As shown in the figure, the throughput of the two protocols reduces when sending interval increases. That is because the number of hops from the source node to the destination node is small in dense topology. It effectively reduces the probability of packet loss during transmission.

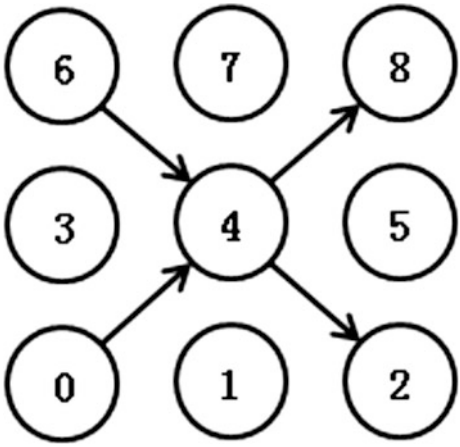


Fig. 2 Dense node topology

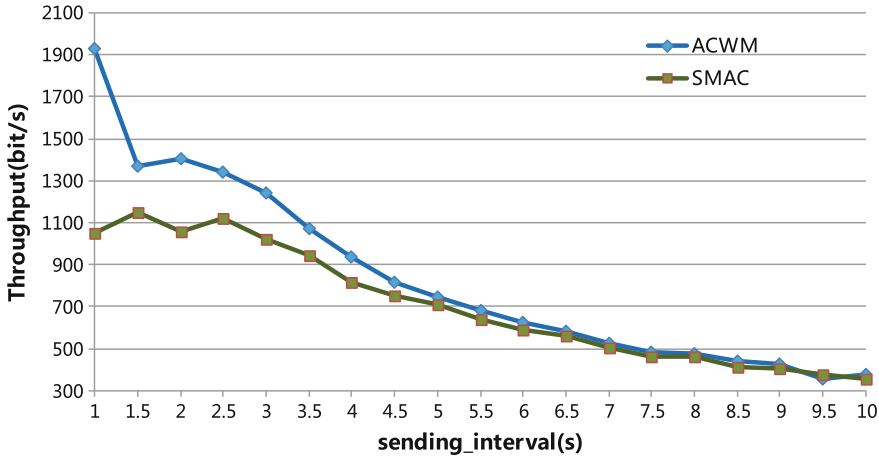


Fig. 3 Throughput variation under different sending interval

Therefore, the throughput is better when the load is heavy. The ACWM protocol has better throughput in the heavy load area. And with the reducing of the load, the increase in throughput is smaller. ACWM protocol adjusts the contention window reasonably when the synchronous cycle is long, so the overall performance is better. ACWM improves the synchronization between nodes and guarantees a large amount of data transmission with a small sending interval. It has no packet loss and can obtain better throughput.

As we can see from Fig. 4, the average delay of the two mechanisms are decreased with the increase of the sending interval. As the number of hops from the source node to the destination node is reduced greatly, time delay will not increases

due to packet loss when the load is heavy. For ACWM, the delay is reduced because ACWM can adjust the contention window adaptively according to the state of the channel occupancy. It can reduce the conflicts among the nodes, therefore the delay of the whole network is reduced. When the load is heavy, ACWM reduces packet loss because of small clock drift. Large amount of data retransmission is not needed. And in the heavy load area, because it can predict the channel state and select a reasonable contention window, therefore ACWM has a better performance in time delay. In the light load area, ACWM reduces the sleeping delay and the waiting delay. So the delay of ACWM is smaller than SMAC in the light load area.

Figure 5 shows the energy consumption of unit data increases with the increase of sending interval. This is because the smaller the interval is, the node can send more packets. But with the increase of sending interval, the node spends a lot of energy in idle listening, so the energy consumption of unit data is increasing. For ACWM, in the heavy load area, the probability that contention window remains maximum is higher. The reasonable contention window reduces the probability of packet retransmission. While in the light load area, due to the detection of the channel, the time of idle listening is reduced. Therefore, the energy efficiency of ACWM is lower than SMAC protocol.

Figure 6 shows that the energy consumption of the two protocols decreases with the increase of the sending interval. This is because the throughput of the network is gradually reduced. While in the heavy load area, as we know from the previous analysis, the throughput of ACWM protocol is greatly increased, therefore the energy consumption of ACWM protocol also increases. Meanwhile in light load area, ACWM adjusts the contention window to reduce the listening time and speeds up the transmission of data packets. So ACWM takes less energy to achieve the same throughput.

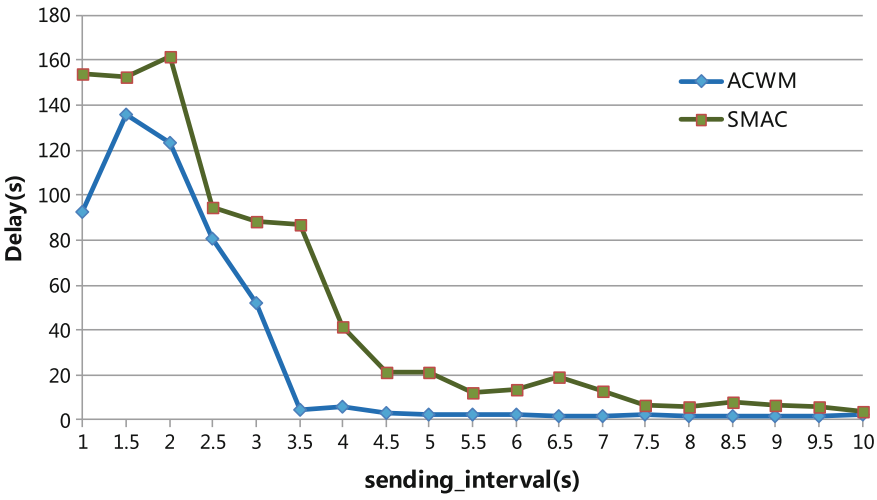


Fig. 4 Delay variation under different sending interval

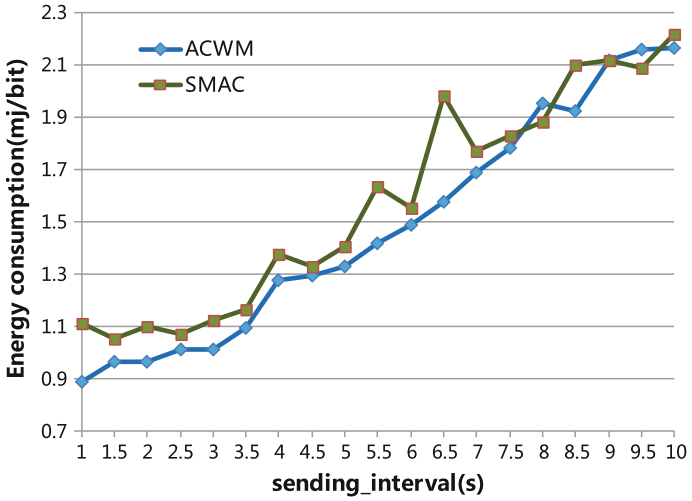


Fig. 5 Energy efficiency variation under different sending interval

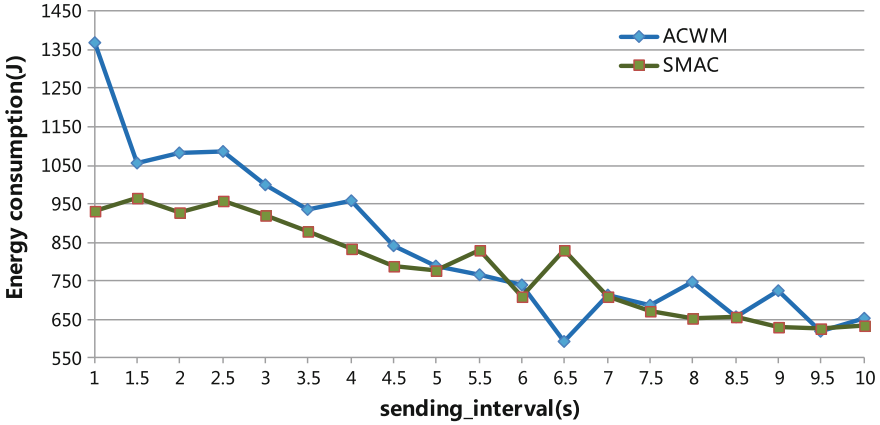


Fig. 6 Energy consumption variation under different sending interval

4 Conclusion

This paper proposes an adaptive contention window mechanism to improve the performance of SMAC protocol in wireless sensor network. This mechanism adjusts contention window by introducing two parameters to measure the state and the fairness of the network. In order to reflect the congestion of the network, ACWM introduces a parameter *cf* which represents the occupancy of the channel. *cf* is the ratio of the conflict time to the length of free time of each node. We use C_0

to measure the fairness among nodes and balance the time that node occupies channel. According to the dynamic changes of the network, the competition window is adaptively adjusted to improve the throughput, reduce the network delay and minimize the additional energy consumption caused by conflicts.

The comparison experiments of ACWM protocol with SMAC protocol via simulation software NS2 in a scenario of dense node topology. Simulation results show that ACWM protocol achieved better performance with higher throughput, lower average delay and higher energy utilization efficiency.

Acknowledgements This paper is financially supported by the program of Guangdong Science and Technology (Grant No. 2015A010103007).

References

1. Mendes LDP, Rodrigues JJPC (2011) A survey on cross-layer solutions for wireless sensor networks. J Netw Comput Appl 34(2):523–534
2. Yick J, Mukherjee B, Ghosal D (2008) Wireless sensor network survey. Comput Netw 52 (12):2292–2330
3. Egea-López E, Vales-Alonso J, Martínez-Sala AS et al (2008) A wireless sensor networks MAC protocol for real-time applications. Pers Ubiquit Comput 12(2):111–122
4. Yang O, Heinzelman W (2012) Modeling and performance analysis for duty-cycled mac protocols with applications to S-MAC and X-MAC. IEEE transactions on mobile computing
5. Feng H, Ma L, Leng S (2010) A low overhead wireless sensor networks mac protocol. Computer engineering and technology (ICCET), 2nd international conference on, 2010
6. Heidemann J, Ye W (2004) Medium access control in wireless sensor networks. Wirel Sens Netw 73–91
7. Estrin D, Ye W, Heidemann J (2004) Medium access control with coordinated adaptive sleeping for wireless sensor networks. Networking, IEEE/ACM transactions on, 2004
8. Wang C, Chen Y, Hou Y (2013) The analysis and improvement of SMAC protocol for wireless sensor networks. Mobile ad-hoc and sensor networks (MSN), 2013 IEEE ninth international conference on, 2013
9. Bianchi G (2000) Performance analysis of the IEEE 802.11 distributed coordination function. Selected areas in communications, IEEE journal on, 2000
10. Xu S, Saadawi T (2001) Does the IEEE 802.11 MAC protocol work well in multihop wireless ad hoc networks? Communications Magazine, IEEE, 2001

Performance Analysis of Encryption Algorithms with Pat-Fish for Cloud Storage Security

M. Usha and A. Prabhu

Abstract In the era of the Cloud, a remote user connected from anywhere, anytime is provided with any form of access to the storage services. Internet of things is growing rapidly in all aspects and Cloud storage has become an essential aspect in the day to day life. Data Science and Big data analytics, and other technologies use the smart devices like personal Laptop, tablet and smartphone and enterprises are interested to store data and the transactions in Cloud data centres. However, cloud storage needs a secured transaction and authentication system. Cloud service providers need to provide high security at their storage level. Our approach combines Blowfish algorithm and Pattern matching to secure the data in cloud data storage. Pattern matching algorithm is the best algorithm in terms of time complexity and space complexity. Blowfish algorithm is a 16-round Fiestal algorithm, which is used to encrypt and decrypt the input files. This paper evaluates the hybrid Pat-Fish algorithm with DES, RSA, and Blowfish methods on text files. The standard evaluation parameters namely encryption time and decryption time are taken for performance comparison. This Pat-Fish approach yields less time for encryption and decryption compared to DES, RSA and Blowfish algorithms. This method is suitable for cloud storage to store the client data with security.

Keywords Encryption · Decryption · DES · Blowfish · RSA
Pat-Fish · Pattern matching algorithm

M. Usha (✉)
Sona College of Technology, Salem, Tamil Nadu, India
e-mail: usha@sonatech.ac.in

A. Prabhu
PSG Institute of Technology and Applied Research, Coimbatore
Tamil Nadu, India
e-mail: prabhuak@gmail.com

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_11

1 Introduction

Every day new technologies are mushrooming as per the industry needs and spreading into all the places, to reduce the human effort and to improve seamless services. Industries like to have much more technology [1] in their business world and have a large number of data to be stored. Recent days "Cloud computing" is a trendsetter which supplements all the existing techniques by storing huge volumes of data in a secure way [2].

1.1 Cloud Computing

Traditional applications are so complicated for the user and industry, and a partial amount is to be invested in the hardware and software to run a business. Cloud computing overcome all barriers for storage of the traditional methods. Cloud-based applications can run through the browser with the help of the internet. It has various benefits like self- serving provisioning, elasticity, pay per use, workload resilience, and migration flexibility.

Types of Cloud Computing Services: The cloud offers three types of services. *Infrastructure as a service.* Example: networks, storage and operating systems.

Platform as a service. Example: Web services, application development beds and search engines.

Software as a service. Example: Internet web browser and applications.

Cloud Computing services are available across the globe. Based on the usage and location, it's divided into many types such as Private, Public Cloud and Hybrid Cloud.

1.2 String Matching Algorithms

The World is full of textual information. To get the information using textual queries or image mapping from websites, books and newspapers are needed. In the view of computer science, search engines use many string algorithms. Matching a similar text is called string matching. There are various types of string matching algorithms available [3].

- Naïve string search algorithm
- Rabin–Karp string search algorithm
- Boyer–Moore string search algorithm
- Knuth–Morris–Pratt algorithm
- Bitmap algorithm

Every algorithm has unique features and capabilities. Mathematical implementations calculate the exact patterns in the millions of the texts. Many attempts have been made to exploit the different pattern matching algorithms in cloud and communication networks [4–7].

1.3 Cryptography Algorithms

Cloud computing faces a lot of security issues in the storage and network platforms. Cryptography algorithm is the process to protect data that are sent through communication networks [8]. It is an art of hiding information by encrypting the message. The readable message is converted into an unreadable format is called cipher-text (Fig. 1).

The fundamental set of cryptography algorithms can be divided into three groups:

- Symmetric encryption algorithm
- Asymmetric encryption algorithm
- Hash functions

2 History of Existing Techniques

Symmetric encryption (Private Key): It's a unique key to encrypt and decrypt the data. Asymmetric encryption (Public Key): It is also called as a public key encryption algorithm. The algorithm uses a pair of keys that help to do encryption and decryption.

2.1 Popular Encryption Algorithms

Many symmetric and asymmetric cryptography algorithms have been proposed (Fig. 2) in literature. *Blowfish Encryption Algorithm*. Blowfish [9] was designed in 1993. It is a symmetric key block cipher. It has a key length from 32 to 448 bits and block size of 64 bits. It's a Feistel network. It was designed by Bruce Schneier to replace the DES or IDEA algorithms. Blowfish is a license-free algorithm.

Data Encryption Algorithm (DES). DES [10] is a symmetric key block cipher algorithm and it was discovered in 1972 by IBM. This method was accepted by the United States of America. The key length is 64 bits (56 + 8 parity bits) and block size is 64-bit length. DES handles different mode operations such as CBC, ECB, CFB and OFB. It has some failures when a weak key is used.

Rivest-Shamir-Adleman (RSA) Algorithm. RSA is founded by Rivest, Shamir and Adelman in 1977 [11]. It's an asymmetric cryptographic algorithm and generates two keys: Public Key and Private key. The public key and private key is used to encrypt and decrypt the message respectively.

Fig. 1 Encryption process

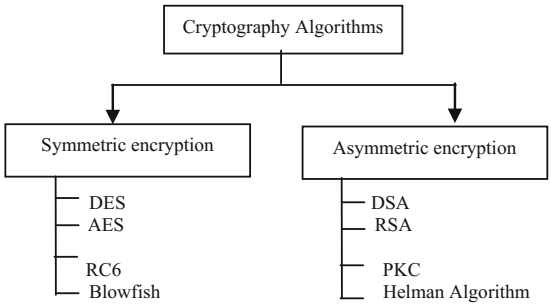


Fig. 2 Types of cryptography algorithms

RSA algorithm consists of three steps:

- Key generation is used to encrypt and decrypt data.
- Data encryption is converting the plaintext to ciphertext.
- The third step is converting the chipper text to plain text.

Key size is 1024 to 4096 bits. RSA is one of the first algorithms and is broadly used for secure data transmission.

Advanced Encryption Standard (AES) algorithm. AES is a symmetric key block cipher algorithm submitted by Joan Daemen and Vincent Rijmen in 1998. AES algorithm supports multiple combinations of key with the length of 128, 192, and 256 bits. 128-bit data are split into basic operational blocks and it's are considered as an array of bytes formed as a 4×4 matrix which is also called as states. For encryption, 10, 12, 14 iterations are used with the key length of 128,192 and 256-bits. Every round of AES uses the order and substitution methods in the network, it is suitable for both hardware and software implementation. Since AES requires more processing power [12] it is not taken for comparison in this work.

3 Evaluation Metrics

The following evaluation metrics are used and compared with the existing techniques which are normally adopted by researchers [13].

i. Encryption Time
ii. Decryption Time
iii. Throughput

3.1 Encryption Time

The number of cycles executed to convert from plain text into cipher text is called encryption time. Encryption time is based on key length and input file size. In our

experimental setup, encryption time is measured in milliseconds and based on the performance of the system.

3.2 Decryption Time

The time taken to convert plain text from cipher text is called decryption time. The decryption time is normally very less than the encryption time. Decryption time is used to measure the performance of the system and it's quite faster than encryption algorithm. Decryption time is measured in milliseconds.

3.3 Throughput

Throughput is calculated based on the file size and execution time. The number of resources (MB/milliseconds) executed at a particular time (MB/milliseconds) is called throughput using the following formula:

$$\text{Throughput} = \text{File size}/\text{Execution time}$$

4 Proposed Method

To enhance the Cloud storage security, a combined approach, Pat-Fish is proposed which is a combination of Blowfish with Pattern Matching. The block diagram of the proposed method is shown in Fig. 3. The Pat-Fish method has two phases:

i. Authentication Phase
ii. Cloud Data encryption Phase

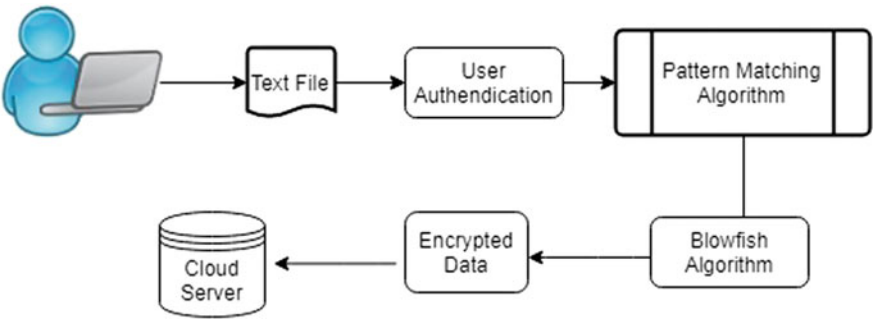


Fig. 3 The block diagram of proposed method

4.1 Authentication Phase

Authentication is an important process to allow the authorized users to access the application. In this process, data sets are compared with the database of authorized users' information within an authentication server.

If the user credentials are matched with the existing data set, then the user is permitted to access the application. During authentication, the user or computer has to get the authorization from the server or client.

4.2 Cloud Data Encryption Phase

In this phase, two existing algorithms, namely the Pattern Matching algorithm and Blowfish algorithm are combined together for fast encryption. The input data is subdivided column-wise using pattern matching technique, the pattern matching takes into account a data 'D' of length 'n' and a pattern of length 'm' with sub-dividing the entire data into the columns. The portioned data is then encrypted with the help of the Blowfish algorithm.

Blowfish Encryption Algorithm. Blowfish is a 64-bit block cipher encryption algorithm. There are two ways to implement the blowfish algorithm through via software and hardware (Fig. 4).

This structure is known as a Fiestal network. Blowfish has 16- Fiestal rounds. The block size is divided into two parts such as left-hand side 32-bits and right-hand side 32-bits. The input divides a 32-bit into four 8 bytes. The output results are added, XORed and swapped in this function. Blowfish algorithm has a set of procedure used to decrypt as well as encrypt the input text. It has the following two functions.

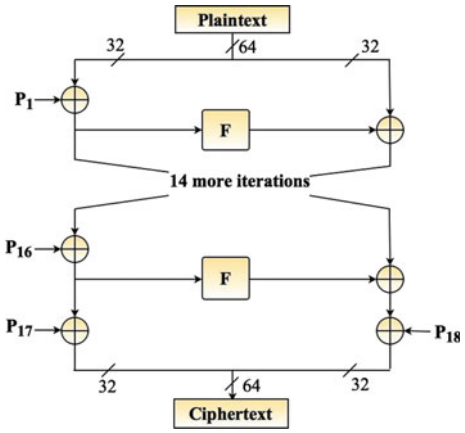


Fig. 4 Graphical representation of blowfish Algorithm

- Key Expansion
- Data Encryption

Key-expansion. The 448 bit key is separated into 4168 bytes. The P-array consists of 18, 32 bits (P1, P2… P18). The following steps are used to calculate the subkeys:

1. P-array and four S-boxes are initialized. The string contains the hexadecimal values obtained from Pi. P1, P2, P3, P4 array values are generated by Key Generator.
2. Right-hand side 32-bits are XORed with P1 of the key, left-hand side 32-bits are XORed with P2 of the key. Encrypt the all-zero string with the Blowfish algorithm.
3. Exchange P1 and P2 (2).
4. Encrypt the exchanged values (3) using the algorithmic procedure with the subkeys.
5. Exchange P3 and P4 (4).
6. Continue the process and concatenate x = mL and mR values.

Data Encryption. The 64-bit input data is divided into two 32-bit halves, which are labeled as the left halves (LH) and right halves (RH). The Blowfish algorithm executes the first 32-bit left half and the P-array performs the XOR function. The outcomes are furnished to the function (FmL). Subsequently, the XOR function is executed for both left halves and the next 32-bit right halves elegantly. This is followed by the swapping of both outcomes. The rest of the round continues until it reaches 16th round.

Four 32-bit S-Boxes consist of 256 entries each:

S1 1, S1, 2 … S1, 256
S2 1, S2, 2 … S2, 256
S3 1, S3, 2 … S3, 256
S4 1, S4, 2 … S4, 256

Process of F_{ml} function. The F_{ml} function executes 32-bit S-boxes, with each one encompassing 256 entries. In this Blowfish technique, the first 32-bit left half is subdivided into four 8-bit blocks such as m, n, o and p. The Eq. (1) gives F(mL) function in detail [9].

$$F(mL_H) = ((S_{b1,m} + S_{b2,n} \bmod 2^{32}) \oplus S_{b3,o}) + S_{b4,p} \bmod 2^{32} \quad (1)$$

Data Encryption algorithm:

Step 1: Get the input file ‘m’
Step 2: ‘m’ is divided into two 32-bit, named as mL, mR
Step 3: For i = 1 to 32, and do the XOR operation with the Key (Pi)

$$mL = mL \text{ XOR } Pi$$

$$mR = F(mL) \text{XOR } mR$$

Step 4: Swap mL and mR

5 Results and Discussion

The experiments were run in NetBeans with CloudSim. The performance of the algorithm compared for encryption time and decryption time with the DES, RSA and Blowfish algorithms. The Pat-Fish approach encryption runtime is compared in Fig. 5.

The bar charts in Fig. 5 show the encryption times for the various file sizes. It is shown that whenever the file size increases the encryption time also increases linearly (Fig. 6).

The decryption time is always low compared with the execution of encryption time. This approach proves the decryption time is lower than the encryption time.

The throughput and average run time of Pat-Fish algorithm are calculated using the encryption time with various files sizes as given in Table 1.

The Tables 2 and 3 show the performance comparison of Pat-Fish's encryption and decryption time using 1, 2, 5 and 10 MB text files. Based on these values Pat-Fish algorithm is far better than DES, RSA, and Blowfish algorithms. Normal encryption method encrypts the text only, but encryption algorithm separates the text file using a pattern matching algorithm. The fragmented file encrypted by Blowfish algorithm and cipher-text file stored in the Cloud server. The authors have

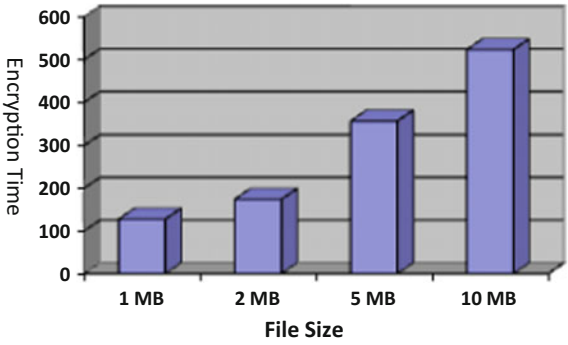


Fig. 5 Pat-Fish algorithm encryption time graph

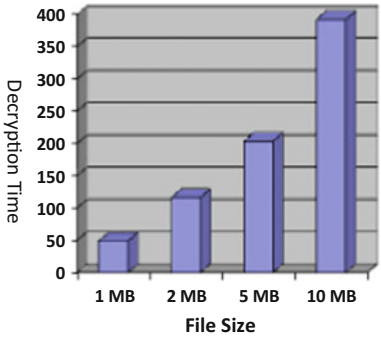


Fig. 6 Pat-Fish algorithm decryption time graph

Tabel 1 Execution time for Pat-Fish

File size	Pat-Fish algorithm (in ms)
1 MB	127.5
2 MB	173.9
5 MB	391.70
10 MB	523.42
Average run time	295.30
Throughput (MB/s)	14.79

Table 2 Performance comparison of encryption time

File size	DES (in ms)	RSA (in ms)	Blowfish (in ms)	Pat-Fish (in ms)
1 MB	136.2	425.6	133.2	127.5
2 MB	269.6	710.9	192.6	173.9
5 MB	665.4	1710.9	373.6	356.4
10 MB	1325.2	2352.24	752.80	523.42
Average run time	599.10	1299.91	363.05	295.30
MB/s	7.51	3.46	12.39	14.79

Table 3 Performance comparison of decryption time

File size	DES (in ms)	RSA (in ms)	Blowfish (in ms)	Pat-Fish (in ms)
1 MB	144.6	375.80	50.2	48.1
2 MB	269.6	541.68	126.3	107.8
5 MB	690.9	995.70	210.7	187.3
10 MB	1874.35	1897.35	521.47	392.18
Average run time	744.86	952.63	227.16	203.84
MB/s	6.04	4.72	19.80	24.47

proposed the time complexity reduction using finite automata [14] in cloud virtualization in their previous work. Combined with the security aspect proposed in this paper, a framework is now established for cloud data centres.

6 Conclusion

In this research paper, DES, RSA and Blowfish algorithms are compared with Pat-Fish. The results show that the Pat-Fish approach's performance is better than other encryption algorithms. 1 MB, 2 MB, 5 MB and 10 MB files are evaluated for the encryption and decryption times. The proposed Pat-Fish method improves the performance by 11% compared to Blowfish technique.

In future, this algorithm will be tuned to improve the efficiency by combining the data center selection [15] too. Space complexity analysis of Pat-Fish and a separate hardware implementation are to be taken up in future.

References

1. Mohammad A (2014) Cloud of things: integrating internet of things with cloud computing and the issues involved. In: Proceedings of 11th IEEE international Bhurban conference on applied sciences and technology IBCAST, pp 14–18
2. Christos S (2016) Secure integration of IoT and cloud computing. Future Generation Comput Syst 78:964–975
3. Priyadarshini P (2016) A comprehensive evaluation of cryptographic algorithms: DES, 3DES, AES, RSA and Blowfish. In: International conference on information security & privacy—Elsevier Procedia computer science, vol 78, pp 617–624
4. Kavitha P (2016) Anomaly based intrusion detection in WLAN using discrimination algorithm combined with naïve Bayesian classifier. J Theor App Info Technol 62(1, 3):646–653
5. Usha M (2017) Anomaly based intrusion detection for 802.11 networks with optimal features using SVM classifier. Wireless Netw 23(8):2431–2446
6. Abdel-Karim Al Tamimi (2008) Performance analysis of data encryption algorithms, available at https://www.cse.wustl.edu/∼jain/cse567-06/ftp/encryption_perf/
7. Prabhu A (2018) A Survey on the pattern matching algorithms in cloud computing. In: National conference on emerging trends in signal and image processing, communication, VLSI design and nano technology (NCSICVN-18), pp 38–43
8. Hussain I (2013) Improved approach for exact pattern matching (Bidirectional exact pattern matching). IJCSI Int J Comput Sci Iss 10(3):59–65
9. Schneier B (1993) Description of a new variable-length key, 64-Bit block cipher (Blowfish). In: Fast software encryption, cambridge security workshop proceedings, Springer, pp 191–204
10. Data Encryption Standard (1999) Federal Information Processing Standards Publication No. 46, National Bureau of Standards
11. Rivest RL (1978) A method for obtaining digital signatures and public-key cryptosystems. Commun ACM 21(2):120–126
12. Sanchez-Avila C (2001) The Rijndael block cipher (AES proposal): a comparison with DES, Security Technology, In: 2001 IEEE 35th international carnahan conference on. IEEE, pp 229–234
13. Bijoy K (2013) Designing and performance analysis of a proposed symmetric cryptography algorithm. In: International conference on communication systems and network technologies-ieee computer society, pp 453–446
14. Prabhu A (2015) Nondeterministic finite automata to reduce the time complexity in cloud virtualization. Int J Appl Eng Res 10(37):28021–28025
15. Prabhu, A A Secured best data centre selection in cloud computing using encryption techniques. Int J Business Intell Data Mining (Accepted for publication). https://doi.org/10.1504/ijbidm.2018.10007299

An Indoor Positioning Scheme Exploiting Geomagnetic Sensor of Smartphones

Young Uk Yun and Youngok Kim

Abstract In this paper, an indoor positioning scheme using geomagnetic sensor of smart devices is proposed. The proposed scheme is based on a fingerprint technique that can estimate the position of the user by comparing the sensed values of geomagnetic sensor with a pre-collected database. To investigate the characteristics of geomagnetic sensor, we have conducted experiments by using only geomagnetic sensors among the built-in inertial sensors of smartphones. Three different experiments were conducted, and firstly we collected the measurements of geomagnetic sensor at the fixed points throughout a day to identify changes in values and patterns. Secondly, 9 points were selected and measured where feature of the geomagnetic value has, and thirdly geomagnetism data was collected while moving a specific route. With the results of experiments, we analyzed the collected data and proposed a positioning scheme to estimate the location by using the geomagnetic values. According to the results, the data measured throughout a day and the data collected at a specific point have a problem that can cause errors, because the measured values change with time. Meanwhile, it is shown that the measured data has a unique pattern while it moves a specific route and there is hardly changes in geomagnetism patterns as time is passed. Therefore, it is concluded that this feature can be utilized in indoor positioning scheme exploiting geomagnetic sensor, rather than the simple fingerprint scheme based on a static database of specific points.

Keywords Positioning · Geomagnetic sensor · Smartphone

1 Introduction

Indoor positioning system (IPS) has emerged as a core technology with many issues for 4th industrial revolution, smart factory, cyber-physical system, on-demand, and pin-tech in recent years. As the result, the demand for IPS is rapidly increasing in

Y. U. Yun · Y. Kim (✉)
Electronic Engineering Department, Kwangwoon University, Seoul, South Korea
e-mail: kimyoungok@kw.ac.kr

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_12

both outdoors and indoors. Global positioning system (GPS) for outdoor usage has been highly developed thanks to satellites for GPS and has brought a lot of convenience to people. Meanwhile, the IPS for indoor usage requires relatively higher precise location accuracy than that for outdoors.

In literatures, various IPS for indoors have been studied and introduced [1–3]. The indoor IPS can be divided according to the sensors and techniques used for position estimation. There are positioning techniques utilizing wireless communication systems such as mobile systems, Wi-Fi, Bluetooth beacon, Ultra-wide band, and Zigbee [4–6]. Vision-based systems through image processing [7] and sensor-based systems, such as geomagnetic sensors, and inertial sensors are introduced [8]. These systems have their own unique characteristics, and there are advantages and disadvantages for different usage and indoor environments.

In case of high performance location accuracy, generally it costs a lot to build infrastructure for positioning purpose. In recent years, however, low-cost and infrastructure-free positioning schemes have been continuously researched and introduced. Especially, the indoor positioning scheme exploiting smartphones has been extensively researched because of the high penetration rate of users of smart phones recently. Since most smartphones have various built-in sensors, in this paper, we use only the geomagnetic sensor among the built-in inertial sensors of smartphones. Not only can we use the inertial sensor with a magnetic sensor to reduce the load on the device, but also it can be designed in stand-alone form to prevent exposure of personal information. It also can be used as a management and control type technology in conjunction with the network. In this paper, we also analyze the magnetic field patterns during the day, at specific points, and during the movements. According to the results, we propose an indoor positioning scheme exploiting the geomagnetic sensor of smart devices.

The rest of the paper is organized as follows. In Sect. 2, the system is briefly described. In Sect. 3, we performed three different experiments to investigate the characteristics of geomagnetic sensor and then an indoor positioning scheme exploiting the geomagnetic sensor of smart devices is proposed. In Sect. 4, concluding remarks are summarized with future works are discussed.

2 System Description

Earth has a huge magnetic field, which can be measured anywhere. However, this magnetic field can be severely distorted by the structure of the steel structure used in the interior construction, the electronic devices, and the power lines in the indoor space. Therefore, a unique magnetic field can be formed at a specific point of each indoor environment, and such a singular point can be utilized for the indoor positioning scheme. The proposed scheme is based on a fingerprint technique that can estimate the position of the user by comparing the sensed values of geomagnetic sensor with a pre-collected database (DB).

2.1 Fingerprint Method for Geomagnetism Value

The fingerprint technique is generally divided into two stages. The geomagnetism value is measured at a specific location in advance and build into a DB as an offline stage. In the next online stage, the estimation of the current position is performed by finding the position where the most similar data value appears through the comparison between the input data and the DB created in advance. To increase the accuracy of the position estimation, it is necessary to collect the offline data at many positions.

On the other hand, it needs to be noticed that the amount of computation cost can be increased because of the comparison with huge data. For computing the similarity between the input data and the DB, various schemes, such as Euclidean distances (ED), Manhattan distances, and Mahalanobis distances, are introduced for the Fingerprint algorithms [9]. In this paper, the ED scheme was used for position estimation, because it has relatively low computational complexity, and it is expressed as follows:

$$ED(\mathrm{x}, \mathrm{y}) = \sqrt{\sum_{i=1}^{n} (\mathrm{x}_i - y_i)^2} \tag{1}$$

where x_i means the i-th data of the pre-collected DB, y_i is the received data in real-time, and n is the total number of comparison objects.

2.2 Application for Smartphone

We developed an Android-based application to control the sensor of the smartphone. Generally, the embedded inertial sensors of smartphones consist of three types of sensor, which are Gyro, acceleration, and geomagnetic sensors, and all have 3-axis of x, y, z. Note that it also referred to as 9-axis sensor. In the proposed scheme, we use the RMS values of the x, y, and z axes of the geomagnetic sensor only. The data rate also can be controlled by using the application and the data are collected by adjusting the speed according to the experiment conditions.

3 Proposed Scheme and Experiment Results

A smartphone of S4 model of S company was used in the experiments and the smartphone application was developed independently for the experiments conducted. Experiments were performed at the 6th floor of a general office building in our University. The corridor was about 6 m wide and 105 m long as shown in Fig. 1. We performed three different experiments to investigate the characteristics of geomagnetic sensor and the positioning performance based on the geomagnetic sensor.

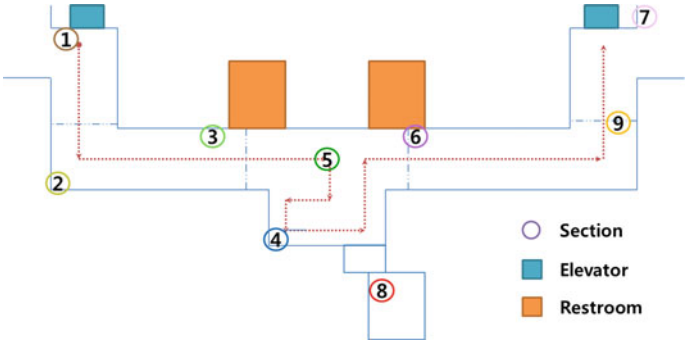


Fig. 1 Schematic of experiment space

The first experiment was performed to investigate the geomagnetic variation during a day. The second experiment was performed to check the coherence time of data measured at specific locations. The third experiment was performed to check the change of patterns of the geomagnetism values while the user is moving.

3.1 Analysis of Geomagnetic Variation for a Day

Experiment carried out at Section 8 of the Fig. 1. The smartphone was placed on the shelf and close to the window frame. The sample rate was 0.5 Hz, which totally collected 43,200 data per day. Figure 2 shows the magnitude variation of the geomagnetic values for 24 h at two different day. The horizontal axis indicates the time in hour and the vertical axis indicates the RMS value of the 3-axis geomagnetic values. As shown in the figure, we can observe that the RMS values change with respect to the hour as well as day. When it was measured on the window frame, the minimum value was about 39 and the maximum value was about 49. Also, it is shown that the difference of RMS values is large in terms of day, even if

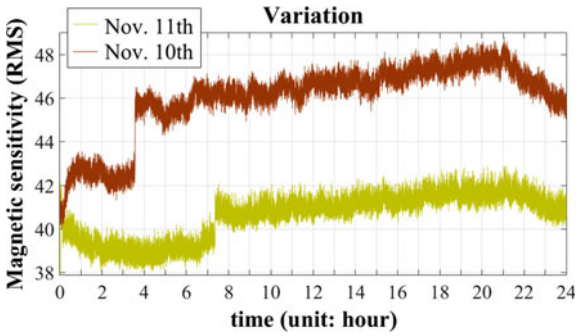


Fig. 2 Variation of geomagnetic values for 24 h at two different days

it is a similar time zone. Since the geomagnetic value changes with time, a large error can be occurred with the fingerprint method if the position is estimated by using the DB, which was collected in a short time period. As for the DB, therefore, data should be accumulated for a long time, and it needs to analyze the valid time for the geomagnetism data of the corresponding point and to update the DB continuously. Thus, there would be difficulties in maintenance, when this DB is used for the indoor positioning scheme.

3.2 *Analysis of Geomagnetic Pattern at Specific Points*

The experiments were conducted to check the coherence time at specific points, 1 to 9 points as shown in the Fig. 1. The representative locations were selected, such as the side of elevator, the front of iron gate, toilet, laboratory room, etc. The sampling rate was 100 Hz for 30 min at each point. A total of 180,000 samples were collected for 30 min. It was collected at a height of 1 m from the bottom and was measured by using a paper support to minimize the effects of distortion from the support.

Figure 3 shows the variation of geomagnetic values at specific points. The horizontal axis is time in minutes and the vertical axis is the RMS value of the 3-axis geomagnetic values. The values within 30 s at the beginning of the measurement can be considered as the results of operation or shaking of the stand. Therefore, the coherence time is considered with the measured data thereafter.

As shown in the figure, the RMS value was similar at positions 1, 2, and 8 and at positions 4, 5, and 6. For the point 8, the RMS value was decreased from 35 to 31 for 30 min, and it showed a change in the magnetic field within a short time. As for the position estimation by using the measured data at specific points, therefore, it is necessary to analyze the data after long time measurement. Thus, as mentioned in the previous subsection, there would be difficulties in maintenance also.

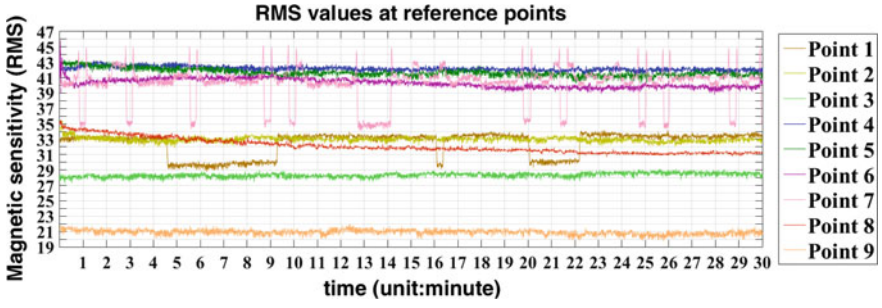


Fig. 3 Variation of geomagnetic values at specific points

3.3 Analysis of Geomagnetic Pattern for Movements

The experiments were conducted to check the variation of geomagnetic values as time passed, while the user was moving along a specific route. Data were repeatedly collected by moving along the same route at different time, and a total of 80 repeated experiments were performed. The specific route is indicated with the dotted-line as shown in the Fig. 1, and the total length of route is about 105 m. The sampling rate was set at 100 Hz, and the data was accumulated while the user moves with the screen of the smartphone up during the experiments.

Figure 4 shows a cumulative graph of 80 experiments. The horizontal axis is time in sample count and the vertical axis is the RMS value for the 3-axis geomagnetic values. Data was collected at 100 Hz, and thus the horizontal axis of 2000 means 20 s in terms of time. The red bar indicates the time when the direction changes. Although the speed of movement for each experiment is different, we can see a similar pattern. To verify the similarity of the patterns, cross-correlation was used to analyze the signals. Figure 5 shows a cross-correlation value between the

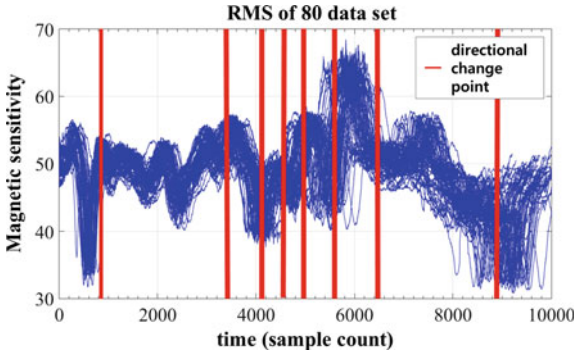


Fig. 4 Cumulative measured data of geomagnetic values during moving along a specific route

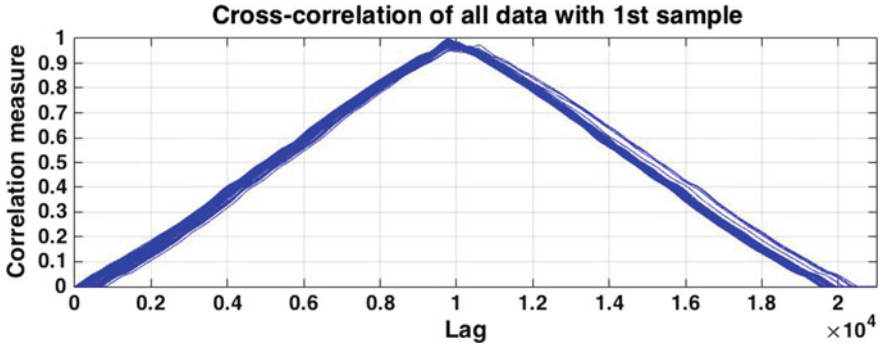


Fig. 5 Cross-correlation results between the first experimental data and the rest of 79 experimental data

first measured data and the rest of 79 data. The vertical axis represents the similarity, which is normalized to a maximum of 1, and the horizontal axis represents the time delay. The time delay is about 10 s, but the similarity is close to 1 that means two signals has almost same pattern. The time interval between the first measured data and the last measured data was about 14 h, but there was no remarkable difference in the data pattern. Through these experiments, it was verified that the possibility of extension to the indoor positioning scheme by using the pattern change of the geomagnetic value while the user is moving.

4 Conclusions and Future Works

In this paper, an indoor positioning scheme using geomagnetic sensor of smart devices was proposed. We performed three different experiments to investigate the characteristics of geomagnetic sensor and the positioning performance based on the geomagnetic sensor. The first experiment was performed to investigate the geomagnetic variation during a day. The second experiment was performed to check the coherence time of data measured at specific locations. According to the results of these two experiments, it is shown that it is necessary to analyze the data after long time measurement, because of the large changes in value even in a short time. Thus, there would be difficulties in maintenance.

The third experiment was performed to check the change of patterns of the geomagnetism values while the user is moving. According to the results, it is shown that the measured data has a unique pattern while it moves a specific route and there is hardly changes in geomagnetism patterns as time is passed. Therefore, it is concluded that this feature can be utilized in indoor positioning scheme exploiting geomagnetic sensor, rather than the simple fingerprint scheme based on a static database of specific points.

Although it is confirmed that there is a unique pattern of the geomagnetism values, while it moves a specific route, the accuracy of proposed scheme is not evaluated enough in more realistic scenarios. As for future works, therefore, we will study about performance analysis of the proposed scheme with long intervals such as weeks, months, etc. In addition, the distortion of the measured geomagnetic values by different users will be evaluated for possible usage of the passive tracking scheme, which can track the user without any smartphones.

Acknowledgements This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the ministry of Education (NRF-2016R1D1A1B03932980).

References

1. Deak G, Curran K, Condell J (2012) A survey of active and passive indoor localization systems. Comput Commun 35(16):1939–1954
2. Shang J, Hu X, Gu F, Wang D, Yu S (2015) Improvement schemes for indoor mobile location estimation: a survey. mathematical problems in engineering
3. Liu Y, Yang Z, Wang X, Jian L (2010) Location, localization, and localizability. J Comput Sci Technol 25(2):274–297
4. Benedetto F, Giunta G, Guzzon E (2011) Enhanced TOA-based indoor-positioning algorithm for mobile LTE cellular systems. In: 8th IEEE workshop on positioning navigation and communication (WPNC), IEEE, Germany, pp 137–142
5. Kim B, Kwak M, Lee J, Kwon TT (2014) A multi-pronged approach for indoor positioning with WiFi, magnetic and cellular signals. In: International conference on indoor positioning and indoor navigation (IPIN), IEEE, Korea, pp 723–726
6. Terán M, Aranda J, Carrillo H, Mendez D, Parra C (20147) IoT-based system for indoor location using bluetooth low energy. In: IEEE colombian conference on communications and computing (COLCOM), IEEE, Colombia, pp 1–6
7. Cooper A, Hegde P (2016) An indoor positioning system facilitated by computer vision. In: IEEE MIT Undergraduate Research Technology Conference (URTC), IEEE, USA, pp 1–5
8. Liu H et al (2007) Survey of wireless indoor positioning techniques and systems. IEEE Trans Syst Man Cybern Part C (Appl Rev) 37(6):1067–1080
9. Farshad A et al (2013) A microscopic look at wifi fingerprinting for indoor mobile phone localization in diverse environments. In: International conference on indoor positioning and indoor navigation (IPIN), IEEE, France, pp 1–10

An IoT Based Smart Berthing (Parking) System for Vessels and Ports

Ahmadhon Kamolov and Su Hyun Park

Abstract Internet of Things, Big Data, Cloud Computing and such last model technologies are taking a pace towards every part of our life. It is worth to mention that novel technologies are being directed towards obtaining more improvements in the field of Marine. Applying the IoT in particular industries makes that industry's working process automation and creates various smart devices in the field. Also, in Marine, the process comes out identically to those industries. Namely, smart ports and smart ships substitute traditional ports and vessels. However, it should be considered that applying the technologies in the fields is not effortless. Perhaps, this process should be done pace by pace. So, building a smart ship and smart port which is expected to be the fundamental basement of shipping requires much time. To obtain it, traditional versions of the objects in the field should be alerted to smart substitutions and the processes also should be automated by dividing them into parts. In this paper, we presented a system which is dedicated to automating the mooring process of ships. The premium target of our IoT based system is finding vacant places automatically by using the sensors fixed at the port and transfer the data to the ships which are going to berth. In the experimental system, we used a smartphone instead of a ship for client-side of the system and made one Android Application that called "Smart Ship Berthing" instead of ship software. And for our proposed server-side system experiment, we utilized Raspberry Pi with Ultrasonic Sensor to detect a ship and modify empty place to berth.

Keywords The IoT · Smart ships · Smart ports

A. Kamolov · S. H. Park (✉)
Graduate School, Dongseo University, Busan, South Korea
e-mail: subak@dongseo.ac.kr

A. Kamolov
e-mail: ahmadxonk@gmail.com

A. Kamolov · S. H. Park
Division of Computer Engineering, Dongseo University, Busan, South Korea

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_13

1 Introduction

In the current time, novel sorts of technologies are covering the whole world. So, the technologies like Artificial Intelligence, Big Data, Cloud, and IoT cannot surprise us as they are living with us in every part of our life. It is undeniable that people are being surrounded by technics and technologies. That is the reason smart technologies are taking all the works which we do in our daily life. So, we are trying very hard to make every branch our world automation and we are nearly succeeding in this as the number of smart devices around us is pretty much and Smart Cars, Smart Watch, Smart Cards, Smart Cameras, Smart Fridge can be a classical example to this. Even our accommodations which we dwell in are turning into Smart houses. In a brief gist, all objects and things are being replaced by their smart substitutions. Some of such substitutions are ships and ports. Above mentioned technologies are causing several premium researches in Marine world too. It should be highlighted that developed countries and large companies making researches on it. Namely, large flourishing companies of Korea, China and Japan are proposing their IoT Solutions to the Smart ships [1–4] and developed countries of Europe such as the Netherlands and Germany are working on the Smart ports. The researches in this field are making on monitoring and controlling the data about the ship and through operating the data providing safety and doing all the work independently without utilizing humans' support and all the mentioned things are done via automating all ships. Modern ships, namely, smart ships focus on not only gathering and controlling the data independently but also they point out creating free communication with other ships and ports. In this pace, while doing such functions the role of diverse sort of sensors fixed at the ports and smart ships as well as hardware and software are really essential. Especially, in forthcoming years it is expected that the demand for software to smart devices will go up extensively. As each process of shipping requires particular automated system. Because in order to make the whole system of ships and ports automated and turn into them smart, their processes should be automated by dividing them into parts and after this the aimed target can be reached. From this point it is summed that in such huge objects firstly some processes are automated and then the whole system becomes smart or automatic monitoring. One of the shipping processes in which modern technologies such as IoT can be helpful and utilizing such technologies can be effective is the process of placing the ships at the port. In a nutshell, it is the berthing (mooring or we can say parking) process of ships. In this process, we may encounter some drawbacks and inconvenience in the system, which are being used now. So, the researches that we are doing are dedicated to the berthing process of ships at the port.

In this paper, we proposed IoT based System that can help ships for berthing. After getting the request of the ships arrived to the port, the system sends information about available places to berth automatically. In order to do this, the system uses the data in the sensors fixed at available berthing places. After going through the introduction of the system in Sect. 1, the rest of the paper includes following

sections: Sect. 2 illustrates Problem Analysis in Ship Berthing, Sect. 3 describes proposed system architecture and design, Sect. 4 is devoted to working scenario of the system, Sect. 5 highlights our experiment implementation and experiment results, and then conclusion for the paper.

2 Problem Analysis in Ship Berthing (Parking)

Utilizing IoT, Cloud, Big Data it is still considered novel process in the field of Marine and Shipping. Contemplating the shipping is one of the huge industries these researches and processes last in a discussive way. Because in the process of turning, ordinary ships and ports into smart versions the field covers a lot of tasks. Ships are usually reckoned as enormous means of transportation. In order to make them smart ships, there are required a big amount of hardware and software technologies. Surely, implementing those demands needs pretty much time as we cannot carry out it in a short period of time. As the implementation of this, we can create simple software to the small parts of a ship gradually.

By means of the research, we are going to share our small system which is really applicable, during the process of arriving and mooring at the port. We highly expect that our proposing system will help ships while reaching the port. Our system helps ships arrived to the port to find a place automatically. Before getting introduced with the system, it is crucial to comprehend the communication between ships and ports in order to get deep understanding about the amenities of our system. In the field of shipping, it is essential not only the communication between ships, but the communication between ships and ports is also vital. The main part of this communication occurs while ships arrive to the port. In these days, ships are wasting pretty much time during the process of berthing at the port. The period may last from two hours till several days. One of the reasons for the problem is the traffic of the ships arrived to the port and the other one is exchanging vast amount of data between ships and ports. The data which is needy for ports composes all information about the name of the ship, the owner, the crew and ship's overall dimensions, its condition, also departure and arrival time of the ship like in [5]. So, this process is requiring additional time, hardship and money as well. Taking into account of it, we highly believe that it is really time to apply modern Hi-Tech technologies to the berthing process of ships. In the following section, we will present our proposed system to the process that we mentioned above.

3 Architecture and Design of Proposed System

The premium target of the system is it gives an amenity to moor the vessels which has arrived at the port and transferred all the data about themselves. The system can carry out several works automatically such as it gets all data about available places

with the help of the sensors installed at the port and after getting the data it transfers it to the vessels looking for a place to moor. The system sorts out online data which are received from the sensors and sends an optimal place which is suitable in location, measure and other criteria to the ship. As most applications of IoT based systems are created using Three Layer architecture our proposed system is also based on this architecture (see Fig. 1).

Application Layer of the Smart Berthing (Parking) System for Ships. This layer consists of the software technology which is suitable for transferring the data about the ship to the port and sending a request for a place to moor.

Network Layer of the Smart Berthing (Parking) System for Ships. This layer provides to create a communication between Application and Sensor layer and its pattern technologies that can be utilized in our system are Wi-Fi, WiMAX, LTE, and LTE-A.

Sensor Layer of the Smart Berthing (Parking) System for Ships. The Sensor Layer of the system can be comprehended as simple sensors that are installed at the port. These sensors serve for clarifying a certain place if there is any ship or not at the port. Then, by means of Network layer it transmits the data to the Application layer namely it sends the data to the ships. Best sensor to identify the place status is Ultrasonic Sensor because it is very cheap and easy to use. That is why in the system we used UR Sensor.

And also we made high-level design for our proposed system (see Fig. 1). In the illustrated design you can see every place at the port that ships can berth shown with particular sensors and numbers. It means at the port every place has its own sensor

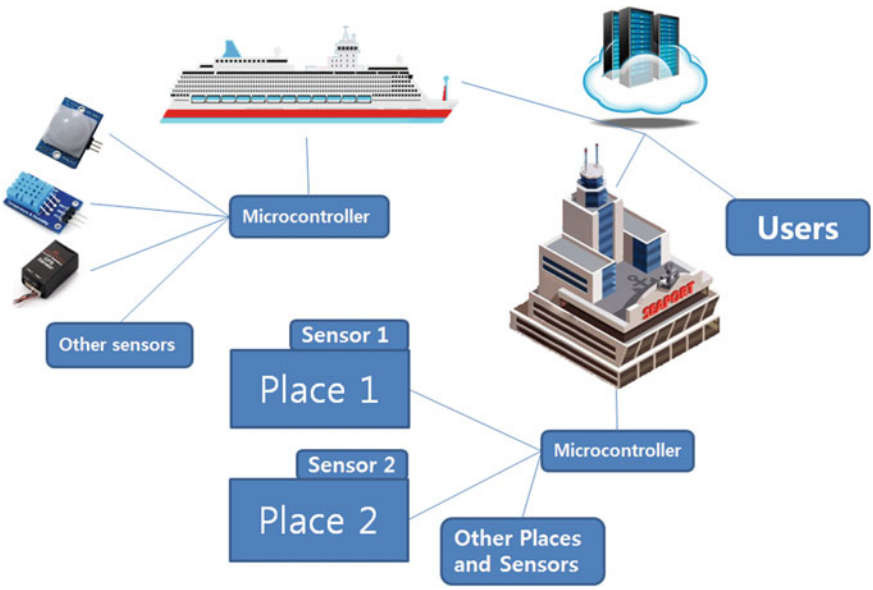


Fig. 1 High-level architecture of the smart berthing system for ships (Color figure online)

and number. In the picture sensor that marked with a red color means the space is busy and the sensor has "busy" information. And a sensor that marked with blue color means that space is empty and the sensor has "empty" information (Fig. 2).

4 System Working Scenario

So, how does the system work? Its working scenario is as following.

Step 1: The ship arrived to the port gets in touch with the system at the port via the network (WiFi, LTE, LTE-A, WiMAX, 5G).
Step 2: After getting a successful communication the ship sends the data about itself and asks a place to moor.
Step 3: The port receives all data transferred from the ships and checks it (see Fig. 3).

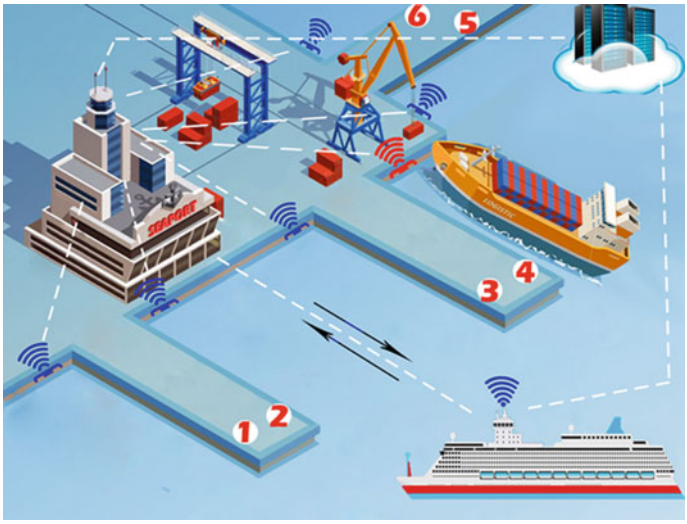


Fig. 2 High-level design of the smart berthing system for ships

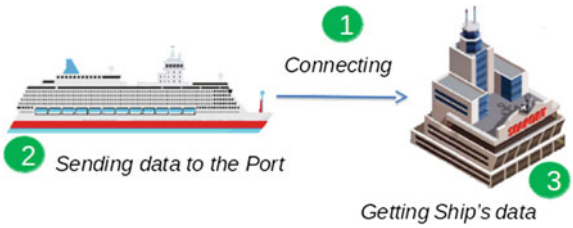


Fig. 3 System working scenario Steps 1–3

Step 4: After receiving the data about the ship, the system at the port gets information about available berthing places using the sensors in certain places. Port system sorts empty places.

Step 5: While getting the data about vacant places the system takes into concern the reserved places which are ordered by the ships arrived to the port earlier (see Fig. 4).

Step 6: All gathered data about vacant places are sorted and the optimal place that is totally suitable in all way is sent to the ship. After sending place information system waits for confirmation messages from the ship.

Step 7: The data about available places for berthing and the area of the place shown on the map of the port are sent to the ship. As soon as the ship reserves a place illustrated in the request message sent from the port, that place gets the same status as reserved places (see Fig. 5).

5 Implementation and Experiment Results

5.1 Experimental Implementation

Carrying out Experiment and implementation in the whole vessel and at the port is so complicated issue. However, the experiments of the system aimed for ships and ports can be carried out in small measure and in small systems. We did our research as illustrated below. The technique and technologies those are needy to carry out our experiment in Table 1.

So, in our experiment, a smartphone which works in android OC does the works of the ships as well as its software. Raspberry Pi and the sensors linked to it are responsible for giving information about the places at the port. So our proposed

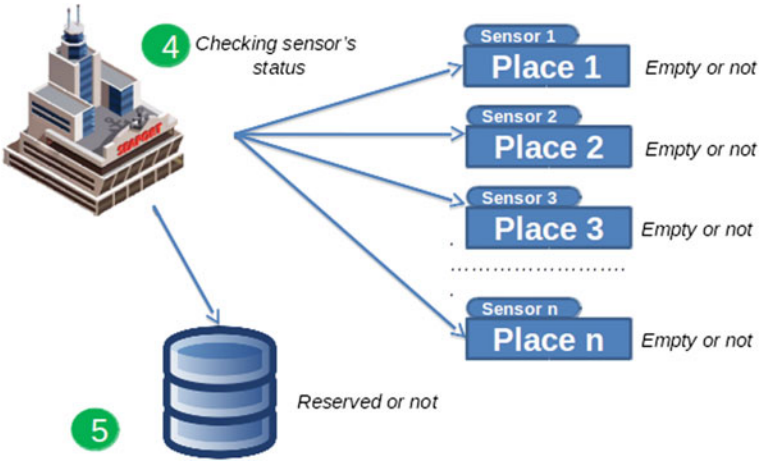


Fig. 4 System working scenario Steps 4–5

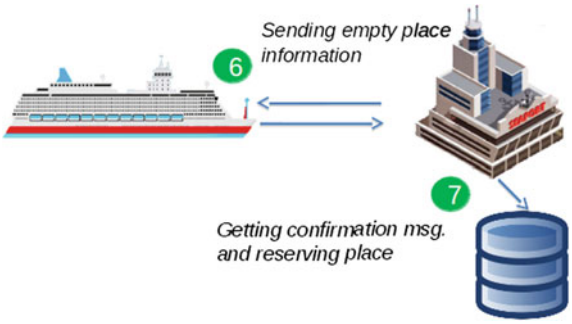


Fig. 5 System working scenario Steps 6–7

Table 1 Requirements of the experiment

Hardware Req.	Software Req.	Platform and Lang.	Others
Android Smartphone	Windows OS	Java & XML	Sensors
Raspberry Pi	Noobs Raspbian (Linux)	Python	Jumper wires
PC	Android OS & Android Studio	MySQL	Resistor
Wi-Fi router			LED
8 Gb CD-card			

system based on Three-layer IoT architecture. During the experiment on our system for the application layer of the system we did an application using Android Studio by means of a smartphone that works in android OS. The applied programming language is Java that is very popular language for developing smartphone Apps. In the second layer of the architecture, for our experiment, we chose Wi-Fi technology. In this point, Wi-Fi according to the conception of IoT architecture, serves as a mean of communication between the application layer (in the experiment it is android app) and sensor layer (raspberry pi and sensors). In the sensor layer, for our experiment all aimed works in the port system are done by Raspberry Pi and also the sensors attached to raspberry. Here, Noobs Raspbian (Linux OS) is mounted to Raspberry Pi. The system is built on the basis of a Python programming language for receiving information from the sensors. The server exchanges information and commands using the sensors connected to the raspberry. And, experimental hardware implementation for the proposed port system illustrated in Fig. 6.

5.2 Experimental Result

From here you can see our experimental smartphone application for the proposed ship system. In this part, we made a high designed application which is working on Android OC. Our application called “Ship Berthing”. Working scenario of the

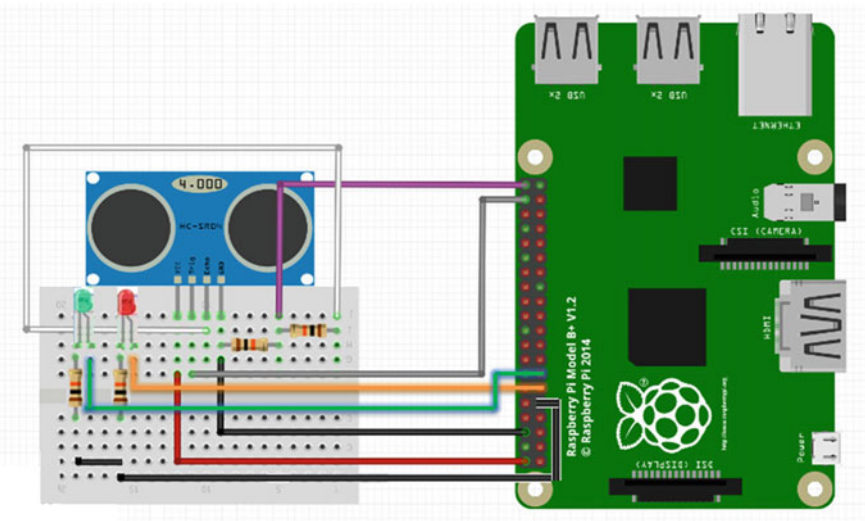


Fig. 6 Hardware implementation of the proposed port system

application is very simple. For ship berthing, users just need to connect to the port system and ask a place to ship berthing. To select the particular port, a user should know about the location of that port like, in which country, in which continent located that port that they are going to berth. The application activities and working scenario are as follows: After running App 1st activity page is "Selecting Continent" that user should select continent which their arriving country and port are located. You can see this activity in Fig. 7a. 2nd activity page is "Selecting Country" that user should select the country that their arriving port is located. You can see this activity in Fig. 7b. 3rd activity page is "Selecting Port" that user should select their arriving port. You can see this activity in Fig. 7c. 4th activity page is main activity page is "Port Activity" that user can send their information (in this case IMO ID Number of Vessel) and asks a place to moor by clicking "MOORING REQUEST" button. And also users can check status and condition of that port that is there place empty or busy by online Port Map. You can see this activity in Fig. 7d. In the online map, red places mean that places are busy and there are ships in the current time. Yellow places mean they are also busy but there are no ships in current time, those places are reserved by new arriving ships that are going to berth recently.

From here you can see our experiment for the proposed port system. In the experiment to check empty place at the port, we used HC-SR04 UR Sensor. This sensor can detect an object in the distance around 4 meters in front of it. By detection results system can modify, there is a ship or not. To reading sensor information we utilized Python Language on Raspberry Pi microcontroller which is working on Noobs OC. And, we used two, red and green LEDs to set sensor status "busy" or "empty". Red LED shows sensor is busy it means there is a ship at the

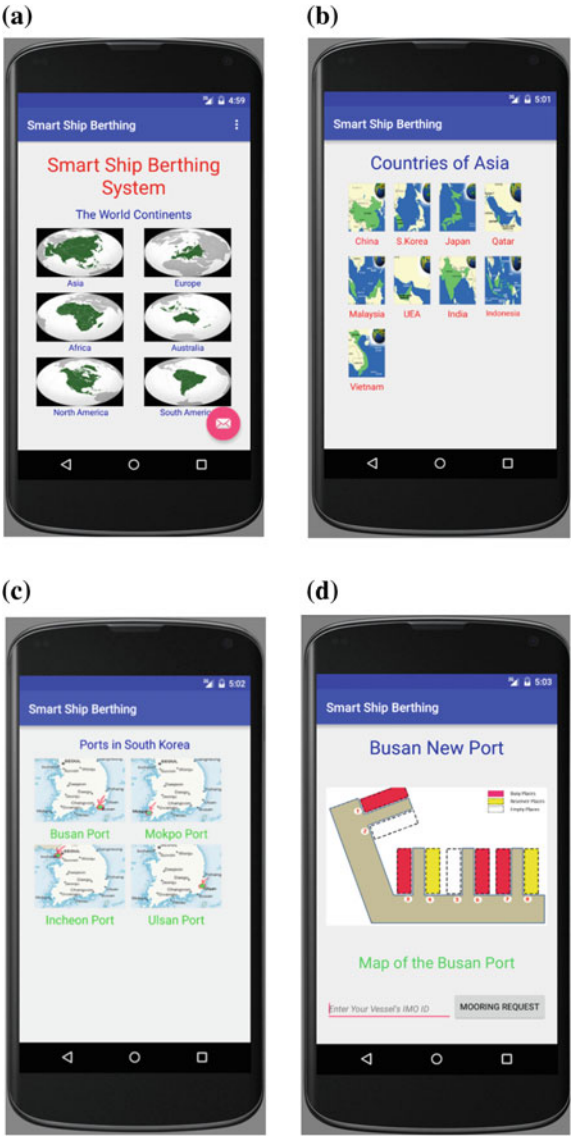


Fig. 7 **a** Selecting continent. **b** Selecting country. **c** Selecting port. **d** Port map (Color figure online)

port. Green LED shows sensor empty which means place also empty to moor. Figure 8a, b illustrated status of the sensor which is mooring place is empty. The green LED is on.

Figure 8c, d illustrated status of the sensor which is mooring place is busy. The green LED is on. Ships should wait to moor. According to the outcome of the

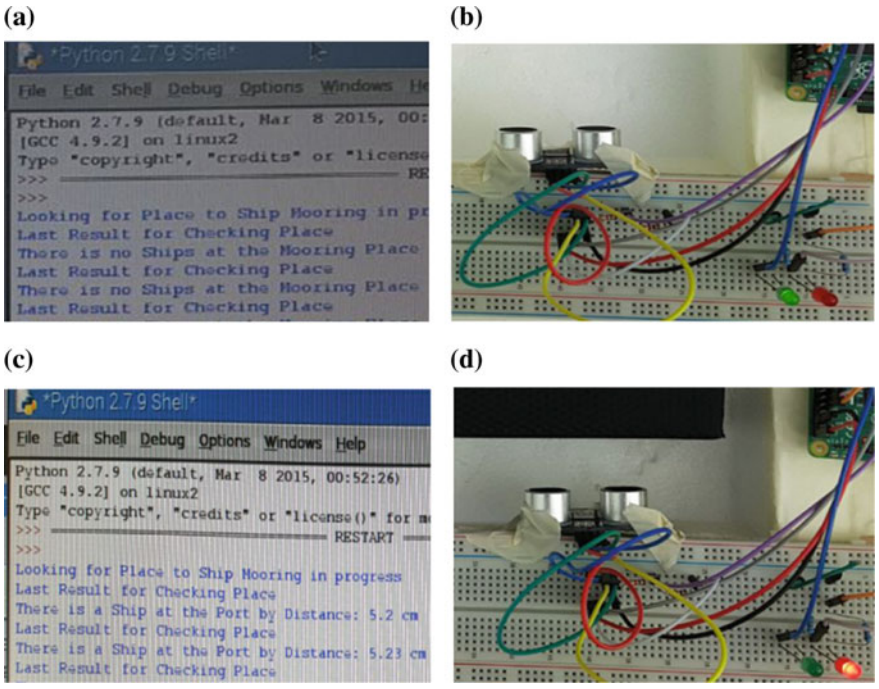


Fig. 8 **a** Sensor status is empty (Msg result). **b** Sensor status is empty (LED result). **c** Sensor status is busy (Msg result). **d** Sensor status is busy (LED result) (Color figure online)

experiment for our proposed system, Sensors and Smart Systems can be used in Marine industry. In the experimental system, we used a smartphone instead of a ship for client-side of the system and made one Android Application that called Smart Ship Berthing instead of ship software. And for our proposed server-side system experiment we utilized Ultrasonic Sensor to detect a ship and modify empty place to berth. The range of the Ultrasonic Sensor is around 4 m and its accuracy is very high that can detect an object in front of the sensor.

6 Conclusion

In this we present an IoT based smart mooring system for vessels and ports. The system is dedicated to automating the mooring process of the ships at the port. The premium target of our IoT based system is by means of the sensors fixed at the port to find vacant places automatically and transfer the data to the ships which are going to take a berth at the port. The intended system gives an amenity to minimize the time, effort and the cost while mooring the vessels.

References

1. Haun E, Hyundai heavy industries debuts ‘smart ship’ solution. In: Homepage, https://www.marinelink.com/news/industries-solution427522, last accessed 10 Oct 2017
2. Japan develops smart ship platform, Homepage, http://www.marinemec.com/news/view,japan-develops-smart-ship-platform_44850.htm, last accessed 15 Oct 2017
3. Hamburg Port Homepage, https://www.hafenhamburg.de/en, last accessed 10 Nov 2017
4. Amsterdam Port Homepage, https://www.portofamsterdam.com/en/shipping/arrivals-and-departures, last accessed 10 Nov 2017
5. Arriving at a port Homepage, http://www.pfri.uniri.hr/∼bopri/documents/21, last accessed 21 Nov 2016

Part II
Workshop on Technology of AI and IoT with Wireless and Advanced Networking

3D Haptic-Audio Enabled Online Shopping: Development and Challenges of a New Website for the Visually Impaired

Kian Meng Yap, Alyssa Yen-Lyn Ding, Hui Yin Yeoh, Mei Ling Soh, Min Wea Tee, Khailiang Ong, Wei Kang Kuan and Ahmad Ismat Abdul Rahim

Abstract The advancement curve of Information Communication Technology (ICT) is growing rapidly over the years; it is essential to look into the application of the current technology to augment its potential benefits in the real world. One such area is how the current technology can be better adapted to contribute to the less privileged such as the visually impaired (VI) community, which is the focus of this project. While people with fully functional abilities are able to adapt to the rapid change in technology, VI individuals often stumble in this area. Due to their restricted visual abilities and heavy reliance on tactile perception which is minimized with the introduction of touchscreens, VI individuals experience difficulties in

K. M. Yap (✉) · A. Y.-L. Ding · H. Y. Yeoh · M. L. Soh · M. W. Tee
K. Ong · W. K. Kuan
School of Science and Technology, Sunway University, Subang Jaya
Selangor, Malaysia
e-mail: kmyap@sunway.edu.my

A. Y.-L. Ding
e-mail: alyssad@sunway.edu.my

H. Y. Yeoh
e-mail: hui.y3@imail.sunway.edu.my

M. L. Soh
e-mail: meilings@sunway.edu.my

M. W. Tee
e-mail: min.t2@imail.sunway.edu.my

K. Ong
e-mail: khai.o@imail.sunway.edu.my

W. K. Kuan
e-mail: wei.k33@imail.sunway.edu.my

A. I. Abdul Rahim
Telekom Research & Development Sdn. Bhd. (TM R&D), TM Innovation Centre,
Lingkaran Teknokrat Timur, 63000 Cyberjaya, Selangor, Malaysia
e-mail: drismat@tmrnd.com.my

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_14

maximizing the benefits of technology. Not only that, it is also a challenge for VI individuals to navigate through websites, especially for the purpose of online shopping. This research looks at creating a new 3D haptic-audio virtual objects website to counter the internet browsing limitations of VI individuals, while enhancing their shopping experience. The challenges encountered in this process are further highlighted and discussed.

Keywords Haptic-Audio shopping website · Visually impaired
Virtual environment · 3D objects

1 Introduction

One of the ways that the human somatosensory system informs the brain about the presence of physical objects in one's surroundings is through the sense of touch (e.g., bumping into a table) [11]. This touch sensation can be further enhanced with the use of haptic technologies as it allows the human somatosensory system to feel objects by providing haptic sensations in both virtual and real environments. There are two primary types of haptic sensations: tactile and kinaesthetic. Tactile sensations are when the skin is in physical contact with an object whereas kinaesthetic sensation is perceiving information about the position and movement of the body part through stimulation of muscles and joints [1]. Theoretically, kinaesthetic information will be perceived and processed 24/7, because, to a certain degree, every movement (even the action of lying down) utilizes the compression of certain muscles. Tactile sensations, on the other hand, are perceived only when the skin is in direct contact with an object [2].

A haptic device is defined as a bidirectional-actuated human-machine interface in which humans interact with a device, and the interaction is displayed in a virtual environment [3]. Its usage and application spans across many areas including medicine, research, training, and education (e.g. [4, 5]). The device acts as a medium that provides tactile and/or force feedback when interaction between human and computer occurs. It allows individuals to touch, feel, and manipulate 3D objects in tele-operated systems as well as virtual environments. This device can tolerate a force up to 3.3 N, allowing a good illusion of tactile rendering.

At present, the applications of haptics have extended to the visually impaired population. According to Sreelakshmi and Subash [3], integrated touch screens have been implemented with haptic feedback in certain devices (e.g. Apple's iPhone 6s and above). Nonetheless, the potential of these devices cannot be fully maximized if a website is not created to support and incorporate haptic technology. As research on haptic-audio feedback interaction is scarce in Malaysia, the current research hopes to address this gap and looks into how haptic-audio technology can be implemented to improve the accessibility, usability and user experiences for the visually impaired (VI) community. The current research is aimed at creating a new

model and algorithm to investigate and evaluate the common preferences of the VI when browsing independently in a 3D virtual learning environment. For example, this new model can be used to develop a haptic-based map of a store's layout. In addition, the new methods employed to create 3D images of products (e.g. grocery) can allow the VI to have a new and enriched online shopping experience. As a result, VI individuals would be able to buy items online with minimal help from sighted others.

This paper aims to accomplish the following goals in order to achieve the proposed major objectives: (1) To investigate how visually impaired users access layout and product information for a website; and (2) To investigate the perception and usability experience of 3D virtual objects through haptic stimuli and compare 3D virtual objects through simulation of different haptic stimuli. This paper presents an initial study by outlining the development of a new haptic-audio website for VI to shop online with minimal guidance from sighted others. It also provides analysis of the possible psychological challenges faced by the VI community as a result from the feedback of the participants while using the haptic device with haptic-audio website.

2 Related Work

2.1 Haptic Assistive Tools for VI Individuals

Research has shown that VI individuals displays a sense of excitement when immersing in the virtual world of the Internet. The Internet provides them with the opportunity to participate and socially connect them with the outside world [6]. However, technology advancement has often proven to be an obstacle that is restricting the prevalent use of Internet among this population. This is particularly obvious when dealing with interaction between VI and machines. In recent years, researchers have been looking into ways to improve the VI's autonomy in navigating the Internet with minimal help. One of the ways is the integration of haptic and audio technology to allow VI to browse more autonomously in a virtual environment. Although haptic technology is slowly evolving, more research and development is still needed in combining visual, haptic, and auditory display to better enhance the VI's experience in web browsing.

Assistive technology has been proven to improve the skills of the visually impaired. Technologies such as VoiceOver, Duxbury Braille Translator, and text enlargers can help improve the quality of life of the visually impaired as it allows them to achieve literacy and also independence [7, 8]. For example, a screen reader is a software application with a speech synthesiser that allows a VI individual to comprehend the contents that is on the current screen. In most cases, the screen reader is able to reiterate the content on the page; but in some cases where there are graphics with no description, there is no current way a screen reader can process

these graphical information. With the rise of assistive tools involving haptic technology, it might be possible to process graphical information on the screen in the time to come.

Wong et al. [9] has previously developed a system that is similar to the current study. They utilized haptic technology to provide a realistic online shopping service for VI individuals. The system, HABOS, allows VI individuals to shop online independently. This system consists of a haptic-audio enabled browser that allows users to receive haptic feedback from the haptic device and obtains information verbally via the speech synthesiser. However, this system was not tested on the VI community, and therefore the inputs and feedbacks from VI individuals are not gathered. The current study goes one step beyond and addresses the limitation of their study so that the prototype invented will be relevant and beneficial to the targeted population.

2.2 Applications of Haptic Technologies

At present, there are various interesting possibilities of the application of haptic technologies that are still being explored. A few areas of examples of these have been highlighted by Boswell [10] and can be seen as follows:

> Entertainment: The possibilities are extremely exciting in the field of games and entertainment with haptic technology, bringing this technology down to a reasonable price point, along with an understandable, dynamic user interface will prove to be a challenge.
>
> Hardware: Perhaps, one day, haptic technology will be able to directly stimulate the central nervous system in order to replicate a touch experience. This could mean brain implants, nerve stimulation, and mechanical interfaces that work in tandem with our bodies.
>
> Wearable haptics: With Google Glass and Kinect tweaks, it looks like we are entering a whole era of wearable body sensors and computer networks. Devices like these can be tweaked to create a completely new paradigm of human and computer interaction, providing guidance for blind people, information and help for those with special needs, or even extra interaction while doing other tasks—cooking, driving, in a meeting, etc. (pa.12)

3 Research Methodology

The development of the haptic-audio technology and website was carried out in two few phases as detailed below: (Fig. 1).

3.1 Phase 1: Background Research

In the initial phase, background studies were carried out regarding the interaction between the visually impaired population and haptic technologies. Besides that, the layout and contents of the website were planned.

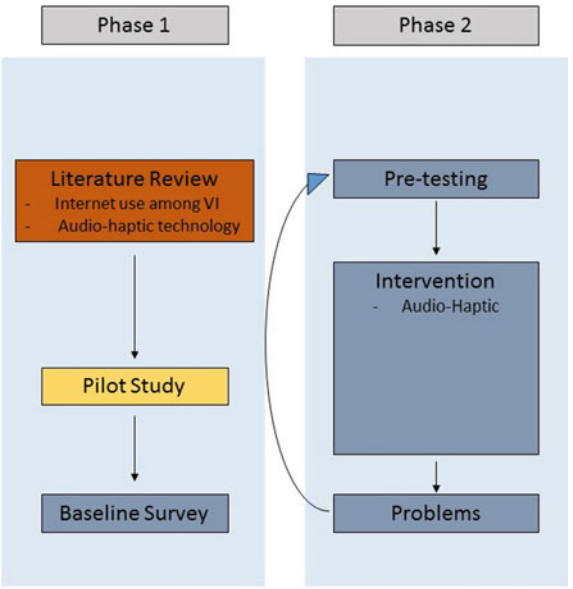


Fig. 1 Flow diagram of research methodology

3.2 Phase 2: Prototype Design and Construction

In this phase, a virtual environment was created so that the virtual object can be supported by the user interface. The development of prototype is executed after some of the background research has been carried out in phase 1 in order to learn about the method for constructing the foundation of the prototype. The system development is based on low-level Application Programming Interface (API) from Novint Falcon, high-level API from open sources such as OpenGL, scene graph API such as CHAI 3D and H3D, and programming language such as C++ and Python. In terms of hardware platform, there are a number of different force feedback devices available today. The haptic device used will be based on several different types of haptic devices such as Novint Falcon, Phantom Omni/Sylus or glove-type haptic device allowing holding and grapping of 3D virtual objects. The special feature allows users to manipulate 3D virtual objects in virtual environment.

3.3 Phase 3: Prototype Evaluation

The prototype is tested and evaluated to obtain feedback in terms of usability. This helps to detect potential user problems at an early stage before the development is completed. For the purpose of efficiency, a list of user problems is produced during the evaluation process. The evaluation process is carried out first by individuals who are blindfolded to simulate the perspective of a visually impaired person.

In this way, feedback from the user should be rather similar to that of a visually impaired individual as participants are blindfolded with the same intensity of darkness to a visually-impaired individual. The data collected will provide information about usability, functionality, efficiency and problems faced by participants during testing and how these problems can be overcome. Rating scale (on a Likert scale) are used to gather feedback to calculate average results for each task to rate their efficiency and usability.

3.4 Phase 4: Application and Feasibility Testing

After modifying the evaluated prototype, the prototype is tested on the visually impaired to gather their feedback of the website and haptic device.

4 Results and Discussion

4.1 Website Development

Jomje (http://jomje.homeislab.com/) is an e-commerce site created for this research project. Jomje is specially constructed for the ease of use of the VI population, whereby this website includes the support of additional features such as the option to increase font size, layouts and elements to ease the reading of the screen reader, and 3D object models that provide haptic feedback to the user with the aid of the haptic device which helps user in identifying the shape of the object that is selected (Fig. 2).

4.2 Modelling of 3D Objects

The 3D model used in this current phase is sample X3D models that are drawn for the purpose of this project. To enable haptic response from the haptic device, codes are modified on the X3D models. The currently added force and surface were <FrictionalSurface/>. This is still being experimented to determine which force works best to deliver the most realistic feeling on touching the virtual object. To simulate the touching of a 3D virtual model, the Touch Haptic 3D Stylus is configured to work with our environment, which will be subsequently replaced by the Force Feedback Haptic Exoskeleton that is being developed simultaneously. The application programming interface used to link 3D models and haptic feedback in our system is H3DAPI, where in H3DViewer, the model with depth information and haptic layer is rendered and presented to the users.

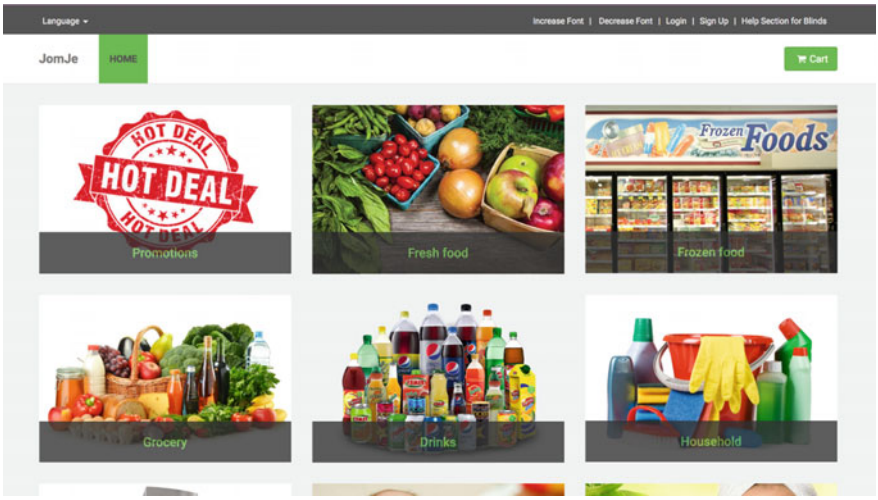


Fig. 2 Jomje homepage

A virtual environment has been created so that the 3D model object can be supported by the user interface. The 3D model object prototype development was created based on low-level Application Programming Interface (API) from Novint Falcon, high-level API from open sources such as OpenGL, scene graph API such as CHAI 3D and H3D, and programming language such as C++ and Python. H3DAPI is chosen because it accepts the X3D file format of 3D model object. Furthermore, X3D is a recognised element tag by World Wide Web Consortium and is included inside HTML5, which is the current version of HTML standard. H3DViewer will be used as a plugin for the website in the future, thus sticking to the standard way a 3D model is rendered in web browser. This will promise better application of our system to the world wide web.

Several changes on the 3D models have been made throughout as well. This includes the pointer which used to be allowed to hover everywhere: it is now only limited to hover on the objects. Apart from that, the starting view which previously begins from the top of the object has been changed to the front of the object. The system is also enabled with a notification system that guides the user when touching the 3D virtual object. In other words, the user will be alerted whenever the pointer is touching the surface of the 3D objects. Figures 3 and 4 show the 3D objects created to enable touch sensations and to provide 3D rendering experience to VI.

4.3 Challenges Identified Through Pre-testing of Website and Haptic Device

The first pre-testing session for the website using the haptic device was conducted among 29 participants from the National Council For The Blind, Malaysia

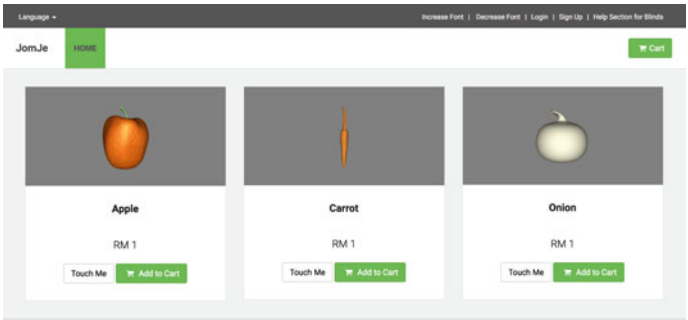


Fig. 3 Product page with 3D objects

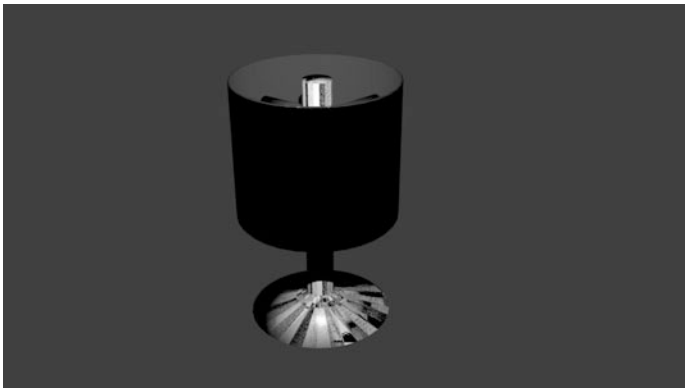


Fig. 4 New 3D objects

(NCBM), St. Nicholas Home, Penang, and Sunway University, Malaysia. The pre-testing includes technical testing of the haptic device and a survey of surfing experiences of the website with haptic technology. The challenges as reported by the VI participants are further highlighted in the sections below:

4.3.1 Language Barrier

The website was originally developed in English. However, during the pretesting session, it was found that there are quite a handful of participants that have very limited English comprehension, thereby restricting their access to the website as they were not able to comprehend the instructions accurately. This narrowed their use of the website with haptic technology, despite them having the interest and basic computer knowledge of the basic working mechanisms of computer navigation. Therefore, it is suggested that, in the future, the website should incorporate other languages such as Bahasa Malaysia which is the national language in Malaysia.

4.3.2 Utilization of Websites for Online Shopping

Based on the pre-testing results, it was found that the VI community only use the internet for online shopping occasionally. This is because, without the haptic device and website with haptic technology, concise descriptions (e.g. brand, size dimensions, weight, colour, quantity, texture, etc.) are required for them to imagine what the product portrayed on the website is actually like. This can be rather frustrating, especially when comprehensive information are not available or lacking. Hence, oftentimes, it is much more convenient to get a sighted individual to help them purchase what they need.

4.3.3 Efficiency of Haptic-Audio Enabled Website

It was also found that the website with haptic technology responded to inputs slower than the VI individuals would have expected it to. Specifically, there was a slight delay (about 1 s) between the website and input received from the haptic device. This could be due to the abundance of information that is available on the website as well as the limitation of the hardware used for testing, thereby causing a slight lag. Work is still needed to ensure that faster feedback can be provided during the interaction between the website and haptic device.

4.3.4 Applicability of Haptic Technology for VI Individuals

Most participants agreed that it was not difficult to learn how to use the website with haptic technology, so long as clear and concise prompts and instructions were available. These participants also agreed that they would recommend the website with haptic technology to their friends; however, this is not without reservations. Participants commented that it was difficult for them to know which part of the object they were touching (e.g. whether it was the front, back, or side of the object) when using the haptic device. In order to counter this limitation, a standardized layout and starting point (e.g. centre of object) of the haptic pen for the objects for all objects on the website needs to be implemented. In general, most of the participants enjoyed using the website with haptic device as it is able to incorporate the sense of touch, and is able to provide additional information such as shape of the object that otherwise would be unknown to them. It is important to note that this website with haptic technology was tested on VI participants that have decent computer literacy as well as some basic knowledge of simple objects. This means that a VI individual who has no prior knowledge in any of these areas may experience difficulties and frustration in navigating through the website and device. Hence, more work is needed to simplify the navigation of the website using haptic technology so that it can benefit the VI community in general as well.

5 Conclusion and Future Directions

In this paper, a new online website using haptic and audio technology is presented. This provides a new type of shopping experience for VI individuals. The work has also investigated the psychological challenges when surfing the website. The initial study has shown that the website is providing a new kind of shopping experience to the participants. In the future, 3D virtual objects that have different characteristics such as damping, stiffness, and frictions would be something useful to look at. This will allow VI individuals to have an enhanced and more accurate perception of objects that they are unable to physically see.

Acknowledgements This Research is fully funded by the Malaysian Communications and Multimedia Commission through the Networked Media Research Grant (Grant No: EXT-FST-CIS-MCMC-2016-01, MCMC(IRLC) 700-8/2/2/JLD.2(9)). St Nicholas Home Penang, National Council for The Blind Malaysia (NCBM), and Malaysian Association for The Blind (MAB) have also contributed invaluable advice, time, support, and expertise such that the objectives of this research can be achieved.

References

1. Demain S, Metcalf CD, Merrett GV, Zheng D, Cunningham S (2013) A narrative on haptic devices: relating the physiology and psychophysical properties of the hand to devices for rehabilitation in central nervous system disorders. Disabil Rehabil Assist Technol 8(3):181–189. https://doi.org/10.3109/17483107.2012.697532
2. Griffin AL (2001) Feeling it out: the use of haptic visualization for exploratory geographic analysis. Cartogr Perspect 39:12–29
3. Sreelakshmi M, Subash TD (2017) Haptic technology: a comprehensive review on its applications and future prospects. Mater Today: Proc 4:4182–4187
4. Liu LM, Li W, Dai JJ (2017) Haptic technology and its application in education and learning. In: 10th international conference on Ubi-media computing and workshops (Ubi-media), pp. 1–6. https://doi.org/10.1109/umedia.2017.8074138
5. Husman MAB, Maqbool HF, Awad MI, Abouhossein A, Dehghani-Sanij AA (2016) A wearable skin stretch haptic feedback device: Towards improving balance control in lower limb amputees. In: 38th annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp. 2120–2123. https://doi.org/10.1109/embc.2016.7591147
6. Williamson K, Wright S, Schauder D, Bow A (2001) The internet for the blind and visually imapired. J Comput-Mediated Commun 7(1). https://doi.org/10.1111/j.1083-6101.2001.tb00135.x
7. Celusnak BM (2016) Teching the iPhone with voiceover accessibility to people with visual impairments. J Vis Impairment Blindness 110(5):369–372
8. Orsini-Jones M (2009) Measures for inclusion: coping with the challenge of visual impairment and blindness in university undergraduate level language learning. Support Learn 24(1):27–34. https://doi.org/10.1111/j.1467-9604.2009.01394.x
9. Wong EJ, Yap KM, Alexander J, Karnik A (2015) HABOS: an exploratory study of haptic-audio based online shopping for the visually impaired. 1–6. https://doi.org/10.1109/have.2015.7359444

10. Boswel W (2013) Making touch more realistic: advances in haptic technology. Retrieved 14 March 2018, from https://software.intel.com/en-us/blogs/2013/05/08/making-touch-more-realistic-advances-in-haptic-technology
11. Tsuchitani C (n.d.) Somatosensory systems. In: Neuroscience Online. Retrieved from http://nba.uth.tmc.edu/neuroscience/m/s2/chapter02.html

An ECC Based Secure and Convenient Rural Medical Care System

Yong-Yuan Deng and Chin-Ling Chen

Abstract As hardware and software technologies have rapidly developed over recent years, many services and applications integrating these technologies into daily life have come to form the Internet of Things. At the same time, medical care requirements also raise rapidly. Governments formulate policies in order to respond the health-care requirements, one of these issues is rural medical care requirement. In recent years, the government has provided patrolling medical vehicle services to take care of residents in these remote areas. On the other hand, due to the rapid development of human physiological sensoring devices, residents with chronic diseases or those who need long-term monitoring of physiological conditions can wear these physiological sensoring devices, and these sensoring data can be provided to rural medical vehicles, so that doctors can diagnose symptoms. For residents with complicated or severe conditions, they can go to large hospitals in the urban areas for further diagnosis and treatment. However, a great deal of the information transmitted during these process are sensitive data, and must therefore be protected with security protocols. This paper proposes an ECC based secure and convenience rural medical care system. The proposed scheme provides mutual authentication, and guaranteeing data integrity, user untraceability and forward and backward secrecy, in addition to being resistant to replay attack.

Keywords Sensor network · Rural medical vehicle · Health-care
Mutual authentication · Privacy · Untraceability

Y.-Y. Deng (✉) · C.-L. Chen
Department of Computer Science and Information Engineering,
Chaoyang University of Technology, Taichung, Taiwan, R.O.C
e-mail: allen.nubi@gmail.com

C.-L. Chen (✉)
School of Information Engineering, Changchun Sci-Tech University,
130600 Changchun, China
e-mail: clc@mail.cyut.edu.tw

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes
in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_15

1 Introduction

Due to the rapid development of LTE, 3G, Wi-Fi, Bluetooth and ZigBee communication technologies, a variety of services and applications have become popular in daily life, one of them is medical care service. In order to respond the long-term health-care requirements of aging populations, governments formulate policies and build a comprehensive medical care system with diverse technologies [1, 2]. These technologies include sensor networks, cloud computation, GPS location, medical vehicle, etc.

Although the development of medical services is widespread today, some residents in the rural areas are still very inconvenient in seeking medical care. Residents of these areas that lacking medical resources need to travel long distances before they can approach hospitals for medical treatment. In recent years, the government has provided patrolling medical vehicle services to take care of residents in these remote areas [3, 4]. However, because the medical vehicle is an interim type of medical service, there is a lack of long-term physiological data for the diagnosis of some long-term chronic diseases. In addition, medical vehicles may lack some high-level medical equipment. For residents with complicated or severe conditions, it is necessary to go to large hospitals in the urban areas for further diagnosis and treatment.

On the other hand, due to the rapid development of human physiological sensoring devices, the size is getting smaller and smaller, and the power can support long-term operation. The physiological information measured by these physiological sensoring devices can be transmitted to the personal mobile device for recording. These data can be provided to rural medical vehicles or medical institutions, so that doctors can diagnose symptoms [5, 6]. However, these physiological sensoring data and physician diagnosis reports are personal privacy. These information must be protected to prevent be accessed illegally by malicious people.

Although some scholars have previously proposed a secure medical system for the Internet of Things, there is still lack of security mechanisms in rural medical care service [7, 8]. Therefore, we propose an ECC based secure and convenience rural medical care system. When the resident go to the rural medical vehicle to seek medical services, doctors will obtain the physiological sensoring data from the resident. After diagnosis, the doctor will inform the resident about the result. For further diagnosis or medical treatment, the resident can go to a large hospital to inform the doctors of the physiological sensoring data and the diagnostic report of the rural medical vehicle. This can achieve higher medical efficiency. The proposed rural medical care system achieves security, privacy and efficiency [9, 10].

The remainder of this paper is arranged as follows. Section 2 presents the proposed ECC based secure and convenience rural medical care system. Section 3 presents security analysis of the proposed scheme. Section 4 offers conclusions.

2 The Proposed Scheme

2.1 System Architecture

The rural medical care system framework of the scheme proposed in this study is shown in Fig. 1. There are four parties in the scheme:

Personal Mobile Device
A personal reading device carried by an individual. It can receive relevant data from a body sensor unit, and transmit the data to the rural medical vehicle or hospital medical device for analysis. It can also keep the inspection report from the rural medical vehicle, and transmit the record to the hospital medical device.

Rural Medical Vehicle
A medical vehicle to rural area for tour inspection. People go to rural medical vehicle, and transmit the data of body sensor unit from the personal mobile device to the rural medical vehicle. The rural medical vehicle makes a basic diagnosis of the resident.

Hospital Medical Device
A device carried by medical staff in a medical facility, or by caregivers in a care center. It can receive relevant data of a body sensor unit, or inspection report of the rural medical vehicle from a personal mobile device for diagnosis by a medical doctor.

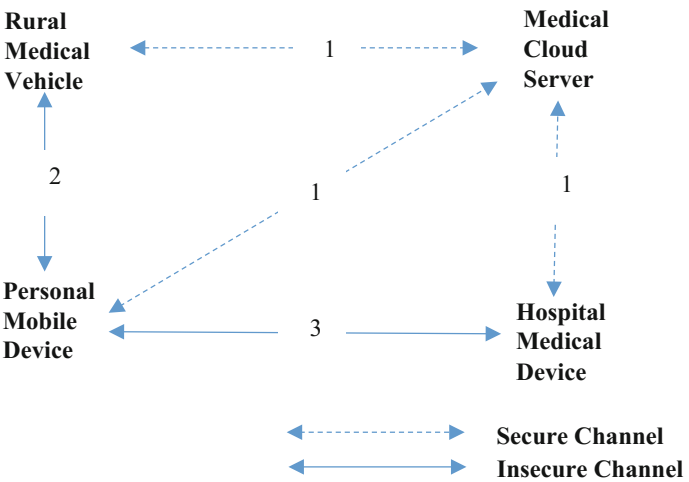


Fig. 1 Rural medical care system framework of the proposed scheme

Medical Cloud Server
A cloud server belonging to a national medical institution. It manages all personal mobile devices, rural medical vehicles, and hospital medical devices. All personal mobile devices, rural medical vehicles, and hospital medical devices must be registered on the medical cloud server.

1. All personal mobile devices, rural medical vehicles, and hospital medical devices must be registered with the medical cloud server through a secure channel. The personal mobile devices, rural medical vehicles, and hospital medical devices send their IDs (e.g. universally unique identifier, UUID) to the medical cloud server. The medical cloud server returns information that includes parameters calculated by Elliptic Curve Group technology.
2. When the resident takes his/her personal mobile device to the rural medical vehicle for health examination, they must authenticate the legality of each other. After mutual authentication between the personal mobile device and the rural medical vehicle, the personal mobile device transmits the data of body sensor unit to the rural medical vehicle. After the basic diagnosis of the resident, the rural medical vehicle transmits the inspection report to the personal mobile device.
3. The personal mobile device sends its ID and parameters calculated by Elliptic Curve Group technology to the hospital medical device for authentication. After mutual authentication between the personal mobile device and the hospital medical device, the personal mobile device sends the encrypted health sensoring data and inspection report to the hospital medical device. The medical doctor gets these data and report, and he/she makes an advanced diagnosis.

2.2 Notations

q	A k-bit prime
F_q	A prime finite field
E/F_q	An elliptic curve E over F_q
G	A cyclic additive group of composite order q
P	A generator for the group G
s	A secret key of the system
PK	A public key of the system, $PK = sP$
$H_i(\,)$	ith one-way hash function
ID_x	x's identity, like an UUID code
r_x, a, b, c, d	A random number of Elliptic Curve Group
S_x:	x's Elliptic Curve Group signature
SEK_{xy}	A session key established by x and y
$E_{\mathrm{x}}(m)$	Use a session key x to encrypt the message m
$D_{\mathrm{x}}(m)$	Use a session key x to decrypt the message m
CHK_x	x's verified message

$A \stackrel{?}{=} B$	Determines if A is equal to B
data	Body sensor's related sensoring information
record	The inspection report established by a doctor
c_i	The session key encrypted sensitive information

2.3 Initialization Phase

In the system initialization stage, the medical cloud server calculates some parameters, and publishes the public parameters for personal mobile devices, rural medical vehicles, and hospital medical devices.

Step 1 The medical cloud server chooses a k-bit prime p, and determines the tuple of Elliptic Curve Group $(F_p, E/F_p, G, P)$.

Step 2 The medical cloud server then chooses s as a secret key, and computes

$$PK = sP \tag{1}$$

as a public system key.

Step 3 Finally, the medical cloud server chooses hash function $(H_1(\cdot), H_2(\cdot), H_3(\cdot))$, and then publishes $(F_p, E/F_p, G, P, PK, H_1(\cdot), H_2(\cdot), H_3(\cdot))$ to all personal mobile devices, rural medical vehicles, and hospital medical devices.

2.4 Personal Mobile Device Registration Phase

The personal mobile device must register with the medical cloud server. The personal mobile device registration phase of the proposed scheme is as follows:

Step 1 The personal mobile device chooses an identity ID_{PMD} (e.g. UUID), and sends it to the medical cloud server.

Step 2 The medical cloud server chooses a random number r_{PMD}, calculates

$$R_{PMD} = r_{PMD}P, \tag{2}$$

$$h_{PMD} = H_1(ID_{PMD}, R_{PMD}), \tag{3}$$

$$S_{PMD} = r_{PMD} + h_{PMD}s, \tag{4}$$

and then sends (R_{PMD}, S_{PMD}) to the personal mobile device.

Step 3 The personal mobile device verifies

$$S_{PMD}P \stackrel{?}{=} R_{PMD} + H_1(ID_{PMD}, R_{PMD})PK. \tag{5}$$

If it passes the verification, the personal mobile device stores (R_{PMD}, S_{PMD}).

2.5 Rural Medical Vehicle Registration Phase

The rural medical vehicle must register with the medical cloud server. The rural medical vehicle registration phase of the proposed scheme is as follows:

Step 1 The rural medical vehicle chooses an identity ID_{RMV} (e.g. UUID), and sends it to the medical cloud server.

Step 2 The medical cloud server chooses a random number r_{RMV}, calculates

$$R_{RMV} = r_{RMV}P, \tag{6}$$

$$h_{RMV} = H_1(ID_{RMV}, R_{RMV}), \tag{7}$$

$$S_{RMV} = r_{RMV} + h_{RMV}s, \tag{8}$$

and then sends (R_{RMV}, S_{RMV}) to the rural medical vehicle.

Step 3 The rural medical vehicle verifies

$$S_{RMV}P \stackrel{?}{=} R_{RMV} + H_1(ID_{RMV}, R_{RMV})PK. \tag{9}$$

If it passes the verification, the rural medical vehicle stores (R_{RMV}, S_{RMV}).

2.6 Hospital Medical Device Registration Phase

The hospital medical device must register with the medical cloud server. The hospital medical device registration phase of the proposed scheme is as follows:

Step 1 The hospital medical device chooses an identity ID_{HMD} (e.g. UUID), and sends it to the medical cloud server.

Step 2 The medical cloud server chooses a random number r_{HMD}, calculates

$$R_{HMD} = r_{HMD}P, \tag{10}$$

$$h_{HMD} = H_1(ID_{HMD}, R_{HMD}), \tag{11}$$

$$S_{HMD} = r_{HMD} + h_{HMD}s, \tag{12}$$

and then sends (R_{HMD}, S_{HMD}) to the hospital medical device.

Step 3 The hospital medical device verifies

$$S_{HMD}P \stackrel{?}{=} R_{HMD} + H_1(ID_{HMD}, R_{HMD})PK. \tag{13}$$

If it passes the verification, the rural medical vehicle stores (R_{HMD}, S_{HMD}).

2.7 Rural Medical Vehicle Authentication and Communication Phase

When the resident takes his/her personal mobile device to the rural medical vehicle for some services, both parties must authenticate each other. After that, the personal mobile device transmits the encrypted health sensoring data to the rural medical vehicle, and the rural medical vehicle transmits the encrypted basic inspection report to the personal mobile device. The rural medical vehicle authentication and communication phase of the proposed scheme is as follows:

Step 1 The personal mobile device chooses a random number a, calculates

$$T_{PMD} = aP, \tag{14}$$

and then sends $(ID_{PMD}, R_{PMD}, T_{PMD})$ to the medical reader.

Step 2 The rural medical vehicle chooses a random number b, calculates

$$T_{RMV} = bP, \tag{15}$$

$$PK_{PMD} = R_{PMD} + H_1(ID_{PMD}, R_{PMD})PK, \tag{16}$$

$$K_{RP1} = S_{RMV}T_{PMD} + bPK_{PMD}, \tag{17}$$

$$K_{RP2} = bT_{PMD}, \tag{18}$$

and the session key is computed as following

$$SEK_{RP} = H_2(K_{RP1}, K_{RP2}). \tag{19}$$

The rural medical vehicle then calculates

$$CHK_{PR} = H_3(SEK_{RP}, T_{PMD}), \tag{20}$$

and sends $(ID_{RMV}, R_{RMV}, T_{RMV}, CHK_{PR})$ to the personal mobile device.

Step 3 The personal mobile device calculates

$$PK_{RMV} = R_{RMV} + H_1(ID_{RMV}, R_{RMV})PK, \tag{21}$$

$$K_{PR1} = S_{PMD}T_{RMV} + aPK_{RMV}, \tag{22}$$

$$K_{PR2} = aT_{RMV}, \tag{23}$$

and the session key

$$SEK_{RP} = H_2(K_{PR1}, K_{PR2}). \tag{24}$$

The personal mobile device verifies

$$CHK_{PR} \stackrel{?}{=} H_3(SEK_{RP}, T_{PMD}) \tag{25}$$

to check the legality of the rural medical vehicle.
If it passes the verification, the personal mobile device calculates

$$c_{PMD} = E_{SEK_{RP}}(data), \tag{26}$$

$$CHK_{RP} = H_3(SEK_{RP}, T_{RMV}), \tag{27}$$

and sends $(ID_{PMD}, c_{PMD}, CHK_{RP})$ to the rural medical vehicle.

Step 4 The rural medical vehicle verifies

$$CHK_{RP} \stackrel{?}{=} H_3(SEK_{RP}, T_{RMV}) \tag{28}$$

to check the legality of the personal mobile device. If it passes the verification, the session key SEK_{RP} between the personal mobile device and the rural medical vehicle is established successfully. The rural medical vehicle calculates

$$data = D_{SEK_{RP}}(c_{PMD}) \tag{29}$$

to get the health sensoring information of the resident. After diagnosis for the resident, the rural medical vehicle generates the encrypted basic inspection report

$$c_{RMV} = E_{SEK_{RP}}(record), \tag{30}$$

and sends (ID_{RMV}, c_{RMV}) to the personal mobile device.

Step 5 The personal mobile device decrypts the received message

$$record = D_{SEK_{RP}}(c_{RMV}), \tag{31}$$

and gets the encrypted basic inspection report from the rural medical vehicle.

2.8 Hospital Medical Device Authentication and Communication Phase

After the resident gets the basic inspection report from the rural medical vehicle, he/she may go to the hospital for advanced diagnosis. When the resident takes his/her personal mobile device to the hospital for some services, both parties must authenticate each other. After that, the personal mobile device transmits the encrypted health sensoring data and basic inspection report to the hospital medical device, and the medical doctor of the hospital makes an advanced diagnosis. The hospital medical device authentication and communication phase of the proposed scheme is as follows:

Step 1 The personal mobile device chooses a random number c, calculates

$$T_{PMD2} = cP, \tag{32}$$

and then sends $(ID_{PMD}, R_{PMD}, T_{PMD2})$ to the hospital medical device.

Step 2 The hospital medical device chooses a random number d, calculates

$$T_{HMD} = dP, \tag{33}$$

$$PK_{PMD} = R_{PMD} + H_1(ID_{PMD}, R_{PMD})PK, \tag{34}$$

$$K_{HP1} = S_{HMD}T_{PMD2} + dPK_{PMD}, \tag{35}$$

$$K_{HP2} = dT_{PMD2}, \tag{36}$$

and the session key is computed as following

$$SEK_{HP} = H_2(K_{HP1}, K_{HP2}). \tag{37}$$

The hospital medical device then calculates

$$CHK_{PH} = H_3(SEK_{HP}, T_{PMD2}), \tag{38}$$

and sends $(ID_{HMD}, R_{HMD}, T_{HMD}, CHK_{PH})$ to the personal mobile device.

Step 3 The personal mobile device calculates

$$PK_{HMD} = R_{HMD} + H_1(ID_{HMD}, R_{HMD})PK, \tag{39}$$

$$K_{PH1} = S_{PMD}T_{HMD} + cPK_{HMD}, \tag{40}$$

$$K_{PH2} = cT_{HMD}, \tag{41}$$

and the session key is

$$SEK_{HP} = H_2(K_{PH1}, K_{PH2}). \tag{42}$$

The personal mobile device verifies

$$CHK_{PH} \stackrel{?}{=} H_3(SEK_{HP}, T_{PMD2}) \tag{43}$$

to check the legality of the hospital medical device. If it passes the verification, the personal mobile device calculates

$$c_{PMD2} = E_{SEK_{HP}}(data, record), \tag{44}$$

$$CHK_{HP} = H_3(SEK_{HP}, T_{HMD}), \tag{45}$$

and sends $(ID_{PMD}, c_{PMD2}, CHK_{HP})$ to the hospital medical device.

Step 5 The hospital medical device verifies

$$CHK_{HP} \stackrel{?}{=} H_3(SEK_{HP}, T_{HMD}) \tag{46}$$

to check the legality of the personal mobile device. If it passes the verification, the session key SEK_{HP} between the personal mobile device and the hospital medical device is established successfully. The hospital medical device decrypts the received message

$$(data, record) = D_{SEK_{HP}}(c_{PMD2}), \tag{47}$$

and gets the encrypted health sensoring information and basic inspection report from the personal mobile device. The medical doctor of the hospital then makes an advanced diagnosis for the resident.

3 Security Analysis

3.1 Mutual Authentication

In the rural medical vehicle authentication and communication phase of the proposed scheme, when the personal mobile device wants to communicate with the rural medical vehicle, they must authenticate each other. The personal mobile device uses

$$CHK_{PR} \stackrel{?}{=} H_3(SEK_{RP}, T_{PMD}) \tag{25}$$

to verify the legality of the rural medical vehicle, and the rural medical vehicle uses

$$CHK_{RP} \stackrel{?}{=} H_3(SEK_{RP}, T_{RMV}) \tag{28}$$

to verify the legality of the personal mobile device.

Thus, the personal mobile device can verify the legality of the rural medical vehicle, and the rural medical vehicle also can verify the legality of the personal mobile device. The rural medical vehicle authentication and communication phase of the proposed scheme thus guarantees mutual authentication.

In the hospital medical device authentication and communication phase of the proposed scheme, when the personal mobile device wants to communicate with the hospital medical device, they must authenticate each other. The personal mobile device uses

$$CHK_{PH} \stackrel{?}{=} H_3(SEK_{HP}, T_{PMD2}) \tag{43}$$

to verify the legality of the hospital medical device, and the hospital medical device uses

$$CHK_{HP} \stackrel{?}{=} H_3(SEK_{HP}, T_{HMD}) \tag{46}$$

to verify the legality of the personal mobile device.

Thus, the personal mobile device can verify the legality of the hospital medical device, and the hospital medical device also can verify the legality of the personal mobile device. The hospital medical device authentication and communication phase of the proposed scheme thus guarantees mutual authentication.

3.2 Data Integrity

To ensure the integrity of the transaction data, this study uses Elliptic Curve Cryptography to calculate the session key SEK_{RP} and SEK_{HP}, and also to protect the

data integrity. The malicious attacker cannot use the signatures (K_{RP1}, K_{RP2}), (K_{PR1}, K_{PR2}), (K_{HP1}, K_{HP2}), and (K_{PH1}, K_{PH2}) to calculate the correct session key SEK_{RP} and SEK_{HP}.

Only a legal personal mobile device or rural medical vehicle can calculate the correct session key SEK_{RP}, and only a legal personal mobile device or hospital medical device can calculate the correct session key SEK_{HP}. Only the correct session key will allow successful communication. Thus attackers cannot modify the transmitted message. Therefore, the proposed scheme achieves data integrity.

3.3 User Untraceability

Another form of privacy attack involves attempting to obtain a person's physical location by tracing any personal device (in this case, the personal mobile device). If the personal mobile device sends the same message continuously, an attacker can trace its location. In the proposed architecture, the session key SEK_{RP} and SEK_{HP} is changed for every communication round in order to avoid location tracing. Thus, location privacy is protected, and user untraceability is achieved.

3.4 Resisting Replay Attack

An attacker may also intercept the message transmitted between the sender and the receiver. They impersonated a legal sender, and then send the same message again to the intended receiver. However, this attack will fail in the proposed scheme, as all messages between the sender and the receiver are protected with the session key SEK_{RP} and SEK_{HP}, and the attacker cannot calculate the correct session key. Because the transmitted messages are changed every round, the same message cannot be sent twice. Thus the replay attack cannot succeed.

3.5 Forward and Backward Secrecy

Even if the session key SEK_{RP} and SEK_{HP} between the sender and the receiver is compromised at any point by an attacker, the system still satisfies forward and backward secrecy. The attacker may use the session key SEK_{RP} and SEK_{HP} for future communication, or use it to obtain previous messages. However, in the proposed scheme, the session key SEK_{RP} and SEK_{HP} is randomly chosen by the sender and the receiver, and may only be used in the current round. The attacker cannot use the same session key SEK_{RP} and SEK_{HP} for future communication, or to obtain previous messages. Thus, a secure rural medical care system achieves forward and backward secrecy.

4 Conclusions

Recent developments in hardware and software communication technologies have given rise to what is known as the Internet of Things. Many services and applications can be provided through these technologies, including rural medical care services. The government arranges rural medical vehicles to rural areas for medical services. Residents with chronic diseases or those who need long-term monitoring of physiological conditions can wear physiological sensoring devices, and these sensoring data can be provided to rural medical vehicles, so that doctors can diagnose symptoms. For residents with complicated or severe conditions, it is necessary to go to large hospitals in the urban areas for further diagnosis and treatment. In order to improve the medical efficiency, the resident can inform the doctors of the physiological sensoring data and the diagnostic report of the rural medical vehicle. However, malicious people may seek to obtain sensitive personal data for various reasons, thus a robust authentication mechanism for rural medical care system is necessary. This study proposes an ECC based secure and convenience rural medical care system. The proposed authentication mechanism ensures mutual authentication, data integrity, user untraceability, forward and backward secrecy, and is secure against replay attack.

References

1. Moosavi SR, Gia TN, Nigussie E, Rahmani AM, Virtanen S, Tenhunen H, Isoaho J (2016) End-to-end security scheme for mobility enabled healthcare internet of things. Future Gener Comput Syst 64:108–124
2. Chiuchisan I, Dimian M (2015) Internet of things for e-health: an approach to medical applications. In: IEEE international workshop on computational intelligence for multimedia understanding (IWCIM), pp 1–5
3. Khemissa H, Tandjaoui D (2015) A lightweight authentication scheme for e-health applications in the context of internet of things. In: International conference on next generation mobile applications, services and technologies, pp 90–95
4. Yang Y, Ma M (2016) Conjunctive keyword search with designated tester and timing enabled proxy re-encryption function for e-health clouds. IEEE Trans Inf Forensics Secur 11:746–759
5. Abbas A, Khan S (2014) A review on the state-of-the-art privacy preserving approaches in e-health clouds. IEEE J Biomed Health Informatics 18(4):1431–1441
6. Yang J, Li J, Niu Y (2015) A hybrid solution for privacy preserving medical data sharing in the cloud environment. Future Gener Comput Syst 43–44:74–86
7. Fortino G, Galzarano S, Gravina R, Li W (2015) A framework for collaborative computing and multi-sensor data fusion in body sensor networks. Info Fusion 22:50–70
8. Gravina R, Alinia P, Ghasemzadeh H, Fortino G (2017) Multi-sensor fusion in body sensor networks: state-of-the-art and research challenges. Info Fusion 35:68–80
9. Yao X, Chen Z, Tian Y (2015) A lightweight attribute-based encryption scheme for the internet of things. Future Gener Comput Syst 49:104–112
10. Yang Y, Zheng X, Tang C (2017) Lightweight distributed secure data management system for health internet of things. J Netw Comput Appl 89:26–37

Water Resource Center Digital Tour and Pre-test and Post-test Systems

Tzu-Chuen Lu, Ting-Chi Chang and Jau-Ji Shen

Abstract This study develops a digital navigation system for the Water Resources Center to introduce the process and knowledge of sewage treatment to the general public. The system can be divided into three main parts, interactive multimedia games, pre and post online tests and 3D guided video. The system also combined with the official LINE account that allows visitors to receive timely results from the pre-test and post-test systems to learn about the sewage treatment-related knowledge. The score competition makes the game funnier that can enhance the entertaining results. The system also provides a completed management interface that assists the factory managers to maintain the entire system, including test questions pre-test and post-test, answer analysis, questionnaire feedback satisfaction analysis. According to the results of the rewinding analysis, the system indeed can promote the sewage treatment knowledge to the visitors. Most of the visitors feel the system is interesting and useful.

Keywords Digital navigation system · Pre-test · Post-test · Water resource center

1 Introduction

In the recent years, with the improvement of economic development and quality of life, water consumption has been increasing year by year. If ignored the problem, the sewage which containing harmful substances is directly or indirectly flow into the ocean that may cause water pollution and environmental degradation.

T.-C. Lu (✉)
Department of Information Management, Chaoyang University of Technology, Taichung, Taiwan
e-mail: tclu@cyut.edu.tw

T.-C. Chang · J.-J. Shen
Department of Information Management, National Chung Hsing University, Taichung, Taiwan

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_16

According to the survey, there are about 200–250 L of sewage by households generated by each person per day.

Most of the domestic sewage that lacks sewage sewers and treatment measures were directly discharged into rivers and severely polluted the natural environment. In the recent years, environmental issues have become more and more prominent. In conjunction with the water resources education and promotion program of Water Resources Department, this study aims to create a set of interactive navigation and learning platform for Water Recycling Centers that work with sewage treatment. Through lively element design, exquisite 3D multimedia video, the combination of process theory of interactive games, and pre-test and post-test system, the visitors can effectively learn the relevant knowledge and understand the importance of water resources.

2 Related Works

There are six major processing procedures to clear the sewage when it flows into the water recycling center [1–3]. The main procedures are described as following.

Pre-treatment equipment. Pre-treatment equipment is divided into two procedures, trash rack and grit chamber.

Trash rack. "Trash rack" is the first procedure when the sewage flows into the recycling center. The main function of the trash rack is to preliminarily screen the plastic, hair, paper, cotton cloth and excrement in the sewage by picking up the rubbish with the crawler rotating the rake and pouring the rubbish into the rear sieve. The sieve filters large garbage out, for example, cartons, and throw it in the junk car to prevent it falling into the next procedure to damage the equipment.

Grit chamber. The "grit chamber" makes the wastewater dropping by using the concept of the free sedimentation theory. At this time, the concentration of the suspended particles in the sedimentation process is not high and there is no interaction between the suspended solids. The particles moved down at the same speed. The sludge and sewage are separated in this procedure.

Primary sedimentation tank. The "primary sedimentation tank" is the second treatment process and is located in the underground sump. Most of the sewage entering the unit is still solid impurities. In this procedure, they use the principle of settling flocculation sedimentation to suspend particles flocculation with each other. Using the concept of mutual aggregation of particles to speed up the settlement process. Then, through a long time to let the sewage sludge with large density, sand and suspended solids sediment at the bottom. On the contrary, the water with lower density and oil will float in the above.

In order to reduce the load of follow-up treatment, the scraping mud machine in the pool scrapes the oil above the scum tank and scraps the sludge to the sludge pipeline, and the separated water flows out through the overflow weir.

Aeration tank. Aeration tank, also known as "bio tank", is the third treatment process. It mainly uses microorganisms present in the water to ingest organic

substances in the water as nutrient sources to maintain the energy of the body and synthesize new cells. After the conversion, most substances become sludge discharge. There are three common biological treatment methods, namely "activated sludge method", "biological rotary disc method" and "anaerobic treatment method."

Second settling tank. The "secondary settling pond", also called the "final sedimentation tank", is the fourth treatment procedure. Here, they use layered precipitation to settle the pellets. The sedimentation is affected by other particles in the surrounding. The relative positions of the pellets remain unchanged. Mainly the aeration tank generated by the colloidal sludge (microbial glue feather), in this precipitation separation, the sludge, oil water and aeration tank microbial sludge generated by the excretion of malodor off anorexia. Finally, the more clarified water will be sent to the chlorinated disinfection tank disinfection, most of the lower sludge sent back to the aeration tank for reusing again.

Sludge dewatering machine. Sludge generated in the sewage treatment process will enter the sludge digestion tank first, and the digested sludge after digestion tank is transported through the filter cloth and squeezed by the roller equipment to squeeze out the water in the sludge to produce sludge cake to reduce the volume of sludge. Sludge cake can become a greening zone composting and brick making recycling.

In order to allow visitors to quickly understand these complex processes, this study developed an interactive navigation system that allows people to understand the principles of wastewater treatment through the game. Furthermore, the system develops a pre-test system to test the prior knowledge of the visitors, and a post-test system to verify that the game indeed enhances knowledge of sewage treatment of visitors. In addition, this study uses feedback questionnaire to understand people's acceptance of the game.

3 Proposed Scheme

In this study, the interactive guided learning game platform is divided into two parts: the viewer side and the manager side.

3.1 The Viewer Side and Manager Side

The viewer side with lively colors and elements designed to attract attention and interest, with six major functions, so that visitors can effectively through this digital navigation platform to understand the water resources and learning to sewage processing related knowledge.

Manager site is divided into 3 different roles, the highest administrator, director, and employee, according to each permission to provide the corresponding function. The highest administrator allows changing the permission function for each role.

This research has designed six functions for the three aspects of "manager," "management topics," and "analysis," so that the factory can manage the entire platform easily, quickly and effectively.

3.2 Research Assumptions

In this study, a questionnaire survey was conducted to investigate whether the system affected the usefulness of the above two factors by incorporating the two factors of perceptual interest and perceived ease of use into the Technology Acceptance Model (TAM) [4, 5]. Thereby affecting people's attitudes toward the use of this system and exploring whether the public can improve their learning outcomes after their use. The research framework is shown in Fig. 1.

According to the science and technology acceptance model proposed by Moon and Kim, this study makes the following five hypotheses:

Assuming H1 perception has a positive effect on perceived usefulness. When the player thinks that the interaction between himself and the game is fun and happy, it will make the player feel that the game can effectively help improve the learning effect and hold the positive attitude. Therefore, this study argues that perceptual interest has a positive effect on perceptual usefulness.

Assuming H2 perceived ease of use has a positive effect on perceived usefulness. When the player feels that the game does not have to spend too much time and energy on how to operate, it will have a positive attitude toward the ease of use of the game and will also have a positive effect on the usefulness of the game. Therefore, this study argues that when the player feels that the game is easy to manipulate, it feels that the game is useful to the player itself.

Assuming the use of H3 perception has a positive effect on the attitude. When the player feels the game can enhance the learning effectiveness, it will have a positive feeling of the game. Therefore, the system considers the game player's perceived usefulness and the use of their own attitude has a positive effect.

Assuming H4 perceived enjoyment has a positive effect on the attitude. Moon and Kim pointed out that the sense of fun is intrinsic motivation, and when the interaction between the player and the game feels interesting, it will enhance

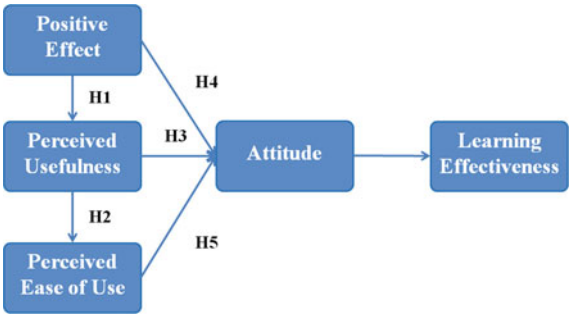


Fig. 1 The diagram of this study

their willingness to use. Therefore, this study believes that to enhance the attitude of using games will have a positive impact.

Assuming H5 perceived ease of use has a positive effect on the attitude. Because the player can spend less time and energy to learn how to operate the game, the game has a more positive feeling. This study argues that when the player has a positive influence on the ease of use and the attitude of using the game, the positive attitude of using. Their intention to use the game will also have a positive impact on their learning outcomes.

3.3 Questionnaire Design

The questionnaire was designed using the Likert 5-point Scale. When the respondents answered the questions on the questionnaire, they specifically rated their approval of the statement. The scores ranged from 5 to 1 with "strongly agree", "agree", "normal", "disagree" and "very disagree" (Table 1).

3.4 Data Analysis

Reliability analysis. This study uses the Cronbach's α coefficient to measure the reliability of this questionnaire. The reliability value of the questionnaire is measured by the measurement system between 0 and 1 to determine whether the questionnaire is consistent or stable so that the questionnaire can be consistent under different conditions or stable result. The formula is as shown in Eq. 1.

$$\alpha = \left(\frac{n}{n-1}\right)\left(1 - \frac{\sum \alpha_{xi}^2}{\sigma_x^2}\right) \tag{1}$$

α is the estimated reliability, n is the total number of tests, α_{xi}^2 is the number of variations in the fraction of i, σ_x^2 is the variation for the total score. The larger Cronbach's α value means the higher the reliability of the questionnaire, at least greater than 0.5, and the best taken in practice is greater than 0.7.

Correlation analysis. Correlation analysis was conducted to understand the correlation between study variables. Pearson correlation was used to measure the correlation (linear dependence) between two variables X and Y, with values between -1 and 1 is calculated as shown in Eq. 2.

Table 1 Likert scale

Strongly agree	Agree	Normal	Disagree	Very disagree
5	4	3	2	1

$$r = \frac{Cov_{XY}}{S_X S_Y} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma_X}\right)\sum_{i=1}^{n}\left(\frac{Y_i - \bar{Y}}{\sigma_Y}\right) \quad (2)$$

Cov_{XY} is the covariance number of X and Y, which is obtained by subtracting the average of X_i and Y_i from each group and multiplying them respectively. The calculated value is located between infinite positive and negative infinity. Divided it by the standard deviation of X and Y, find the correlation coefficient r. The correlation coefficient is between −1 and +1. When $r > 0$, it indicates the positive correlation between the two variables. On the contrary, when $r < 0$, it means the two variables are negatively correlated. When the coefficient is closer to −1 or +1, the higher the correlation of research variables, the closer to 0 means no correlation.

Regression analysis. Regression analysis refers to the correlation between dependent variables and dependent variables, the purpose is to understand whether two or more variables are related, related to the direction and intensity.

4 Experimental Results

This study performs the system to several senior and middle schools in Taichung, let students of different ages to actually use the games and platforms, and applying questionnaires to explore whether the system affects the students' perception of interest, perceived ease of use, and thus the perceived usefulness. The total number of participated students is 98. The basic information is shown in Tables 2, 3, 4. There are 85 questionnaires were returned. After checking whether there was an

Table 2 Basic information of students

	Category	Number
Sex	Male	43
	Female	42
Age	Under 15	18
	16-19	67

Table 3 Questionnaire results

Item	Number	Percent (%)
Valid questionnaire	85	86.73
Invalid questionnaire	13	13.27
The total number of questionnaires	98	100

Table 4 Reliability analysis of this study

Item	Cronbach's α value
Perceived enjoyment	0.881
Perceived usefulness	0.908
Perceived ease of use	0.889
Attitude	0.864

incomplete answer or no change in the questionnaire, the result was 13 invalid questionnaires. The rate of 86.73%, questionnaire results are shown in Table 3.

4.1 The Results of the Reliability Analysis

In this study, Cronbach's α coefficient was used to measure the stability of the questionnaire. The results are shown in Table 4, and the α values of all the facets were all greater than 0.8 after reliability analysis. It shows that the questionnaire in this study is very credible.

4.2 Correlation Analysis of This Study

The average number of research facets of each subject was analyzed by Pearson Correlation to explore the interaction between the faces to confirm the hypothesis of this study. The results are shown in Table 5. Significant values of all facets are less than 0.01, indicating that each structure is "medium and highly positive correlation".

4.3 Regression Analysis Results of This Study

Due to the small number of samples in this study, we conducted further regression analysis for each hypothesis to further verify.

Table 5 Relevant analysis of each research plane

	Perceived usefulness	Perceived ease of use	Perceived enjoyment	Attitude
Perceived usefulness	1			
Perceived ease of use	0.591**	1		
Perceived enjoyment	0.488**	0.636**	1	
Attitude	0.441**	0.709**	0.648**	1

ps:*means p<0.05** means p<0.01***means p<0.001

The impact of "Perceived Enjoyment" on "Perceived Usefulness". In this study, perceived enjoyment as an independent variable, perceived usefulness as a variable, the results are shown in Table 6.

As shown in Table 6, Beta is positive that indicating perceived enjoyment has a positive effect on perceived usefulness, R squared explaining 23.8% of the variance, with $P < 0.001$ being significant, and therefore H1 hypothesis is valid.

The impact of "Perceived Ease of Use" on "Perceived Usefulness". In this study, perceived ease of use as an independent variable, perceived usefulness as a variable, the results are shown in Table 7.

As shown in Table 7, Beta is positive that indicating perceived ease of use has a positive effect on perceived usefulness, R squared explaining 34.9% of the variance, with $P < 0.001$ being significant, and therefore H2 hypothesis is valid.

The impact of "Perceived Usefulness" on "Attitude". In this study, perceived usefulness as an independent variable, attitude is the dependent variables, the results shown in Table 9.

As shown in Table 8, Beta is positive that indicating perceived usefulness has a positive effect on the attitude of use. R squared explained 19.4% of the variance, with $P < 0.001$ being very significant. Thus, the H3 study Assumptions are valid.

The impact of "Perceived Enjoyment" on "Attitude". In this study, perceived enjoyment as an independent variable, attitude is the dependent variables, the results shown in Table 9.

Table 6 Perceived enjoyment and perceived usefulness of regression analysis

Variables	Beta
Perceived enjoyment	0.488***
R square	0.238

Dependent variable Perceived usefulness

Table 7 Perceived ease of use and perceived usefulness of regression analysis

Variables	Beta
Perceived ease of use	0.591***
R square	0.349

Dependent variable Perceived usefulness

Table 8 Perceived usefulness and attitude of regression analysis

Variables	Beta
Perceived ease of use	0.441***
R square	0.194

Dependent variable Attitude

Table 9 Perceived enjoyment and attitude of regression analysis

Variables	Beta
Perceived enjoyment	0.648***
R square	0.419

Dependent variable Attitude

Table 10 Perceived ease of use and attitude of regression analysis

Variables	Beta
Perceived ease of use	0.709***
R square	0.503

Dependent variable Attitude

As shown in Table 9, Beta is a positive number, indicating that the perceptual interest has a positive effect on the attitude of use. R-square accounts for 41.9% of the variation which $P < 0.001$ is very significant. Therefore, H4 study Assumptions are valid.

The impact of "Perceived Ease of Use" on "Attitude". In this study, perceived ease of use is an independent variable, attitude is the dependent variable, the results are shown in Table 10.

As shown in Table 10, Beta is positive that indicate that perceived ease of use has a positive effect on the attitude of use, R square explains 50.3% of the variance, with $P < 0.001$ being very significant and therefore H5 The research hypothesis is valid.

4.4 Verify the Research Hypothesis

Summarized the results of sub-Sects. 5.2 and 5.3, regression analysis of the results of the study for the five results are shown in Table 11.

Table 11 Study hypothesis verification results

Research hypothesis	Validation results
H1: "Perception of interest" has a positive effect on "perceptual usefulness"	Valid
H2: "Perceived ease of use has a positive effect on" perceived usefulness"	Valid
H3: "Perceived usefulness has a positive effect on" attitude toward use"	Valid
H4: "Perception of interest has a positive effect on" attitude towards use"	Valid
H5: "Perceived ease of use has a positive effect on" attitude toward use"	Valid

Table 12 The comparisons of a digital tour and traditional tour

Comparisons	Traditional guide	Digital tour
Area	Restricted	Unrestricted
Timeliness	Restricted	Anytime
Interactivity	Passive	Initiative
Interesting	Boring	Interesting
Learning effectiveness	Smaller	Larger

5 Conclusion

The information technology is booming, and the media is gradually presented in the form of digitization. From the aspects of space, mobility, extensibility and information presentation, the digitalized personal navigation has the characteristics of lightness and portability and establishes an interactive mode. The navigation system allows viewers to control the order and progress of viewing.

Comparing with traditional navigation methods (as shown in Table 12), it is found that digital tour guides give more interesting and has no geographical and time constraints. It can also greatly enhance learning outcomes.

The purpose of this study is to enable people who come to visit to enhance their learning interest and effectiveness through the multimedia navigation system. With the introduction of staff in the plant area, this study will enable them to better understand the procedures and processes of domestic wastewater treatment and deepen the learning impression through pre-and post-test analysis learning result. Finally, the use of interactive games for the public after the use of feelings to use SPSS TAM model analysis. From the "Perceived Enjoyment" and "Perceived Ease of Use" on the "Perceived Usefulness" validation analysis of the data show that people prefer to use the game because it is fun and easy to operate, it is relatively considered that the game can help their learning. From the "Perceived Enjoyment", "Perceived Ease of Use" and "Perceived Usefulness" on the "Attitude" verification analysis of the data show that when people feel the game is fun and easy to operate, it will be relatively considered that the game can help learning, and then increase the attitude of the public to use the game, will want to use this game. The results show that our students are generally aware that our platform is easy to use and useful and that they have also succeeded in arousing their strong interest and sense of participation. They are also willing to use this platform to learn and hope to bring more technology and navigation to the future. The combination, such as AR, VR, etc., to make the tour richer content is not limited to the factory.

References

1. Biological treatment method. http://www.ecaa.ntu.edu.tw/weifang/water/NEW2/%E7%94%9F%E7%89%A9%E8%99%95%E7%90%86%E6%B3%95.htm
2. Domestic sewage, family sewage control everyone together. http://enews.open2u.com.tw/~noupd/book_up/2605/13912.htm
3. Chen W Futian water recycling center tour. https://market.cloud.edu.tw/resources/web/1630120
4. Chen R, Ou J Willingness to buy smartphones with technology acceptance model—taking iPhone as an example
5. Yu S (2003) Interactive design of personalized digital guided tour—a case study of the palace museum "Emperor Qianlong's cultural undertakings"

Adversarial Attacks on SDN-Based Deep Learning IDS System

Chi-Hsuan Huang, Tsung-Han Lee, Lin-huang Chang, Jhih-Ren Lin and Gwoboa Horng

Abstract In recent years, software defined networking (SDN) has become a novel network architecture and design by employing manageable software between the control and data planes. Many SDN-based intrusion detection systems (IDS) have been proposed in recent researches. On the other hand, deep learning has emerged as an explosive growth in research and industry. The application of deep learning on IDS has been studied in many security-critical scenarios. However, some reports has indicated the vulnerability of adversarial attacks on deep learning IDS system with intentional perturbation injection or manipulation. It is important for the empowered IDS system employing deep learning to be responsible for not leading misclassification. Therefore, exploring the impact on the adversarial attacks over SDN networks or security-critical scenarios applying deep learning is an important step to confront such urgent issues. In this paper, we will conduct the SDN-based experiments on adversarial attacks for deep learning detecting system. We propose a novel class of adversarial attacks that exploits the vulnerability of the deep learning classifiers in SDN environment. Three typical deep learning models combining with four different adversarial testing will be conducted in the simulation for complete analysis.

C.-H. Huang · G. Horng
Department of Computer Science and Engineering,
National Chung-Hsing University, Taichung, Taiwan
e-mail: d102056016@mail.nchu.edu.tw

G. Horng
e-mail: gbhorng@cs.nchu.edu.tw

T.-H. Lee · L. Chang (✉) · J.-R. Lin
Department of Computer Science, National Taichung
University of Education, Taichung, Taiwan
e-mail: lchang@mail.ntcu.edu.tw

T.-H. Lee
e-mail: thlee@mail.ntcu.edu.tw

J.-R. Lin
e-mail: bcs105103@gm.ntcu.edu.tw

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_17

Keywords Adversarial attacks · SDN · Deep learning · IDS

1 Introduction

The Software Defined Networking (SDN), decoupling the control and data plans from network devices, provides a global view of the entire network with flexible, configurable and programmable design in the centralized controller. The networking devices, SDN switches, on the other hand provide the packets [1] forwarding with the flexible functionality. By configuring different network applications from the SDN controller, the SDN switches could be functioned as routers, switches or with networking specific functions, such as firewalls.

Many SDN-based Intrusion Detection Systems (IDS) have been proposed in recent researches. The port scan [2] is one of evasion techniques to evade IDS which consequently leads to a challenge to provide SDN-based IDS services. A port scanning attacker, such as network mapper (Nmap) [3], sends client requests to a number of sequential ports addressed on a host or server to find an active port or acquire advanced information about what services or operating system are behind them. Port scan is noisy because it causes the IDS system raising an alarm due to the logging of the sender IP. Usually, the source IP in the intrusive packets is spoofed to keep IDS from tracing original source using attacking packets. A large number of request packets with sequential port numbers result in Packet-In message request flood sending from SDN switches to the SDN controller, and consequently cause network performance degradation.

On the other hand, deep learning has emerged as an explosive growth in research and industry. Deep learning has achieved great success in many domains, such as image detection and speech recognition. There are a significant amount of deep learning based applications or network deployment being used or planned to be implemented in SDN networks or IDS systems currently. The empowered applications employing deep learning need to be responsible to be secure and accurate during and after learning process, respectively.

However, the deep learning system have been reported to be vulnerable to well-designed adversarial examples in SDN networks or IDS systems [4]. The deep learning system can be easily tricked in the deploying or testing process by adversarial examples. That means the deep learning system is vulnerable to well-designed adversarial input samples with little perturbations which however are imperceptible to users. For the application of deep learning in SDN networks or security-critical scenarios, it is very important to take into serious consideration about the vulnerability of the adversarial examples. The attacks and defenses on adversarial examples would draw great attention on research and implementation.

The traditional signature or rule based approaches of intrusion detection usually focuses on the defined attack patterns which require network behavior analysis and domain knowledge. The application of machine learning or deep learning on the intrusion detection will release the dependency on domain knowledge and heavy

load of data collection and analysis. It actually provides more flexibility and scalability on intrusion detection issues.

A deep learning approach for SDN-based port scan detecting system has been proposed in this paper to solve the port scan issue in SDN network using deep learning. Furthermore, exploring the impact on the adversarial attacks over SDN networks or security-critical scenarios applying deep learning is an important step to confront such urgent issues. We conduct the SDN-based experiments on adversarial attacks for deep learning detecting system. The experiments will collect the Packet-In messages and STATS reports received by the controller to create the SDN-based IDS dataset. The proposed port scan attack detecting system is majorly based on the frequency of received Packet-In messages and pre-defined features for the deep learning. Three typical deep learning models, Multilayer Perceptron (MLP) [5], Convolutional Neural Network (CNN) [6] and Long Short-Term Memory (LSTM) [7], combining with four different adversarial testing will be studied. In this paper, we will use Fast Gradient Sign Method (FGSM) [8], Jacobian-based Saliency Map Attack (JSMA) [9] and JSMA reverse (JSMA-RE) schemes as perturbation function for the adversarial attacks. As comparison, we will also examine the result of normal testing dataset. So, we will have totally 4 different processing. The detailed discussion of the adversarial attacks and analyses will be addressed in the next sections.

The rest of this paper is organized as follows. In Sect. 2, we address some related works corresponding to this research. We then describes the experimental deep learning models, architecture, and design in Sect. 3. Section 4 presents the adversarial attacks and system analyses, followed by the conclusion addressed in Sect. 5.

2 Related Works

We will discuss some related topics and works corresponding to this research in the next sub-sections.

2.1 OpenFlow Protocol in SDN

With the essential ideas of decoupling the data and control planes, using switches and controllers respectively, from traditional networks the SDN designs an architecture which provides programmable and flexible platform with global visibility to reduce the complexity of networking devices. The OpenFlow protocol [10], being the communication standard for SDN networks, defines the communication between the controllers and switches. It is defined as the south bound communication of SDN networks. The management and maintenance costs could be lessened in SDN networks by deploying the SDN controller with centralized control which dynamically configures the traffic forwarding rules to and collects statistic data from OpenFlow-enabled switches at per port or per flow granularity levels.

We use Ryu controller [11] in this research for SDN networking environments. Ryu controller with OpenFlow-enabled switches are widely used for SDN networking because they are easy to be deployed and expanded with simple architecture which is especially popular for lightweight traffic communication and control in small to medium-sized networking architectures.

2.2 Deep Learning Systems and Architectures

Deep learning network, defining multiple simple features to solve the complicated problems, is a machine learning algorithm with many layers of neural networks. Deep learning has achieved successful results in various areas such as image recognition, speech recognition, natural language processing and so on.

In this research, we conduct three different deep learning models, MLP, CNN and LSTM for our experiments using TensorFlow open source library as deep learning platform.

2.3 Flow-Based DDoS Detection

In the research [12], the authors proposed a detection scheme to identify the distributed denial of service (DDoS) attacks by analyzing the statistics of each flow in the SDN flow table. In SDN network, the traffic is treated as flow-based which is different from the traditional packet-based mechanism. To identify the abnormality of traditional packet-based traffic, the system needs to collect a large number of packets for analysis. The flow-based mechanism on the other hand only requires the statistical data of each flow to determine the irregularity of the traffic. However, the generation of the flow statistics may cause some delay resulting in the victim host being attacked until an abnormal flow is detected.

2.4 Adversarial Examples in Deep Learning

The adversarial attacks against deep learning detecting system in an effort to reduce its accuracy and effectiveness or extract information from it is a study of adversarial examples in deep learning. All aspects and phases of the deep learning processes, including learning model, training and testing dataset, could be attacked by an adversary. Researches have shown that a small variations being imperceptible to humans on an image could lead to misclassification of the deep learning model. On the other hand, it was studied that some noise-like or various non-described patterns could fool a deep learning model to classify them in a specific class with a high confidence.

Some studies have been carried out to generate adversarial samples which produce a small perturbation in the original sample. The Fast Gradient Sign Method (FGSM) proposed by Goodfellow [8] is one famous scheme to generate adversarial attacks with relatively effective computation. Jacobian-based Saliency Map Attack (JSMA) [9] is another famous scheme which provides the ability to generate adversarial samples with less distortion degree. JSMA however, needs more computational resource than FGSM because JSMA changes the targeted class by exploring one attacked input feature or dimension for each iteration.

In FGSM, a perturbation ρ of the generated adversarial example can be obtained by computing the gradient of the cost function J with respect to the original input sample x, as shown in the following equation.

$$\rho = \varepsilon \, \mathrm{sign}(\nabla \mathrm{x} \mathrm{J}(\theta, \mathrm{x}, \mathrm{y}))$$

where ε is a very small scalar number to restrict the small perturbation, θ are model parameters, x is the input to the model, y is the label associated with x, and J is the cost function used to train the deep neural network. FGSM utilizes backpropagation in the neural network for gradient component computation so that the scheme could be processed very fast.

3 Deep Learning Models, Architecture, and Design

3.1 Deep Learning Models

Deep learning network, defining multiple simple features to solve the complicated problems, is a machine learning algorithm with many layers of neural networks. In this paper, we conduct three different deep learning models, MLP, CNN and LSTM for our experiments using TensorFlow open source library as deep learning platform.

3.2 Simulation Architecture

The experimental topology is illustrated in Fig. 1, where Mininet [13] was used to simulate SDN networking environments. Mininet is a lightweight SDN/OpenFlow network simulator that supports the OpenFlow protocol and allows users to create independent network nodes. Several OpenFlow-enabled switches, such as Open vSwitch [14], can be simulated on a physical host to create a complex SDN network. Special attributes such as bandwidth and delay can be added or modified from the controller to switches for specific flow handling and actions. As shown in Fig. 1, the network topology, created from Mininet simulator, consists of a host with Ryu controller and deep learning system, an OpenFlow switch, a host serving

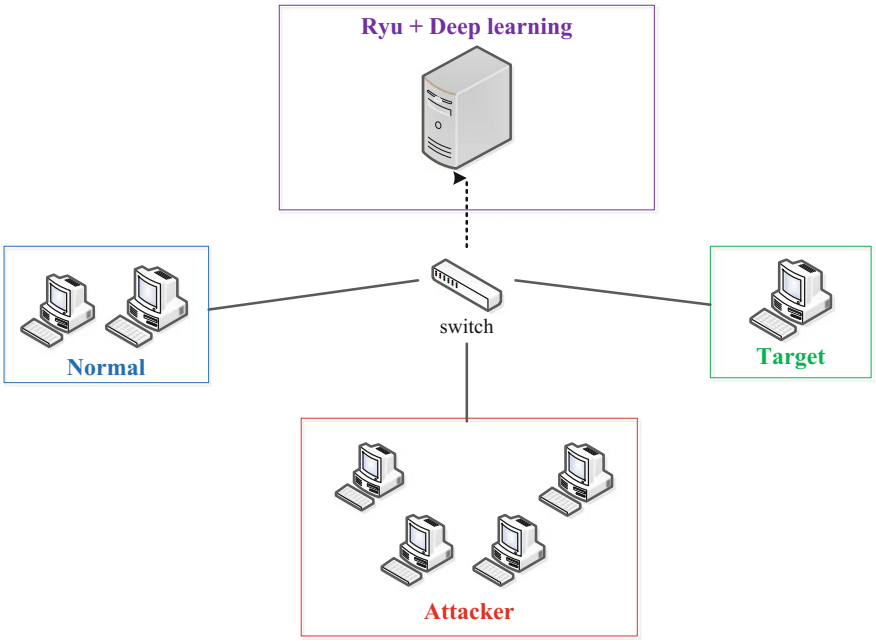


Fig. 1 Experimental topology

as targeted server, two hosts generating normal packets, and four hosts acting as port scanning attackers.

We integrated Ryu controller with three different deep learning models, MLP, CNN and LSTM, using TensorFlow open source library as deep learning platform for different experiments and analyses. The port scanning attackers employ Network Mapper (Nmap) [3] to generate a large number of probing attacks on the targeted server. Nmap is a security port scanner which sends special crafter packets to the targeted host or even adapting to network conditions, such as congestion and latency, during the scanning process to provide advanced service detection or build a map of the scanned network.

3.3 Experimental Design

The port-scan attack detecting system proposed in this paper is based on the frequency of Packet-In messages received by SDN Ryu controller. For instance, if an SDN switch sends more than 200 Packet-In messages, being received by Ryu within 2 s, the controller will initiate the detection mode from the deep learning system. The deep learning system will extract the pre-defined features from the Packet-In messages and conduct the attack detecting process to determine whether

the messages are attacks. The process flow of the proposed deep learning detecting system is shown in Fig. 2. The details of the procedures and mechanism of deep learning detecting system are discussed as following.

Port scan attack detecting mechanism: Based on the original OpenFlow protocol, the SDN switch will initiate a Packet-In message to the controller when the arriving packets do not match the entries in the Flow Table. In this paper, we set the Flow Table rules using network and transport layers, such as IP addresses and transport port numbers, instead of using MAC address in link layer only. This setting will allow the switch to initiate all Packet-In messages, with respect to the port scan packets, to the controller because the switch treats all port scan packets as different flows. The controller will count the number of Packet-In messages for each specific port number. If the number for one port is greater than 200 within 2 s, the controller extracts the pre-defined features from all corresponding Packet-In messages for deep learning prediction test.

Feature extracting mechanism: Each Packet-in message received by the controller is extracted pre-defined features in the deep learning detecting system. For the port scan attacks in this research, we define six features, including packet length, udp packet count, tcp packet count, icmp packet count, udp port count, and tcp port count, for deep learning detection.

Dataset collection: The deep learning models were trained and optimized with SDN training dataset and predicted with testing dataset as described in port scan attack detecting mechanism above [15, 16]. Besides of the detecting mechanism for the controller to collect the Packet-In messages, the controller will monitor the switch using statistic messages (STATS). We implement the switch a counter for each port. The switch will send an STATS report to the controller in every 2 s. By collecting the STATS reports as well as Packet-In messages, the controller passes these datasets to the deep learning detecting system for training processing and model building. To collect SDN-based port scan attack dataset, we employ wireshark to capture all the Packet-In messages as well as STATS reports in the

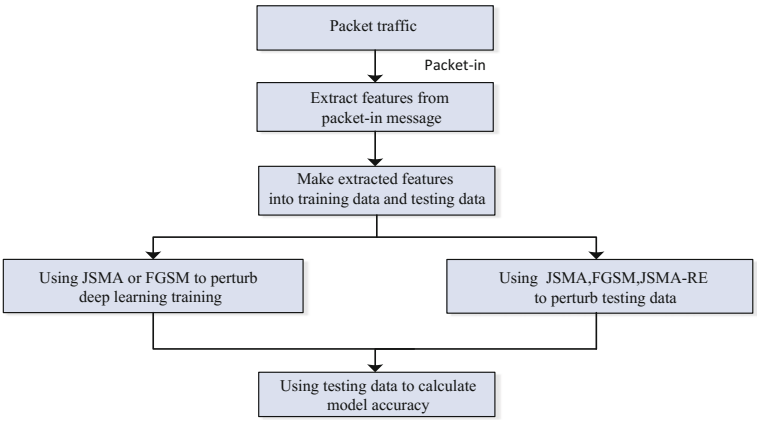


Fig. 2 The process flow of the proposed deep learning detecting system

controller and store them with pcap file format. By using tshark tool, we further transfer the pcap file into csv file format with pre-defined features as discussed in the feature extracting mechanism above. The collected dataset could be applied to both training and testing procedures of deep learning detecting system.

Adversarial attacks: As mentioned above, the deep learning models used in this research are MLP, CNN, and LSTM, which therefore are source models for the generated adversarial samples. With the knowledge of system model or the capability of training dataset access, these deep learning models could be the target models during the training phase. In this research, we will apply FGSM, JSMA and JSMA reverse (JSMA-RE) mechanisms as perturbation function for the adversarial attacks. The detailed discussion of the adversarial attacks and analyses will be addressed in the next section.

4 Adversarial Attacks and System Analyses

4.1 Deep Learning and Adversarial Attacking Models

In this research, we apply three different adversarial attack mechanisms, FGSM, JSMA and JSMA-RE, to three different deep learning models, MLP, CNN, and LSTM, to evaluate the impact of the adversarial attacks for different scenarios. Before the deep learning detecting system begins to detect attacks, we must provide a large number of normal and attack traffic datasets collected from the SDN network to train the deep learning models. The deep learning models are capable of detecting port scan attack by using the Packet-In messages and STATS reports as training dataset. Table 1 shows the training parameters used in the deep learning system during training stage [17].

4.2 Adversarial Attacks and Processing

Three different adversarial attacks, FGSM, JSMA and JSMA-RE schemes, plus a normal testing as comparison were conducted in this paper. As described in the related works, JSMA scheme exhibits adversarial characteristics by generating a small perturbation with the input dimensions or features in the original sample to create a targeted class change.

Table 1 Training parameter

Parameters	Values
Learning rate	0.1
Batch size	64
Epoch	256

For the port scan attacks, JSMA scheme basically focuses on the perturbation of positive samples meaning that JSMA will only generate small perturbations on the port scan attacking packets but not on the normal packets. Therefore, in this research, we further modify the program and function of JSMA and design the JSMA reverse (JSMA-RE) scheme which on the other hand generates small perturbations on the normal packets to create fault negative results. That means JSMA-RE provides the perturbation from normal packets to be classified as port scan attacking packets, while JSMA trying to change the features of the testing dataset from attacking packets to normal packets.

In this research we conduct adversarial sample generation with four different adversarial test datasets, including normal test dataset, FGSM adversarial test dataset, JSMA adversarial test dataset, and JSMA-RE test dataset, for three different deep learning models. Only JSMA has two different types of adversarial test datasets because it is more suitable for targeted misclassification. With the knowledge of system model or the capability of training dataset access, these deep learning models could become the target models during the training phase. Hence, we apply FGSM and JSMA schemes to the original dataset collected from SDN-based environment to perturb the inputs during training phase.

4.3 System Analyses

More than 3400 samples, including 1166 port scan attacks and 2238 normal packets, were collected as the SDN-based training dataset. For testing dataset, we collected 896 port scan attacks and 693 normal packets. The experimental result for the normal adversarial phase is shown in Table 2. The rightmost column is the training time using normal SDN-based training dataset. Three deep learning models combining with 4 different adversarial tests result in a 3×4 data matrix. Similar results for JSMA and FGSM adversarial phases are tabulated in Tables 3 and 4, respectively. We will have the following discussion about experimental results.

First of all, JSMA adversarial attacks cause a relatively significant impact, from 14 to 42% on different deep learning models. JSMA-RE did not reduce any accuracy for both CNN and LSTM, however, 35% accuracy drop was observed for MLP model. Secondly, FGSM in general did not perform well for testing dataset

Table 2 Normal adversarial phase

Model	Phase				
	Normal adversarial phase				Training Time(s)
	Normal	jsma	fgsm	jsma-re	
	Predictions				
MLP	0.8610	0.725	0.8369	0.614	87.3
CNN	0.9791	0.5631	0.9970	0.9876	21
LSTM	0.9806	0.5637	0.4209	0.9913	102

Table 3 JSMA adversarial phase

Model	Phase				
	Normal adversarial phase				Training Time(s)
	Normal	jsma	fgsm	jsma-re	
	Predictions				
MLP	0.9217	0.9166	0.8944	0.8013	771.3
CNN	0.9979	0.9967	0.9934	0.9978	592.7
LSTM	0.9983	0.9986	0.5923	0.996	1223

Table 4 FGSM adversarial phase

Model	Phase				
	Normal adversarial phase				Training Time(s)
	Normal	jsma	fgsm	jsma-re	
	Predictions				
MLP	0.704	0.48391	0.6393	0.7167	280.7
CNN	0.9983	0.9986	0.5923	0.996	1223
LSTM	0.4833	0.4834	0.7131	0.4453	327.7

perturbation, while causing a significant drop of accuracy, more than 50%, for LSTM. Thirdly, both FGSM and JSMA perturbation models spend much more training time (10 times more) than that of normal adversarial phase. However, FGSM completes the training with less time than JSMA. Fourthly, regardless of which model, JSMA adversarial testing dataset provides some impacts on accuracy when training accuracy is as high as 98% under normal adversarial phase.

Lastly, the proposed JSMA-RE adversarial testing dataset provides a better effect on dropping MLP model accuracy for both normal and JSMA adversarial phases. From the overall results, JSMA is more suitable for use in perturbing testing data, nevertheless FGSM is more suitable in perturbing training model.

5 Conclusions

In this paper, we have conducted the SDN-based experiments on adversarial attacks for deep learning detecting system. We have collected the Packet-In messages and STATS reports received by the controller to create the SDN-based IDS dataset. The proposed port scan attack detecting system is majorly based on the frequency of received Packet-In messages and pre-defined features for the deep learning. Three typical deep learning models, MPL, CNN and LSTM, combining with four different adversarial testing have been studied and analyzed. In general, JSMA adversarial attacks make a relatively high impact on different deep learning models. Especially, significant drops of accuracy for JSMA adversarial testing were observed with higher accuracy under normal adversarial phase. On the other hand, FGSM did not perform well for testing dataset perturbation. The proposed JSMA-RE scheme provides about 35% accuracy drop for MLP model while showing no significant

influence on both CNN and LSTM models. Both FGSM and JSMA perturbation models spent more than 10 times on training time than that of normal adversarial phase. Judging from the overall experimental results and discussion, we can conclude that JSMA is more suitable for use in perturbing testing data for port scan attacks, nevertheless FGSM is more suitable in perturbing training model.

In terms of future research, we will further conduct more studies on adversarial attacks on different SDN-based Deep Learning IDS Systems, such as DDoS, U2L and R2L. This paper will also be extended to the attacks and defenses on the design and implementation of SDN-based deep learning.

Acknowledgements This research was supported by research grants (MOST 105-2221-E-142-001-MY2 and 105-2221-E-142-002-MY2) from Ministry of Science and Technology, Taiwan.

References

1. Software Defined Network [Online]. https://www.opennetworking.org/sdndefinition/
2. TCP Port Scan with Nmap [Online]. https://pentest-tools.com/network-vulnerability-scanning/tcp-port-scanner-online-nmap
3. Nmap [Online]. https://nmap.org/
4. Almeida LB (1997) Multilayer perceptrons, in Handbook of Neural Computation, IOP Publishing Ltd and Oxford University Press
5. LeCun, Yann. LeNet-5, convolutional neural networks. Retrieved 16 November 2013 Convolutional Neural Network (CNN)
6. Sepp H, Jürgen S (1997) Long short-term memory. Neural Comput 9(8):1735–1780. 10.1162/neco.1997.9.8.1735. PMID 9377276
7. Yuan X, He P, Zhu Q, Bhat R, Li X (2017) Adversarial examples: attacks and defenses for deep learning at https://www.researchgate.net/publication/321936593
8. Goodfellow IJ, Shlens J, Szegedy C (2014) Explaining and harnessing adversarial examples. arXiv preprint. arXiv:1412.6572
9. Papernot N, McDaniel P, Goodfellow I, Jha S, Celik ZB, Swami A (2016) Practical black-box attacks against deep learning systems using adversarial examples. arXiv preprint. arXiv:1602.02697
10. OpenFlow [Online]. https://github.com/OSE-Lab/Learning-SDN/tree/master/Protocols/OpenFlow
11. Ryu SDN Framework [Online]. https://osrg.github.io/ryu/
12. Braga R, Mota E, Passito A (2010) Lightweight DDoS flooding attack detection using NOX/OpenFlow. In: Annual IEEE conference on local computer networks, Oct. 2010
13. Mininet [Online]. http://mininet.org/
14. Open vSwitch [Online]. https://www.openvswitch.org/
15. Tang TA, Mhamdi L, McLernon D, Zaidi S, Ghogho M (2016) Deep learning approach for network intrusion detection in software defined networking. Wireless Networks and Mobile Communications, Oct. 2016
16. Niyaz Q, Sun W, Javaid AY (2016) A deep learning based DDOS detection system in software-defined newtoking (SDN). ArXiv:1611.07400
17. Maria Rigaki (2017) Adversarial deep learning against intrusion detection classifiers. http://www.diva-portal.org/smash/get/diva2:1116037/FULLTEXT01.pdf

Indoor Localization with Adaptive Channel Model Estimation in WiFi Networks

Yung-Fa Huang and Yi-Hsiang Hsu

Abstract This paper investigates the indoor localization based on channel estimation with the received signal strength (RSS) of three access points (APs) in Wi-Fi networks. In this paper, the proposed adaptive channel estimation for the path loss exponent (PLE) is based on the root mean square error (RMSE) of RSS of the APs. The PLE effects are investigated and then an adaptive PLE estimation is proposed to improve the indoor localization. Experimental results show that the proposed adaptive modeling of PLEs can effectively decrease the localization error.

Keywords Indoor localization · Received signal strength · Path loss exponent Root mean square error

1 Introduction

In the applications of wireless communication, the localization with global position systems (GPS) has become a very important technique for mobilization issues for the military and commercial [1] in the outdoor environment. However, in indoor environments, the GPS can't be effectively utilized with the house blocking the communication links. Therefore, the indoor localization methods become the important research topics [2].

Developments in the Internet of things (IoT) [3] makes it possible to build smart cities in which traffic congestion, parking, street lighting, and urban noise are monitored and managed in real time. Heterogeneous information systems can be used for environmental monitoring, military surveillance, and object tracking [4]. However, the above applications require accurate positioning information.

Indoor or outdoor location-based services (LBSs) are expanding rapidly [5]. The global positioning system (GPS) is widely used outdoors [6]; however, it does

Y.-F. Huang (✉) · Y.-H. Hsu
Department of Information and Communication Engineering,
Chaoyang University of Technology, Taichung 41368, Taiwan
e-mail: yfahuang@cyut.edu.tw

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_18

not perform well indoors, due to the blocking of radio waves from satellites. Furthermore, indoor environments are prone to interference from moving bodies, multi-path effects and shadowing, which has necessitated the development of positioning systems especially for indoor environments [7, 8].

In wireless channels, the signal will be subject to different shadow fading effect [9], thus this article will explore the effects of different shadow fading effects on localization estimation errors. In [10], the radio channel fingerprints method has been performed for the localization with long-term evolution (LTE) systems. However, in indoor localization, the shadowing fading effect will be deteriorated for LTE received signal. Therefore, this article will focus RSS based localization. Moreover, the root mean square error (RMSE) distance estimation method is proposed to adaptively compensate the multipath fading effect in Wi-Fi networks.

2 RMSE Based Localization

In the range-based localization method, we estimate the unknown location based on the estimated distance between the test point to from three reference nodes. The three distance can be estimated by the channel model [11]

$$\hat{d}_m^j(n) = 10^{\frac{P(d_0) - P_m^j(d_m^j)}{10n}} \tag{1}$$

where $\hat{d}_m^j$ is the estimated distance between the mth test point TP_m and the jth access point AP_j, d_0 is a reference distance (d_0 = 1 m), $P(d_0)$(in dBm) is the received signal strength (RSS) at distance d_0 to AP_j, $P_m^j(d_m^j)$(in dBm) is the measured RSS at distance d_m^j between the mth test point TP_m and the jth access point AP_j, n is the path loss exponent (PLE) for n = 1.2–3.0.

In order to minimize the location error, we investigate the RMSE of distance estimation for different PLEs, n by

$$d_{m,RMSE}^j(n) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{d}_{m,i}^j(n) - d_m^j)^2} \tag{2}$$

where N is the number of measurements. Then the adaptive channel model of PLEs n_{ad} can be find from the minimum d_{RMSE} for the channel of the mth test point TP_m to the jth access point AP_j by

$$n_{m,ad}^j = \arg\left\{\min_n d_{m,RMSE}^j(n)\right\} \tag{3}$$

In the estimation of distance between the test points and the reference APs, the RSS of the test point will be affected to varying degrees of PLE [10]. Therefore, the minimum mean square error (MMSE) algorithm can be used to minimize the

localization errors. The RSS at the receiver is attenuated with the distance in wireless communication channels [9]. To construct the MMSE method for localization problems, we formulate the estimation error equation for test position (x_0, y_0) by

$$\hat{d}_n = \sqrt{(x_n - x_0)^2 + (y_n - y_0)^2}, \tag{4}$$

where x_0 and y_0 are the coordinates of the test node to be estimated, x_n and y_n are the known coordinates of the nth reference APs, $\hat{d}_n$ is the estimated distance for the distance between the test node and the nth AP node. When the number of reference nodes is 3, the index n would be 1, 2, 3. There are three simultaneous equations in (4).

Then through mathematical operations, we can derived by

$$\hat{d}_1^2 - \hat{d}_2^2 = -2x_1x_0 - 2y_2y_0 - x_2^2 - y_2^2 + 2x_2x_0 - 2y_2y_0 + x_1^2 + y_1^2 \tag{5}$$

and

$$\hat{d}_1^2 - \hat{d}_3^2 = -2x_1x_0 - 2y_3y_0 - x_3^2 - y_3^2 + 2x_3x_0 - 2y_3y_0 + x_1^2 + y_1^2 \tag{6}$$

Using the Least Square (LS) method, we obtain the estimated coordinates of the test node by

$$\hat{\mathbf{b}} = \begin{bmatrix} \hat{x}_0 \\ \hat{y}_0 \end{bmatrix} = \mathbf{X}^{-1}\mathbf{w}. \tag{7}$$

where

$$\mathbf{X}^{-1} = \begin{bmatrix} 2(x_2 - x_1) & 2(y_2 - y_1) \\ 2(x_3 - x_1) & 2(y_3 - y_1) \end{bmatrix}^{-1} \tag{8}$$

and

$$\mathbf{w} = \begin{bmatrix} \hat{d}_1^2 - \hat{d}_2^2 + x_2^2 + y_1^2 - x_1^2 - y_1^2 \\ \hat{d}_1^2 - \hat{d}_3^2 + x_3^2 + y_3^2 - x_1^2 - y_1^2 \end{bmatrix} \tag{9}$$

3 Experimental Results

To investigate the RMSE performance of the trilateral localization, we perform experiments for the localization. In the experimental area, there are nine test nodes (marked by TP1, TP2, TP3, TP4, TP5, TP6, TP7, TP8, TP9) to be examined for three reference nodes AP1, AP2 and AP3 in a room. The coordinates of the test points and reference nodes AP1, AP2 and AP3 are shown in Fig. 1. The other parameters are shown in Table 1.

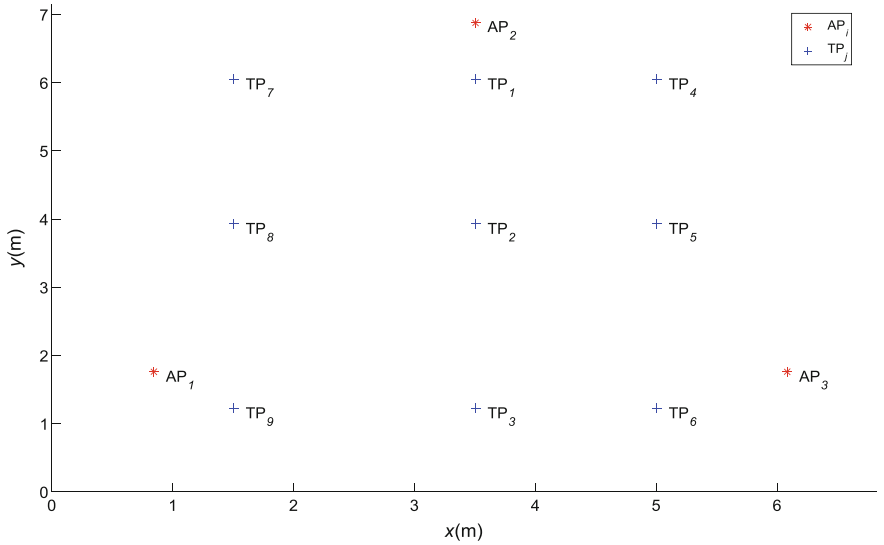


Fig. 1 The localization area with three reference nodes and seven estimated positions

Table 1 Simulation Parameters

Parameters	Value
Number of measurements (N)	300
Coordinates of the reference nodes (x_0, y_0)	AP1: (0.84, 1.76), AP2: (3.50, 6.88) AP3: (6.08, 1.76)
Coordinates of the test points (x_0, y_0)	TP1:(3.5, 6.05), TP2: (3.5,3.93), TP3: (3.5, 1.22), TP4: (5, 6.05), TP5: (5, 3.93), TP6: (5, 1.22) TP7 (1.5, 6.05), TP8 (1.5, 3.93), TP9 (1.5, 1.22)
Path loss exponent (n)	1.2–3
Transmitting power (dBm)	5

At first, we examine the RMSE of the nodes TP_m, m = 1,…,9. To investigate the effect of distance between the unknown node and reference node, we find the reference RSS, $P(d_0)$ at d_0 = 1 m far from the AP_j for TP_m to AP_j as shown in Table 2.

To find the suitable PLE for various TP_m to AP_j, we estimate the distance $\hat{d}_m^j$ between TP_m and AP_j by using different PLE n = 1.2–3. Figure 2 shows that the distance estimation examples for TP_m to AP_j with different PLE**s** n = 1.2, 2, and 3 by (1). From Fig. 2, it is observed that the distance estimation of n = 2 is closer to the actual distance than n = 1.2 and 3 for TP2 to AP3. Thus according to (3), we select the adaptive PLE by $n_{1,ad}^1 = 2$ for the channel of TP2 to AP3. Similarly, we can examine all the adaptive PLEs as shown in Table 3.

Table 2 The reference RSS, $P(d_0)$ at d_0 = 1 m far from the AP_j for TP_m to AP_j

$P(d_0)$ (dBm)	AP1	AP2	AP3
TP1	−26.885	−26.18	−26.6617
TP2	−26.885	−26.18	−26.4383
TP3	−26.885	−26.18	−28.9133
TP4	−26.885	−29.3857	−28.54
TP5	−28.7283	−31.58	−28.6617
TP6	−34.0117	−31.58	−29.5483
TP7	−26.885	−37.3267	−26.4383
TP8	−26.885	−31.58	−26.4383
TP9	−39.7367	−31.58	−32.54

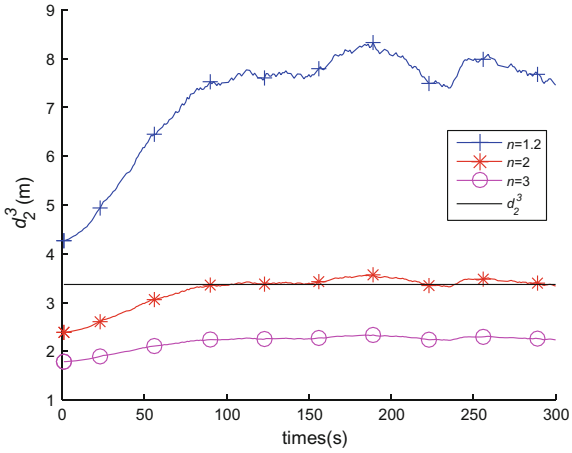


Fig. 2 The distance estimation examples: $\hat{d}_2^3$ for TP2 to AP3 with different PLEs

Moreover, we further investigate the location error comparison for adaptive PLE and constant PLE n_c = 1.2, 1.5, 2, 2.5, 3, for nine different TP_m, respectively, as shown in Table 4. From Table 4, it is observed that the localization distance errors of adaptive PLE for nine test points are all smaller than that of constant PLEs. Most over, the average localization error of n_{ad} is 1.53 m smaller than that of minimum localization error 2.85 m of n_c = 2.5.

Table 3 The adaptive PLE for different channel between the various TP_m to AP_j

$n_{m,ad}^{j}$	AP_1	AP_2	AP_3
TP1	2.2	3	2.5
TP2	1.2	2.5	2
TP3	2.8	2.1	1.4
TP4	1.7	3	3
TP5	2.1	1.4	2.7
TP6	1.5	1.5	3
TP7	2	3	1.4
TP8	3	2.7	1.6
TP9	3	1.2	2.5

Table 4 The location error comparison for adaptive PLE and constant PLE for nine different TP_m, respectively

e_n(m)	n_c					n_{ad}
	1.2	1.5	2	2.5	3	
TP1	22.42	8.89	2.22	1.86	2.75	1.33
TP2	7.54	3.49	1.6	1.8	2.2	0.48
TP3	17.7	7.08	1.61	1.62	2.5	0.25
TP4	22.87	9.59	3.21	2.67	3.29	1
TP5	11.19	5.31	2.49	2.39	2.73	1.53
TP6	3.11	3.44	4.15	4.57	4.82	3.07
TP7	11.79	5.29	3.7	4.18	4.63	2.34
TP8	14.77	7.03	3.29	2.88	3.13	2.21
TP9	17.56	7.47	3.55	3.66	4.14	1.58
Average	14.33	6.40	2.87	2.85	3.35	1.53

4 Conclusion

In this paper, we investigate the indoor localization based on RMSE PLE estimation method. Experimental results show that the proposed adaptive PLE can decrease the average distance error to 1.53 m with nine test positions and three reference AP nodes.

Acknowledgements This work was funded in part by Ministry of Science and Technology of Taiwan under Grant MOST 106-2221-E-324-020.

References

1. Akyildiz IF, Su W, Sankarasubramaniam Y, Cayirci E (2002) A survey on sensor networks. IEEE Commun Mag 40(8):102–114
2. Royer EM, Toh CK (1999) A review of current routing protocols for Ad-hoc mobile wireless networks. IEEE Pers Commun 6(4):46–55

3. Boman J, Taylor J, Ngu AH (2014) Flexible IoT middleware for integration of things and applications. In: IEEE international conference on collaborative computing: networking, applications and worksharing, Miami, Florida, pp 481–488
4. Xiong SM, Wang LM, Qu XQ, Zhan YZ (2009) Application research of WSN in precise agriculture irrigation. In: International conference on environmental science and information application technology, Wuhan, pp 297–300
5. Gui L, Val T, Wei A, Taktak S (2014) An adaptive range-free localization protocol in wireless sensor networks. Int J Ad Hoc Ubiquitous Comput 15(1/2/3):1–20
6. Lin A, Zhang J, Jiang X, Zhang J (2016) A reliable and energy-efficient outdoor localization method for smartphones. Int J Ad Hoc Ubiquitous Comput 23(3/4):230–242
7. Kharidia SA, Ye Q, Sampalli S, Cheng J, Du H, Wang L (2014) HILL: ahybrid indoor localization scheme. In: International conference on mobile Ad-hoc and sensor network, Maui, Hawaii, pp 201–206
8. Huang H, Zhou J, Li W, Zhang J, Zhang X, Hou G (2016) Wearable indoor localisation approach in Internet of Things. IET Netw 5(5):122–126
9. Dag T, Arsan T (2018) Received signal strength based least squares lateration algorithm for indoor localization. Comput Electr Eng 66:114–126
10. Ye X, Yin X, Cai X, Pérez Yuste A, Xu H (2017) Neural-network-assisted UE localization using radio-channel fingerprints in LTE networks. IEEE Access 5(5):12071–12087
11. Huang Y-F, Jheng Y-T, Chen H-C (2010) Performance of an MMSE based indoor localization with wireless sensor networks. In: The sixth international conference on networked computing and advanced information management (NCM 2010), Seoul, Korea, pp 671–675

Traffic Analysis of Important Road Junctions Based on Traffic Flow Indicators

Hung-Chi Chu and Chi-Kun Wang

Abstract Intelligent transportation systems can be used as an indicator of urban sophistication. Therefore, many countries devote themselves to developing an intelligent transportation system. The construction of intelligent transport system in Taiwan is still in infancy. The allocation of traffic signal lights requires a lot of manpower and time to observe the site. Based on collected information, data integrity was inadequate due to human negligence. The traffic congestion problems caused by the mistake can't be improved. In this article, the traffic flow based on the intersection traffic flow indicators analysis is proposed. This method is based on the relationship between data to judge, to avoid the problem of empirical judgment error without human intervention. Traffic flow are collected by the vehicle detector to improve data integrity. Using the indicator for road junction association analysis can find the association rules between road junctions. Finally, a deep neural network is used as its classifier, so the accuracy is over 95%.

Keywords Intelligent transportation system · Association analysis
Deep neural network

1 Introduction

As IT technology advances, more and more applications of Internet of Things are in the cities [1]. The future will be placed more equipment throughout the city for data collection and environmental monitoring. The information returned by these devices is very helpful for understanding the current traffic conditions, enabling intelligent transportation systems [2] to get an accurate picture of current traffic

H.-C. Chu (✉) · C.-K. Wang
Chaoyang University of Technology, Taichung, Wufeng District 41349, Taiwan
e-mail: hcchu@cyut.edu.tw

C.-K. Wang
e-mail: s10530605@gm.cyut.edu.tw

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_19

conditions and carrying out the allocation of traffic lights for the current situation [3]. The construction of intelligent transport system in Taiwan is still in infancy, and there are no clear guideline on how to deploy the equipment in the cities, and how to effectively use the information returned by the devices.

As a result, the current configuration of the existing traffic lights will require a great deal of manpower and time for the data collection and statistics, and will not ensure the integrity and correctness of the data as a whole, and it needs to rely on more experienced observers to determine the form of this intersection to configure initial traffic light cycle configuration, so for the more complex road junctions or the higher complexity of the traffic network system, there will be judgment errors.

In addition, the length of data collection and collection of time will also affect the observer's judgment [4]. When the configuration of traffic light cycle is based on the above method, the configuration can't meet the current traffic flow, leading to the overall traffic network performance decline. As the overall traffic network performance decreases, the operational efficiency of the city will also decrease. Traffic network performance is usually reduced due to serious congestions of many major junctions, which in turn affects the traffic congestion at the adjacent junctions and has a ripple effect. Therefore, it is quite important to determine the interdependence of the main intersections. It is necessary to find out the range of influence between major junctions and infer the correlation between intersections. The traffic lights cycle configuration can effectively relieve the overall traffic congestion based on the association, rather than only according to the current congestion at the junction traffic lights allocation.

Based on the above problems, this article presents an important road junction traffic flow indicators analysis. This indicator analysis is based on traffic growth at each road intersection. The indicators will be applied to all road junctions. Then, according to this index to divide the road junctions of related degree and deduce the correlation degree of each major road junction. Finally, we use the deep neural network [5] to build the model based on the indicator. The purpose is to configure the equipment at the secondary junctions in the future that can collect information about their intersection. Through the model, it can quickly determine the indicators of the road junctions. Through the indicators of the road junction analysis, the overall relevance will be more enhanced.

The dataset used in this article is provided by the government open data of Department of Transportation (DOT), Taipei City Government, with the data range of 230 important road junctions in Taipei City in the year 2016, and the hours from 7 to 9 am and from 5 to 7 pm, and sampling interval of 15 min. Through the method of this study, different index levels can be effectively divided, through which the correlations among various major intersections can be clearly identified with the final deep neural model accuracy of more than 95%.

2 Related Work

2.1 R Language

R language [6] is a cross-platform free software programming language and operating environment, and evolved from the S language. R is an integrated data processing software and statistical software, but also a graphics software. In the R language, the result output can be stored as object to provide subsequent calculations. This is also different from many statistical software such as: SAS, SPSS.

R language, with a number of built-in statistical and numerical analysis functions provided to the user, is also available on the official website to download many packages to enhance more ways to calculate. In this article, mxnet package is used to establish a deep neural network model. R language has a very powerful visualization capabilities that can effectively present the result of statistical or data analysis, and R language can improve computational efficiency through parallel computing.

2.2 Data Mining

The purpose of data mining [7] is to mine the information hidden in the data and is part of Knowledge Discovery. Data mining uses many methods of statistical analysis and modeling to look for useful features and associations in the data. Data mining involves aspects such as database and database management, data preprocessing, model and result inference considerations, and complexity considerations, finding structures and visualization post-processing.

When using traditional methods for analyzing large amounts of data, the way of analysis needs to be judged by the past experience of analysts. Therefore, when the initial steps are poorly selected, the overall analysis is wrong direction and the efficiency of the analysis is reduced. In addition, when conducting a large amount of data analysis, many of the information hidden in the data may not be able to be found by analysts, thus omitting some important information. Through the way of data exploration, it can make use of the selected method to mine the hidden information in the data. Data mining can construct the following six models: classification [8], regression [9], time series [10], cluster [11], association [12] and sequence [13]. Classification and regression are mainly used for forecasting. Associations and sequences are used to describe the behavior. In the absence of a clear category label that can be applied, the use of cluster can be used for unsupervised learning. Finally, if forecasting for the time-sensitive data, it can choose the time series of ways to analyze.

This article use the association analysis to analyze the important road junctions in Taipei to find the correlation between important junctions. Then this association

through the neural network establish a classification model to provide the next major road intersection information classification.

2.3 Deep Learning

Deep learning [14] is a branch of machine learning. An algorithm based on characterization of data. Data sources can be presented in a variety of ways. The advantage of deep learning is the use of unsupervised, semi-supervised learning features and hierarchical feature to extract high-performance algorithm instead of hand-made features. Therefore, the feature extraction at the road junction can be performed in this way, without losing important features.

The framework of deep learning evolved from the notion of neural networks of the class of the 1990s. Because at that time the computing performance of the computer was low and the amount of data was insufficient, the neural network was not effectively optimistic, and even once considered to be impossible to practice.

In 2006, however, there was a shift, mainly because Geoffrey Hinton, the father of deep learning, proposed Deep Belief Networks (DBN) to use unsupervised pre-training to optimize the initial weight of the network, followed by Fine-Tune used to truly implement multi-layer neural networks. Deep neural networks at that time were called deep learning. The existing deep learning framework includes deep neural networks, convolutional neural networks, deep belief networks and recurrent neural networks and many more. A wide range of applications are: video recognition [15], speech recognition [16], natural speech processing [17] and life applied to many related fields. This article uses deep neural networks to model the road junction traffic flow indicator classification and achieve the accuracy of more than 95%.

3 Traffic Flow Index Analysis

The dataset used in this article is provided by the government open data of Department of Transportation (DOT), Taipei City Government, with a data range of 230 important road junctions in Taipei City, 2016, the hours from 7 to 9 am and from 5 to 7 pm, and sampling interval of 15 min. Figure 1 shows the intersection of the 230 road junctions after the use of administrative divisions in Taipei.

According to this dataset, we can understand Traffic flow ratio of various road junctions at the same time point and the traffic growth ratio of a single intersection. Due to the continuity of the traffic flow, there will be a continuous relationship. So the traffic growth rate in the affected area will be similar. Based on this concept, road junction traffic flow indicators are analyzed. First of all, there are two kinds of normalization methods to process dataset, described below:

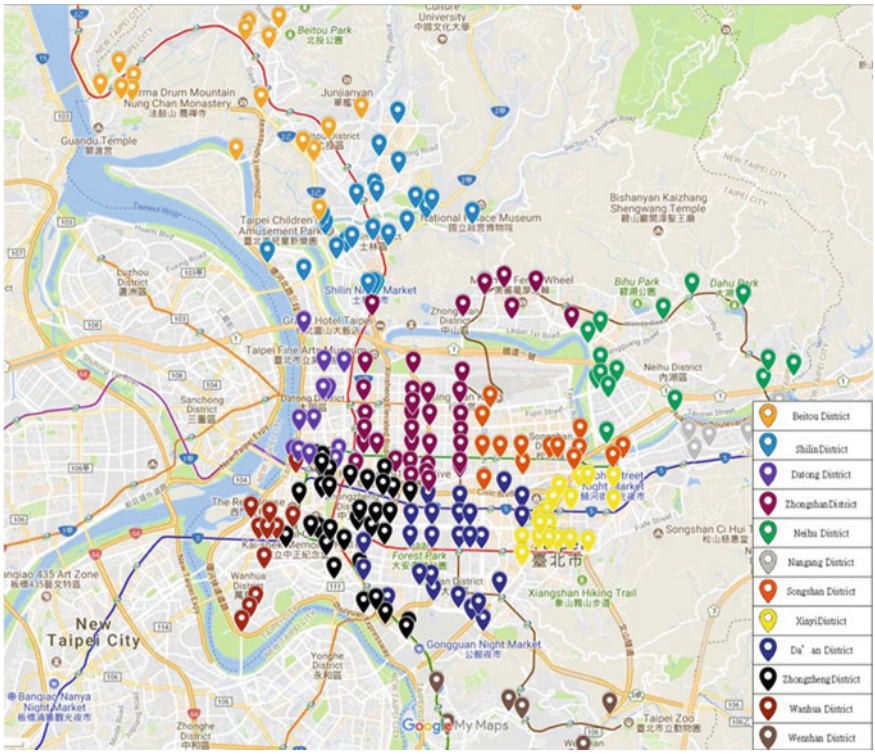


Fig. 1 Using administrative divisions for the road junction location map

(1) Normalize according to the overall road junction
The road junction are normalized under the same time interval. Therefore, it can calculate traffic flow ratio of various road junctions and the overall road junctions at the same time interval. The same way of normalization is done for each period of time. The data are normalized to set between 0 and 1. Each data type in the data set is changed to traffic flow ratio of various road junctions and the overall road junction. According to the ratio, it can be informed of the road junction at different periods of traffic growth under the value not only by considering the intersection, but also by integrating the overall road junctions for analysis. The normalization formula is as follows:

$$N_{(x,t)} = \frac{d_{[x,t]} - min(d^t)}{max(d^t) - min(d^t)} \tag{1}$$

where d is data set, x is the number of data, and t is the time period.

(2) Normalize according to the single road junction
The road junction are normalized at different time period. Therefore, it can calculate different time periods of growth value at each road junction. The data are normalized to set between 0 and 1. It is different from the first formalization.

This method is mainly to know the grow value of the road junctions. The normalization formula is as follows:

$$N'_{(x,t)} = \frac{d_{[x,t]} - min(d^x)}{max(d^x) - min(d^x)} \tag{2}$$

Select the road junction traffic flow with more than 30,000 as the basis for index classification. According to the selected intersections, the growth of traffic flow under two kind of normalization methods is calculated. Through the first normalization, it can get the growth value of the road junction relative to the overall traffic flow. Through the second normalization calculation, it can get the car traffic growth value at this road junction. The two growth values will be added as the threshold traffic classification index.

If only considering the first normalization, there is a possibility that the value of the road junction growth value will decrease, but the decline will be small. Therefore, at the ratio of the total intersections, the growth of the junction will not be clearly drawn.

On the contrary, if only considering the second normalization, we only have the growth value of the single road junction. However, for the traffic network, we must consider the overall network vision so as not to adjust the traffic efficiency of a single road junction affecting the traffic efficiency of the surrounding junctions. Therefore, this article integrated two kinds of growth values to analyze indicators so to effectively meet the current situation of traffic flow. The formula is as follows:

$$G = G_1 + G_2 \tag{3}$$

$$\sum_{n=1}^{t-1} G_1 = G_1 + \left(N_{(x,n+1)} - N_{(x,n)}\right) \tag{4}$$

$$\sum_{n=1}^{t-1} \mathrm{G}_2 = \mathrm{G}_2 + \left(N'_{(x,n+1)} - N'_{(x,n)}\right) \tag{5}$$

where G is the sum of two kinds of normalized car traffic flow growth values, G_1 and G_2 are the growth values of road junction traffic flow after the first normalization and the second normalization, respectively.

After the traffic flow growth values of selected road junctions are calculated, the road junction traffic flow index is classified, and the road junction traffic flow indicator is divided into eight levels. Level 1 is the lowest car traffic growth, and level 8 is the highest traffic growth. It means that when the index level is 8, the road junction traffic flow is very high and also the most congested intersection level. Each level corresponds to a color as shown in Table 1.

After the establishment of the road junction traffic flow indicators, it will assign the remaining road junction to these eight levels. Each road junction will only correspond to a road junction traffic flow indicator. Finally, the corresponding

Table 1 Road junction traffic flow indicator level correspondence

Traffic flow indicator	Range	Color
1	$G \leq -1.175$	Blue
2	$-1.175 < G \leq 1.510$	Green
3	$1.510 < G \leq 3.455$	Yellow
4	$3.455 < G \leq 5.365$	Red
5	$5.365 < G \leq 6.483$	Brown
6	$6.483 < G \leq 6.752$	Gray
7	$6.752 < G \leq 7.865$	Dark gray
8	$7.865 < G$	Black

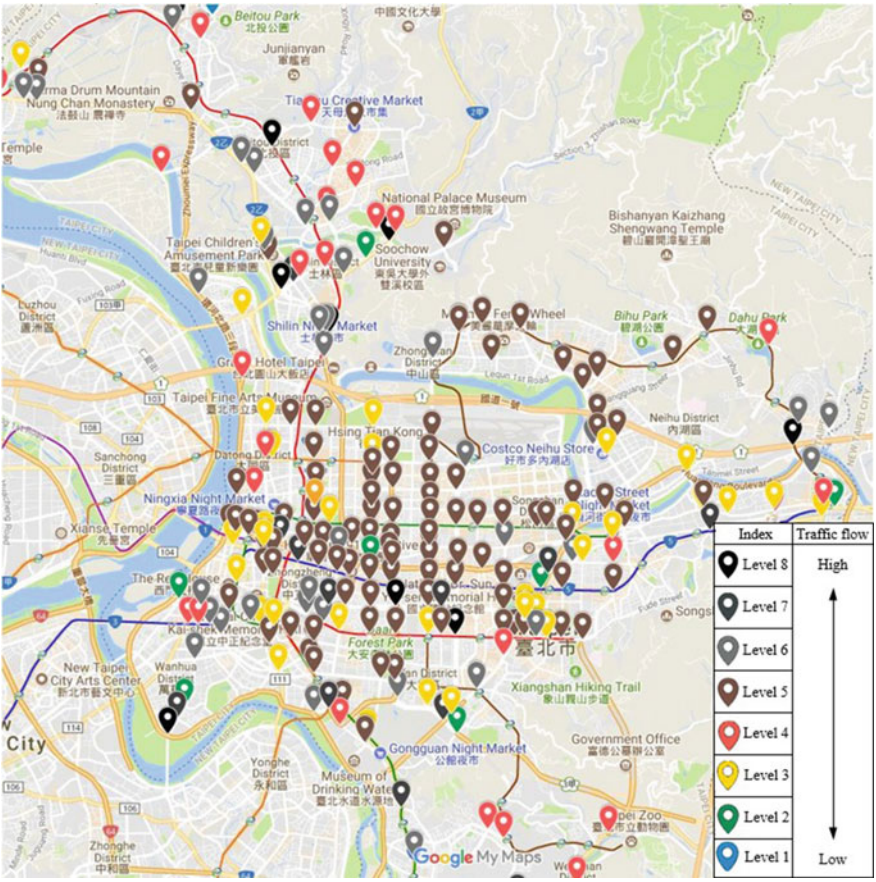


Fig. 2 Road junction traffic flow indicator location map

relationship is presented to the location map as shown in Fig. 2. Figure 2 can clearly see the relationship between high-traffic road junctions and low-traffic road junctions.

4 Road Associative Analysis

According to Fig. 2, it can obtain road junction traffic flow indicator position correspondence that can find the relationship between the road junctions. According to the traffic pattern correlation analysis, we can look for the most congested road junction in the area as center of association analysis. High traffic congestion at the road junction of the highest traffic flow will have the highest level of road junction traffic indicators.

According to traffic flow characteristics, the most congested road junction will gradually spread traffic flows to the surrounding road junctions. Therefore, the traffic flow indicators of surrounding road junctions will gradually decline. When the road junction traffic flow indicators of the road junction are down to the lower level, meaning that the traffic congestion of the most congested is gradually evacuated to the this road junction. Therefore, the next road junction has the least influence for the congestion points. According to this way as the threshold of road junction association analysis, the following uses Fig. 3 as an example.

Fig. 3 Road junction association analysis

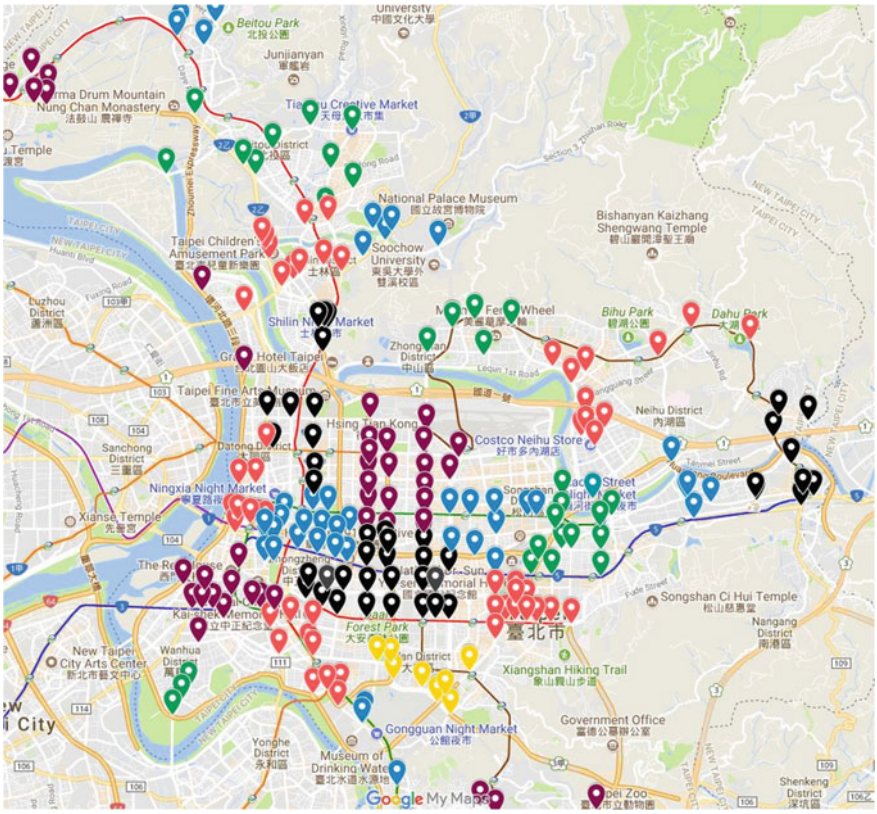


Fig. 4 Road junctions associated with the corresponding map

In Fig. 3, the red dotted line and the blue dotted line show the different relationship areas of road junctions. In the red area, first of all, select the highest intersection traffic flow indicators in the area as the associated center. Then according to traffic flow pattern, traffic flows will gradually spread to the surrounding junctions. Therefore, the road junction of the surrounding traffic flow indicators will be reduced. The red dotted line is the diffusion area of the red area. In this way, we can use the road junction traffic flow indicators to get the relationship between the road junctions. Road junctions are associated with the location of the corresponding map as shown in Fig. 4. The associated road junction shows the same color. Each area is independent, and different color repetition in different regions does not represent the connection.

According to this relationship, it can be found that traffic soothing directionality is rather multi-directional. When most of the traffic signal lights are being designed in Taiwan, only the current maximum pass rate of major roads are taken into consideration. Most of consideration are based on horizontal or vertical traffic light configuration considerations. Therefore, designing traffic on the more complex

traffic lights, we need to consider if the configuration of the traffic lights will produce too large or too small scope If the scope of consideration is too small, it will lead to a lack of ease of congestion. As a result, traffic congestion will continuously rise. If the scope of consideration is too large, it may take into account unnecessary junctions, resulting in the design flaws of the secondary junctions that cannot ease traffic congestion. The ripple effect caused by the above problems will make the entire traffic network become paralyzed during rush hours and cannot effectively handle traffic flows.

This article proposed the road junction association analysis, and the information used is the traffic flow of all important road junctions at rush hour. Through the traffic volume of each road junction, the traffic flow road junction traffic indicators are analyzed, and the results of its analysis are applied to all road junction, which can effectively find out the relationship of each road junction. This method does not produce the above-mentioned considerations of too small or too large scope. Because it is based on the status of the current traffic flow for association analysis to be able to put appropriate practical traffic conditions, not just through the experience of personnel to judge.

5 Classification Model Build

In the future, if joining other junctions to find road junction traffic flow indicators, the current data can be used to build a model for subsequent intersections to classify road traffic indicators. The model is used for classification. According to the model feature extraction method, the classification basis of the traffic flow indicators at different intersections can be obtained.

This study uses a deep neural network as an algorithm to establish a classification model of traffic flow road junction indicators. There are many places in Deep Neural Networks that are different from the original neural network. In the traditional neural network, when the network layer increases, it will be more and more prone to local optimal phenomenon. As a result, the traditional neural network can only be modeled using shallow neural networks.

Deep neural networks use the Rectified Linear Unit (ReLU) and maxout functions to overcome the problem of gradient disappearance and then add the concept of Dropout during training. Dropout can set how many neurons to discard each time, and prevent all feature selectors from interacting each time, avoiding always zooming in or out of certain features. Adjust the neurons for each training to prevent over-fitting situation from happening. Based on the above reasons, this study uses deep neural network as an algorithm to establish classification model of road junction traffic flow indicators.

The deep neural network used in this study has a total of six layers, one input layer, four hidden layers, and one output layer. The hidden layer nodes are 500, 400, 300, 200 respectively. Dropout settings for each hidden layer are 0.5, 0.4, 0.4, and 0.3. Depth neural network is set to fully connected state, using the Rectified

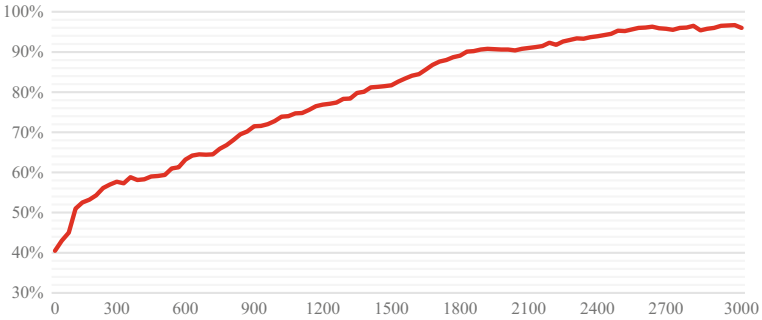


Fig. 5 The model training process

Linear Unit (ReLU) function as the activation function, with the learning rate set to 0.08, and the momentum factor set to 0.9. After 3000 epochs training, the accuracy can be as high as 95% and the loss value below 0.1. The model training process is shown in Fig. 5.

As can be seen from Fig. 5, the initial training was 40% accuracy; through continuous training when the epoch reaches the set 3000, the accuracy of the model has been as high as 95%. The model can accurately determine the classification of road junction traffic flow indicators at each road junction according to the data set used. According to the established model, in the future, when other intersection information can be used, this model can be used to classify road junction traffic flow indicators.

6 Conclusion

This article uses 230 important road junctions in Taipei City road for junction traffic flow indicators analysis. According to the proposed road junction traffic flow indicator method to analyze the road junction traffic flow, the road junction traffic flow indicator is divided into eight levels, and then according to the indicator conduct road junction association analysis. After confirming relationship of road junctions, the deep neural network is used for learning and modeling. Through this model, it can be used as a classifier for road junction traffic flow indicators with a precision of more than 95%.

The method of intersection traffic flow index proposed in this paper is based on traffic flows at each juncture during peak hours. Therefore, it can meet the current state of traffic flows to find its relevance. Compared with the traditional way to classification is personnel according to past experiences after they arrived at the scene. There are better time-cost-effectiveness and the ability to extract features that are not visible with artificial feature extraction. In the future, when configuring the traffic light cycle configuration, the cycle configuration may be performed according to the characteristics of the correlation model so as to avoid the problem of oversized or undersized considerations and find a better traffic light configuration.

Acknowledgements This study was supported by the Ministry of Science and Technology (No. MOST 105-2221-E-324-009-MY2) of Taiwan.

References

1. Anagnostopoulos T, Zaslavsky A, Kolomvatsos K, Medvedev A, Amirian P, Morley J, Hadjieftymiades S (2017) Challenges and opportunities of waste management in IoT-enabled smart cities: a survey. IEEE Trans Sustain Comput 2(3):75–289
2. Xu X, Wang W, Liu Y, Zhao X, Xu Z, Zhou H (2016) A bibliographic analysis and collaboration patterns of ieee transactions on intelligent transportation systems between 2000 and 2015. IEEE Trans Intell Transp Syst 17(8):2238–2247
3. Younis O, Moayeri N (2017) Employing cyber-physical systems: dynamic traffic light control at road intersections. IEEE Internet Things J 4(6):2286–2296
4. Contreras S, Kachroo P, Agarwal S (2015) Observability and sensor placement problem on highway segments: a traffic dynamics-based approach. IEEE Trans Intell Transp Syst 17 (3):848–858
5. Li J, Mei X, Prokhorov D, Tao D (2016) Deep neural network for structural prediction and lane detection in traffic scene. IEEE Trans Neural Netw Learn Syst 28(3):690–703
6. Martinez C, Velasquez JD (2016) An efficient new scheme of fitness evaluation in genetic programming using the R language. IEEE Latin America Trans 14(4):1866–1869
7. Witten IH, Frank E, Hall MA, Pal CJ (2016) Data mining: practical machine learning tools and techniques, 4th edn. Morgan Kaufmann, San Mateo, CA, USA
8. Chakravorti T, Patnaik RK, Dash PK (2018) Detection and classification of islanding and power quality disturbances in microgrid using hybrid signal processing and data mining techniques. IET Signal Proc 12(1):82–94
9. Lughofer E, Pratama M (2017) Online active learning in data stream regression using uncertainty sampling based on evolving generalized fuzzy models. IEEE Trans Fuzzy Syst 26 (1):292–309
10. Wang J, Zhang F, Liu F, Ma J (2016) Hybrid forecasting model-based data mining and genetic algorithm-adaptive particle swarm optimization: a case study of wind speed time series. IET Renew Power Gener 10(3):287–298
11. Wu W, Peng M (2017) A Data mining approach combining k -means clustering with bagging neural network for short-term wind power forecasting. IEEE Internet Things J 4(4):979–986
12. Sheng G, Hou H, Jiang X, Chen Y (2016) A novel association rule mining method of big data for power transformers state parameters based on probabilistic graph model. IEEE Trans Smart Grid 9(2):695–702
13. Munk M, Drlík M, Benko L, Reichel J (2017) Quantitative and qualitative evaluation of sequence patterns found by application of different educational data preprocessing techniques. IEEE Access 5:8989–9004
14. LeCun Y, Bengio Y, Hinton GE (2015) Deep learning. Nature 521:436–444
15. Ullah A, Ahmad J, Muhammad K, Sajjad M, Baik SW (2017) Action recognition in video sequences using deep bi-directional lstm with cnn features. IEEE Access 6:1155–1166
16. Gelly G, Gauvain JL (2017) Optimization of rnn-based speech activity detection. IEEE/ACM Trans Audio Speech Language Process 26(3):646–656
17. Lai YH, Chen F, Wang SS, Lu X, Tsao Y, Lee CH (2016) A deep denoising autoencoder approach to improving the intelligibility of vocoded speech in cochlear implant simulation. IEEE Trans Biomed Eng 64(7):1568–1578

A Priority-Based Fast Forward Scheduling Function in 6TiSCH Networks

Tsung-Han Lee, Ji-Liang Chen and Lin-Huang Chang

Abstract The technology of Internet of Things is gradually maturing. IoT has applied in many fields and extended to industrial networks. Thus, the IETF has proposed 6TiSCH (IPv6 over the TSCH mode of IEEE 802.15.4e) to establish the link between Internet and IWSNs (Industrial Wireless Sensor Networks). The main purpose of this research is to enhance the performance of Multi-hop 6TiSCH network by using the proposed Priority-based Fast Forward Scheduling Function (PFFSF). According to the simulation results, the proposed PFFSF has lower end-to-end transmission delay and higher throughput then the original 6TiSCH SF0 mechanism.

Keywords 6TiSCH · SF0 · 6top · IEEE 802.15.4e

1 Introduction

Low-power wireless sensor network (WSN) is widely used for Internet of Things (IoT) applications, such as smart home, smart city and Industrial internet. However, most of IoT devices are operating in the 2.4 GHz ISM frequency band and often face the co-channel interference from other wireless technologies, such as Wi-Fi and Bluetooth. Thus, IETF recently proposed 6TiSCH [1] protocol which uses the IEEE 802.15.4e [2] time slotted channel hopping (TSCH) MAC layer under the IPv6 addressing to reduce the co-channel interference from other wireless communications. In 6TiSCH, protocol dynamically assigns bandwidth resources to the

T.-H. Lee (✉) · J.-L. Chen · L.-H. Chang
Department of Computer Science, National Taichung University of Education, Taichung, Taiwan
e-mail: thlee@mail.ntcu.edu.tw

J.-L. Chen
e-mail: bcs104101@gmail.ntcu.edu.tw

L.-H. Chang
e-mail: lchang@mail.ntcu.edu.tw

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_20

nodes in the network according to the application requirements. The 6top [3] is a new sub-layer for 6TiSCH which responsible for the adjacent node monitoring and management nodes scheduling. However, the default Scheduling Function Zero (SF0) [4] in 6TiSCH uses a simple random slot selection which does not suitable for 6TiSCH multi-hop networks. Therefore, a Priority-based Fast Forward Scheduling Function (PFFSF) has been proposed in this paper to enhance the transmission performance for 6TiSCH multi-hop Networks. We also have compared the transmission performance between the original 6TiSCH SF0 and the proposed PFFSF in the OpenWSN [5]. The simulation result shows the proposed PFFSF has the lower end-to-end transmission delay and higher throughput than the original 6TISCH SF0 mechanism.

The rest of paper is organized as follows. The 6TiSCH standard and related works are reviewed in Sect. 2. The Sect. 3 describes the 6TiSCH SF0 and the proposed PFFSF. Section 4 presents the simulation results to analysis the performance of the proposed PFFSF mechanism as well as the original 6TiSCH SF0. Finally, Sect. 5 concludes this paper along with future works.

2 Related Work

This section takes a deep discussion of the related researches on 6TiSCH and related scheduling functions in 6TiSCH networks.

2.1 6TiSCH

IETF proposed 6TiSCH architecture, which integrate with IETF 6LoWPAN [6], ROLL [7] and 802.15.4e TSCH. In the 6TiSCH architecture, UDP protocol is used on upper-layer communication. The routing strategies may take control with centralized faction by PCE [8] (Path Computation Element) or a distributed routing design RPL to manage the route-path between neighbor nodes. The 6top is a new sub-layer in the 6TiSCH stack, which responsible is maintain the node schedule. A default Scheduling Function Zero (SF0) enables neighbor nodes to manage scheduled timeslots, so the receiver and the transmitter can communicate with each other.

2.2 Related Researches on 6TiSCH Scheduling Functions

The authors in [9], the On-The-Fly (OTF) [10] mechanism has been proposed to allocate the required bandwidth by added or deleted cells between the pair of nodes. A fast scheduling method called LLSF is proposed in paper [11], which reallocates

cells to enhance transmission performance in multi-hop 6TiSCH network. In LLSF, the transmitter selects a TX cell randomly to transmit a packet and forwards the received packet in the next slot immediately. The receiver in next hop, will select the closest unused cell to forward the received packet. However, we focus on the issue for how to scheduling different service requirements in the multi-hop 6TiSCH network. Therefore, a priority-based fast forward scheduling function (PFFSF) has been proposed in this paper to enhance the transmission performance for emergency packets in 6TiSCH networks. We will conduct the simulation to compare the performance between the original 6TiSCH SF0 and the proposed PFFSF.

3 The Schedule Functions in 6TiSCH

3.1 Schedule Function Zero (SF0)

In the IEEE802.15.4e TSCH, IETF proposed a 6TiSCH Operation Sublayer (6top) to monitor the transmission bandwidth between nodes. The SF0 is the default scheduling function of 6TiSCH which communicates through 6P Protocol [12]. The SF0 takes a random schedule to complete the individual schedule table called slotframe. The cell of a timeslot contains slot offset and channel offset and defined as TX, RX, Share and serial Rx states to manage the communication of node.

In Fig. 1, a child node sends a 6P request to the parent node. The cells list contains 3 cells, which is {[2, 7], [4, 9], [3, 7]}. Once, the parent node will compare the schedule table with cells list. Finally, the slot offset and channel offset (4, 9) can be set into the 6P response and sent to the child node.

The Fig. 2 shows the multi-hop forwarding route path from node 15 to node 1, and the node 7 and node 3 are forwarding nodes of the route path. The RX cell of node 3 (from node 7 to node 3) is located after the outgoing TX cell (from node 3 to node 1). Thus, node 3 must buffer the corresponding packet during the next closest TX cell in the next slotframe, thereby significantly increasing the end-to-end delay.

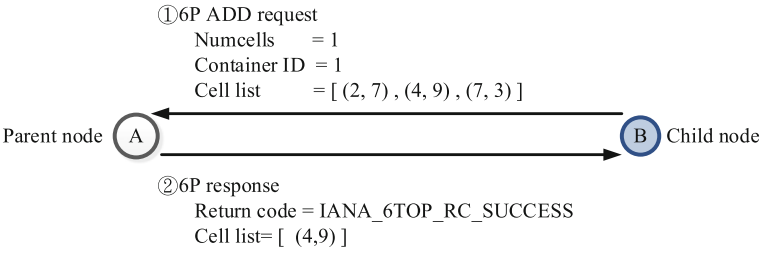


Fig. 1 6P negotiate process

Fig. 2 SF0 schedule table

3.2 *Priority-Based Fast Forward Scheduling Function (PFFSF)*

The proposed Priority-based fast forward scheduling function can be divided into the fast forward scheduling function and the priority-based queuing design.

Fast forward scheduling function (FFSF)

Figure 3 shows the flowchart for FFSF, a node receives a requested from 6P to setup a RX cell, it will check the RX + 1 cell first to confirm whether it can be used to forward packet immediately in the cells list in order to establish a fast forwarding path in the multi-hop 6TiSCH networks.

Figure 4 illustrates the FFSF schedule in a 3 hops 6TiSCH network. In the route path from node 15 to node 1, the outgoing cell in node 7 is located next the incoming cell in the second hop. Thus, node 7 can fast forwarding the corresponding packet in the same slotframe in order to reduce the end-to-end delay.

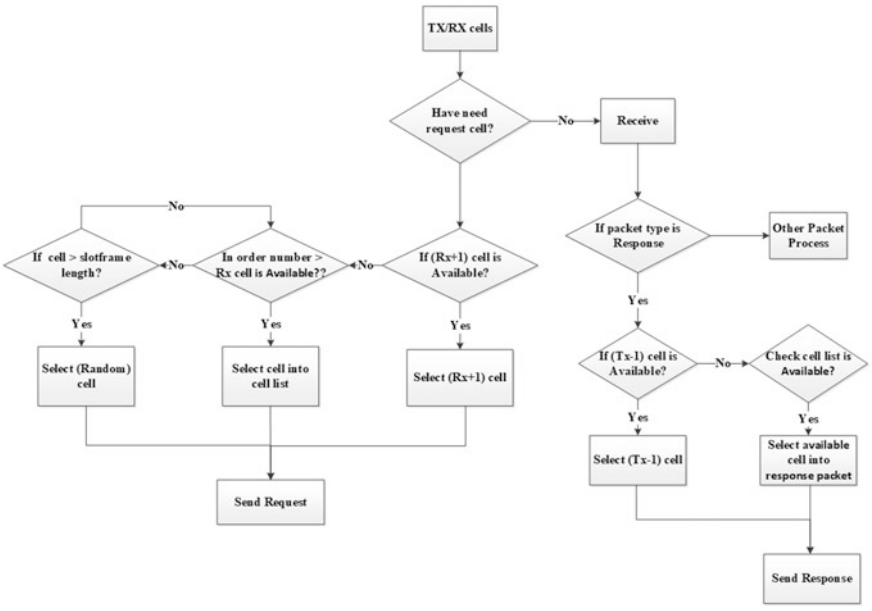


Fig. 3 PFFSF schedule process

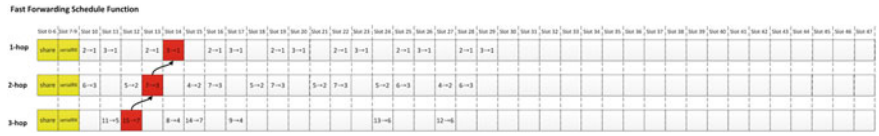


Fig. 4 FFSF schedule table

The priority-based queuing design

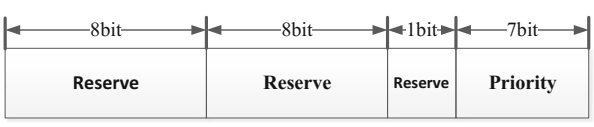
In the PFFSF, we classify packets into emergency and best-effort packets by the priority tag in flow label field. Figure 5 shows the 7 bits of priority tag in the flow label field. The value of priority tag for best-effort packets is 0×00 and the value for emergency packets is 0×01. The emergency packet will be first forward according to the priority-based queuing design once the queue has emergency packets. Otherwise, the best-effort packets will follow the FIFO queue.

The Fig. 6 shows the process of priority-based queuing design. In the emergency packet, the emergency tag will insert into the flow label field. The proposed priority-based queuing design will check is there any emergency packets in the queue when the TX cell is encountered. If so, the sending node will rearrange the emergency packet to the first position of the queue to ensure the emergency packet will be send by the closest TX cell. Thus, the forwarding node will fast forward the emergency packet in the nearest TX cell once an emergency packet has received in the RX cell.

4 Performance Analysis

4.1 Simulation Configuration

In this paper, we conduct the openWSN 6TiSCH simulator to compare the performance for SF0 and proposed PFFSF. Figure 7 shows the 3-hop full binary trees network topology in 6TiSCH network. The simulation parameters are summarized in Table 1. The packet length of MAC frame is 127 bytes and a slotframe is 47 slots (about 705 ms). Table 2 shows the pre-defined packet delivery ratio for selected channel 15, 20, 25 and 26 for our simulation.



Priority	Value
Emergency	0x01
Best effort	0x00

Fig. 5 The 7 bits of priority in the flow label field

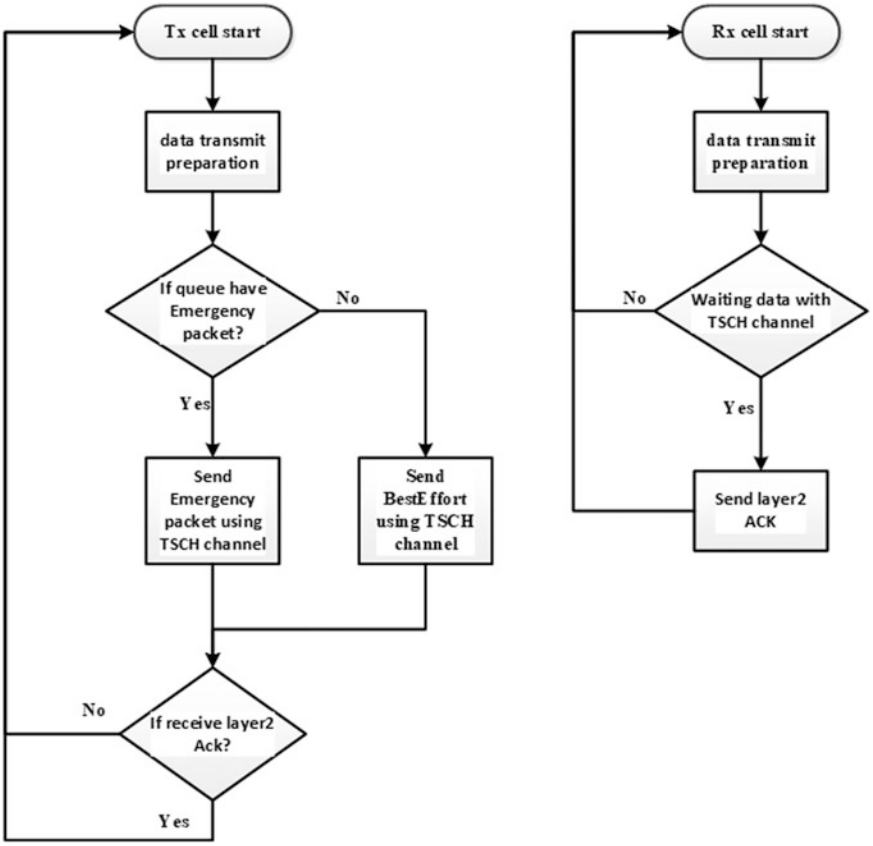


Fig. 6 The priority-based queue process

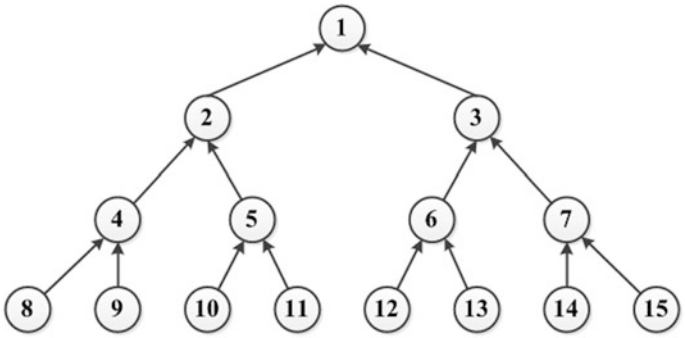


Fig. 7 The Multi-hop network topology in OpenWSN

Table 1 Simulation parameter

Selected channels	15, 20, 25, and 26
Number of nodes	15
UDP period	1000 ms
packet length	127 bytes
UDP packet length	73 bytes
The length of slotframe	47 cells
Maximum MAC retries	2

Table 2 The PDR of selected channels

Channel	PDR
15	0.63
20	0.70
25	0.92
26	0.81

4.2 Simulation Results

Figure 8 shows the average end-to-end delay for UDP packets sending from child nodes to the parent node. In PFFSF, the emergency packets are transmitted from node 8 to node 1. Form the simulation result, both PFFSF and FFSF can reduce the end-to-end delay in the 3-hop network topology and reach the destination more quickly. In addition, the end-to-end delay of node 8 in the proposed PFFSF is 0.88 s which has lowest end-to-end delay than SF0 and FFSF. It is because the emergency packets sending from node 8 can take the advantage in our proposed priority-based queuing design in PFFSF to reduce the average end-to-end delay compared with other best-effort packets.

The simulation results of the average throughput for SF0, FFSF and PFFSF is illustrated in Fig. 9. Both PFFSF and FFSF have higher average throughput then SF0. In node 8, the average throughput of SF0 is 0.83 kbps and the FFSF is 0.86 kbps. However, the average throughput of emergency packets for PFFSF with priority-based queue design is 1.15, which is the highest average throughput than other best-effort packets in 3-hop 6TiSCH network.

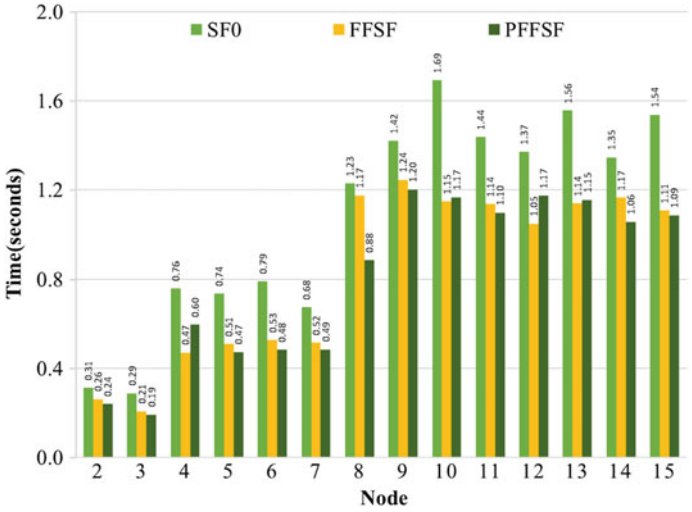


Fig. 8 The average end-to-end delay by using SF0, FFSF and PFFSF

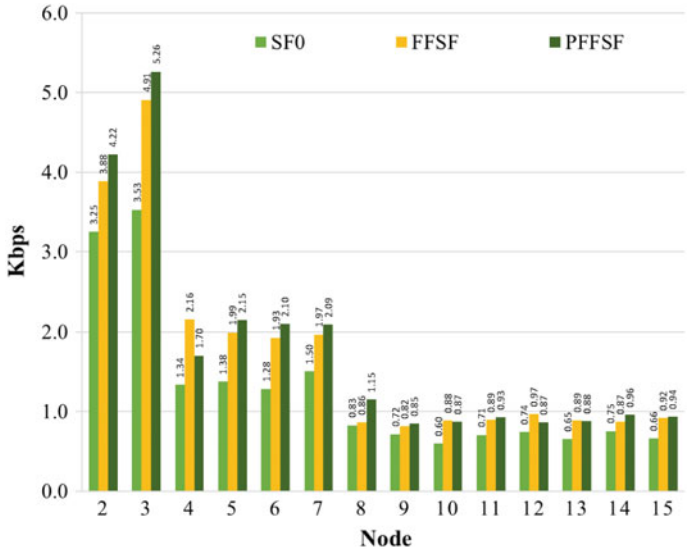


Fig. 9 The average throughput by using SF0, FFSF and PFFSF

5 Conclusions

In this paper, a Priority-based Fast Forward scheduling Function based on 6TiSCH networks is proposed. We have compared the original SF0, FFSF and PFFSF for end-to-end delay and average throughput in OpenWSN. From simulation results, the proposed PFFSF has high transmission performance than the original 6TiSCH SF0.

In future work, we consider completely integrated the QoS of IOT devices into the 6TiSCH. Thus, how to using the QoS-based schedule functions in 6TiSCH to achieve a bidirectional QoS mechanism between IWSNs and the Internet is the goal of further research.

Acknowledgements This research was supported by research grants from Ministry of Science and Technology, Taiwan 105-2221-E-142-002-MY2.

References

1. Thubert P (2016) An architecture for IPv6 over the TSCH mode of IEEE 802.15.4, draft-ietf-6tisch-architecture-10(work in progress)
2. IEEE standard for information technology, IEEE std. 802.15.4e, Part. 15.4: low-rate wireless personal area networks (LR-WPANs) Amendment 1: MAC sublayer. April 2012
3. Wang Q, Ed X (2015) Vilajosana. 6TiSCH operation sublayer (6top) interface
4. Dujovne D, Grieco L, Palattella M, Accettura N (2015) 6TiSCH 6top scheduling function zero (SF0), draft-ietf-6tisch-6top-sf0-02 (work in progress)
5. OpenWSN. https://openwsn.atlassian.net/wiki/spaces/OW/pages/29884474/Home. Last Accessed 26 Feb 2018
6. Kushalnagar N et al (2007) IPv6 over low-power wireless personal area networks (6LoWPANs): overview, assumptions, problem statement, and goals. IETF RFC 4919
7. Vasseur J (2010) Terminology in low power and lossy networks, draft-ietf-roll-terminology-04 (work in progress)
8. Vasseur J, Le Roux J (2009) Path computation element (PCE) Communication protocol (PCEP), RFC 5440, IETF
9. Palattella MR (ed0 On-the-fly bandwidth reservation for 6TiSCH wireless industrial networks. IEEE Sensors J
10. Dujovne D (ed) 6TiSCH on-the-fly scheduling, IETF Std. draft-dujovne-6tisch-on-the-fly-06
11. Chang T (ed) (2016) LLSF: low latency scheduling function for 6TiSCH networks. In: international conference on distributed computing in sensor systems, May 2016
12. Wang Q, Vilajosana X (2016) 6top Protocol (6P), draft-ietf-6tisch-6top-protocol-03, October 2016

An Adaptive Bluetooth Low Energy Positioning System with Distance Measurement Compensation

Hung-Chi Chu and Ming-Fu Chien

Abstract In recent years, smart phones are widely used in indoor positioning services, making the indoor positioning accuracy increasingly important. However, most of existing positioning technologies have high errors or require some extra special hardware. This study proposes an algorithm to change the environmental attenuation factor and compensation mechanism according to the current situation of the environment, and reduce the signal anomaly caused by the decrease of positioning accuracy. To ensure the positioning system has a certain positioning accuracy to achieve system efficiency, the method consists of three stages: First of all, calculate the current environmental attenuation factor by distance. Secondly, compensate the lack of signal distance according to the compensation mechanism. Finally, the triangulation method and the revised signal are used to locate. The experimental results show that the average positioning error of the proposed method is less than 0.5 m, which is better than some of the existing positioning methods, and provides accuracy and practicality to the positioning system.

Keywords Indoor positioning · Attenuation factor · Compensation mechanism

1 Introduction

In recent years, people who own smart phones worldwide have been growing rapidly, and wireless network technology has become popular. People live more and more time indoors, so the indoor environment of location-based services has gradually been taken seriously. Because people spend most of their time indoors like shopping malls, exhibitions, airports, hospitals, MRT stations, college cam-

H.-C. Chu (✉) · M.-F. Chien
Chaoyang University of Technology, Taichung Wufeng District 41349, Taiwan, R.O.C.
e-mail: hcchu@cyut.edu.tw

M.-F. Chien
e-mail: S10530606@gm.cyut.edu.tw

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_21

puses, etc., it is very important to provide high accuracy positioning in indoor environment for many people and commercial applications.

Since wireless network technology has the advantages of high mobility, high elasticity, high transmission rate, etc., it can only be applied to a small range of homes and offices, but also extended to a wide range of intelligent buildings, shopping malls and school campuses and other environments. Wireless network technology can sense things in the environment through some components to sense temperature, humidity, carbon dioxide, etc., so it can be used in industry for processes and equipment monitoring, and commercial applications such as air-conditioning temperature control, detection of building fire or guiding the escape path. In these multitudinous applications, the indoor positioning is one of relatively important applications. The equipment such as ZigBee, Bluetooth, wireless access point and other devices can be used to complete the indoor environment positioning system.

The well-established technology in the positioning system is that Global Positioning System (GPS) [1] can provide convenient and reliable space location services. However, in the indoor environment, satellite signals received by GPS will be interfered and affected by the shielding, making it impossible to effectively operate satellite signals so that in the indoor environment GPS positioning accuracy is not enough to successfully complete the positioning of the situation. When the indoor environment can't use GPS technology to complete the positioning, it needs other positioning technology to complete the indoor positioning service. Common indoor location technologies include infrared [2], ultrasonic [3], radio frequency identification (RFID) [4], Bluetooth [5, 6], Wi-Fi [7, 8], ZigBee [9]. However, these technologies have some bottlenecks like positioning time, accuracy and calculation complexity. Therefore, the research of relative indoor positioning technology is still very important.

In this paper, we propose a novel positioning system to improve the positioning accuracy. In indoor environment, we use a smart phone to receive the signal strength indicator from the beacon [10] and use the algorithm we proposed to consider the environment attenuation factor and the signal distance compensation so that the wireless signal quality in the environment can change effectively to adjust the positioning of the parameter values to ensure the positioning system has a certain positioning accuracy and system efficiency. The remainder of this paper is organized as follows. In the Sect. 2 related work is briefly reviewed. The environmental attenuation factor selection algorithm is discussed in detail in Sect. 3. Section 4 shows the experimental results in a real-world scenario. Finally, the conclusion of this paper is given in Sect. 5.

2 Related Works

Wi-Fi (Wireless Fidelity) positioning technology [7, 8] development direction can be broadly divided into two types: One is the triangulation method through which a mobile device obtains three or more AP signal intensities, targets and distances or

angles of AP. The features of the triangulation geometry method is used to compute positioning information such as time of arrival (TOA) [11], time difference of arrival (TDOA) [12], angle of arrival (AOA) [11] and received signal strength indicator (RSSI), etc. The other is the fingerprint positioning method. The fingerprint positioning method must in advance measure the signal characteristics (Wi-Fi signal strength in each direction) in the environment everywhere and save fingerprints in database. The positioning of the current received signal characteristics is matched with the fingerprint positioning database to determine the location information. Wi-Fi positioning has advantages such as fast speed and low hardware costs, and can reduce the possibility of radio frequency (RF) interference. Wi-Fi positioning technology can be used in a wide range of applications to achieve complex large-scale positioning, monitoring and tracking tasks. For Wi-Fi positioning shortcomings, the accuracy of indoor positioning can only reach about 2 m without achieving accurate positioning.

Bluetooth positioning technology [5, 6] is a short-distance and low-power wireless transmission technology, which is done by setting up the appropriate Bluetooth LAN device in the environment, configuring the network as a base network connection mode for multiple users, and ensuring that the Bluetooth local area network device is always the master of this piconet, so it can be connected through multiple Bluetooth devices at the same time, receiving the signal strength of the Bluetooth device to estimate the positioning information of the target. Bluetooth indoor positioning technology has advantages such as low power consumption, low cost, high integration and easy integration of applications in mobile devices and consumer electronics, and it is also relatively unaffected by visual distance. However, the disadvantage is that in the complex space environment, the stability of Bluetooth system is inferior and easy to be disturbed by noises. Bluetooth technology has extended the technology for iBeacon [13, 14]. IBeacon is a new technology Apple introduced in 2013. It is based on Bluetooth 4.0 low-energy communication technology, also known as Bluetooth 4.0 or Bluetooth Smart. IBeacon operation uses iBeacon signal strength as the distance classification divided into three categories. When the calculated distance ranges from 0 to 0.5 m, it is classified as “immediate”; when ranging from 0.5 to 2 m, it is classified as “near”; when ranging from the distance 2 to 30 m, it is classified as “far”; otherwise recorded as unknown. The iBeacon technology is to send the broadcast packet periodically through the mobile phone receiving iBeacon. The broadcast packet contains Universally Unique IDentifier (UUID), major, and minor identifiers as the basis for identification. When the device (smart phone) enters the area of the iBeacon signal coverage, the corresponding software receives the signal strength, and use the signal strength to obtain the distance between the device and iBeacon. Finally, using the positioning algorithm to estimate the final position of the device information. In this paper, we use the Esimote Company’s Beacon and Apple’s proposed iBeacon function of the same device to achieve the positioning system of this paper. IBeacon has advantages of Bluetooth interference and low power characteristics. It can also be deployed in large real-world applications to optimize existing Wi-Fi positioning systems to provide higher positioning accuracy.

3 Compensation Adaptive Ble Positioning Method

In our proposed positioning method, there are three main steps to obtain the environmental attenuation factor n, correction signal error, the implementation of positioning algorithm. When positioning needs RPN (Receive Position Node, RPN), each user carried by the mobile device can be used in indoor environment to receive the position information and signal strength sent out by beacon issues a positioning request and will receive the beacon node's signal strength and beacon location information to perform the three steps of the positioning method. (1) Obtain the environmental attenuation factor n: Choose the environment factor suitable for the current environment. (2) Correct the signal distance compensation: the corresponding distance is compensated for the attenuation signal when the signal is received. (3) Implement positioning algorithm: Use the circular intersection method in triangulation algorithm to estimate the position information.

In this chapter, we will introduce the environmental attenuation factor change selection and signal distance compensation mechanism in our proposed algorithm of environmental attenuation factor compensation. This algorithm is used to observe the environment change in the indoor field. According to our environmental attenuation factor compensation, when the signal is unstable or inferior, it can provide different environment attenuation factors to improve the positioning accuracy. The algorithm consists of the following three steps. (1) Collect signal strength from different distances to do signal distance correction. (2) Use standard deviation for signal filtering. (3) Compensate signal distances.

(1) Collect signal strength from different distances to do signal distance correction: The signal range correction is shown in Fig. 1. Before selecting the environmental attenuation factor, a single BN(Bluetooth Node, BN):The node is set up in the area according to demand, its main BN regularly sends broadcast packets, including their own positioning information and signal strength. The nodes are used as calibration data to measure the signal strength at three fixed distances respectively (i.e. d_1 = 0.5 m, d_2 = 1.5 m, d_3 = 2.5 m). The sampling period for each range signal is 300 ms and 10 signal strength data are sampled. The signal strengths measured at each distance is stored in the database.
(2) Use standard deviation for signal filtering: The signal strength measured at various distances are sorted first according to the signal strength, and signals strength after sorting is taken as a set. The median value and the standard deviation Formula (1) are used to calculate the signal strength of the sorted signals, and the median and standard deviation are used to calculate its threshold value for the signal filtering.

$$S = \sqrt{\frac{1}{k-1}\sum_{i=1}^{k}(X_i - \bar{X})^2} \tag{1}$$

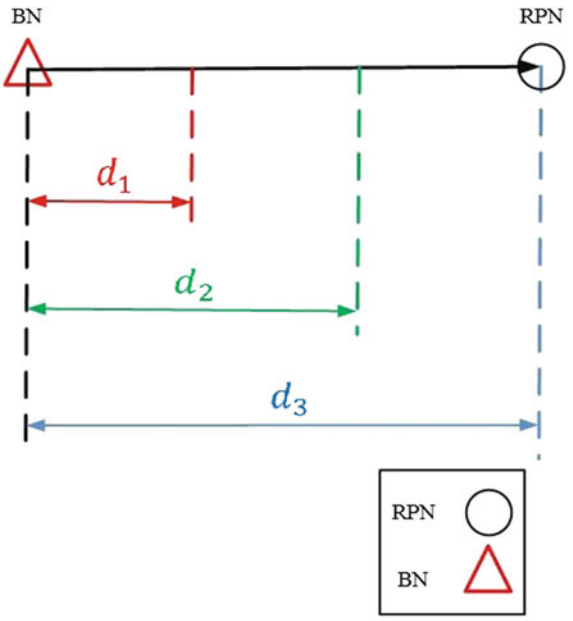


Fig. 1 Distance signal correction

S is the standard deviation; k is the number of received signal strength; X_i is the signal strength value of the k; $\overline{X}$ is the average of the total signal strength.

The measured signal strength filtered signals are averaged into the signal strength conversion distance by Formula (3). Use formula (2) [15] for derivation to complete Formula (3). The corresponding distance of all environment attenuation factor n is calculated respectively at 0.5, 1.5, 2.5 m. The error between the measured filter signal and the environment attenuation factor n of all range values is calculated to obtain errors between the estimate distance and the actual distance measured, 0.5, 1.5 and 2.5 m. The sum of the error values calculated by the rule are added to the total to find the minimum errors from the sum of the range error values for all environmental attenuation factors n, which is the best environment attenuation factor n in the current environment.

$$RSSI = -10n \log(d) + A \tag{2}$$

$$d = 10^{\left(\frac{A-RSSI}{10\times n}\right)} \tag{3}$$

$RSSI$ is the received signal strength; n is the environmental attenuation factor (generally set in free space environment between 2 and 5); A is the RSSI received at the signal transmission distance of 1 m; d is the distance between sending and receiving points.

(3) Compensate signal distance: The environment attenuation factor n and the signal range are sorted before the signal distance is compensated. The first classification is based on the currently calculated environmental attenuation factor n. The second classification is based on the signal intensity received by the measurement. Finally, the corresponding signal distance is compensated by classification twice.

Selecting the current environment attenuation factor n method includes the following main steps.

(1) In the environment, three different distances are selected to measure the signal strength of each distance for a single BN node. The measured signal strength is a set $S_{d_j}(t)$. $S_{d_j}(t)$ sequenced by signal strength is defined as S_j. d_j is the distance of the j measurement.
(2) Signal filtering is processed after the signal strength is sorted. Use Formula (4) and (1) to set different signal filtering thresholds, k is the number of sampling signals.

$$Med_j = \begin{cases} \frac{S_j\left(\frac{k}{2}\right) + S_j\left(\frac{k}{2}+1\right)}{2}, & if\ k\ is\ even \\ S_j\left(\frac{k+1}{2}\right), & else \end{cases} \tag{4}$$

(3) Using step (3), the filter signal gets the set of signal intensity after filtering S'_j. The filtered signal strength is brought into the Formula (6) to calculate the average value. Std_j is the standard deviation for calculating the received signal strength of the array.

$$S'_j = \{S_j | Med_{dj} - Std_j \le S_j \le Med_{dj} + Std_j\} \tag{5}$$

$$Avg_j = \frac{\sum S'_j}{\#ofS'_j} \tag{6}$$

(4) Calculates the distance (d'_j(n) corresponding to the attenuation factor n in different environments.

$$d'_j(\mathrm{n}) = 10^{\left(\frac{A - \mathrm{Avg}\left(S'_{dj}\right)}{10 \times n}\right)}, \mathrm{n} = 2.0 \sim 5.0,\ \mathrm{A} = -80 \tag{7}$$

(5) Calculate the error between the measured distance and the estimate distance, sum all error values, and select the minimum error value in the range of n values, environmental attenuation factor for the current environment.

The signal compensation is according to using environmental attenuation factor and receiving signal intensity. The definition of the difference before and after the compensation distance is shown in Fig. 2, where d_A, d_B, d_C are, respectively, the distance that the RPN measures the signal strength of three BN nodes. The coverage of the three BN nodes is represented by dashed circle. Figure 2 can see that the BN_C node signal range does not cover the RPN node, as a result, three BN nodes can't be successfully positioned. Therefore, the signal distance compensation mechanism must be used to solve the problem that the signal coverage to does not cover the RPN node. After using the signal distance compensation mechanism to update the original d_C into d'_C. The compensated signal coverage is indicated by a solid circle. The three BN nodes measure the signal conversion distances that are covered to the RPN node, improving the accuracy of positioning. In the actual positioning system, we do not know which BN node measurement signal is converted to a distance that does not cover the RPN node. Before the signal distance compensation, it must know that the BN node's circle has no intersection point. The BN node without the intersection gives the signal distance compensation. Preventing all BN nodes from giving signal distance compensation leads to the decrease of positioning accuracy. In the following, the compensation types of several compensation mechanisms are described.

Type 1: Three signals have only two circles intersecting. As shown in Fig. 3, BN_A and BN_B node signal ranges intersect. The BN_C node does not intersect with the signal range of any node, compensation mechanism compensating the distance of the BN_C node makes three nodes intersect.

Type 2: Three signals each intersect with different circles. As shown in Fig. 4, BN_A, BN_B and BN_C node signal ranges intersect. The BN_B and BN_C node signal range does not intersect, compensation mechanism compensating the distance of the BN_B and BN_C nodes makes three nodes intersect.

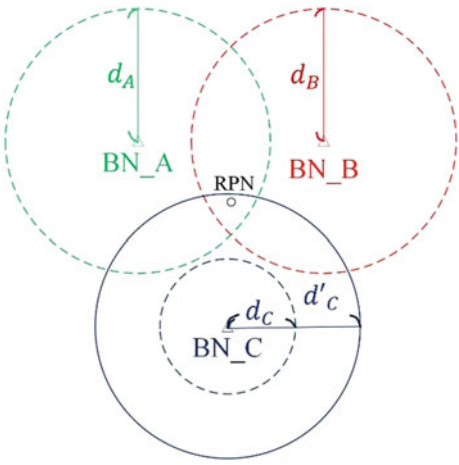


Fig. 2 Signal distance compensation

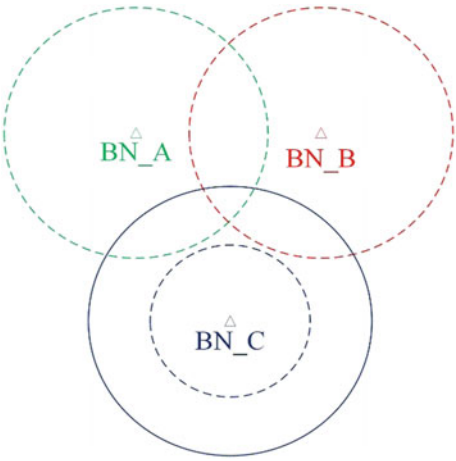


Fig. 3 Three signals have only two circles intersecting

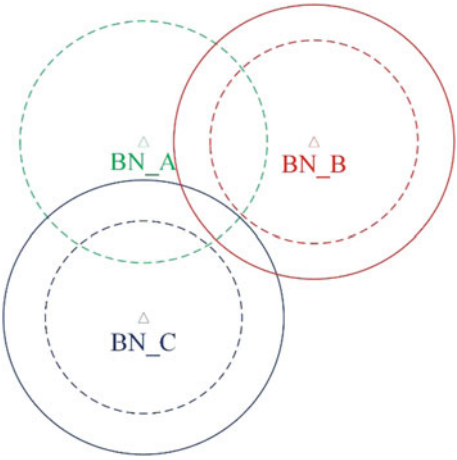


Fig. 4 Three signals each intersect with different circles

Type 3: There is no intersection of three signal ranges. As shown in Fig. 5, BN_A, BN_B and BN_C node signal ranges do not intersect. The compensation mechanism compensating for the smaller signal ranges, BN_B and BN_C nodes for the distance makes three nodes intersect.

4 Experimental Environment and Results

4.1 Experimental Environment

In order to test the performance of the proposed environmental factor compensation method, an experiment in an indoor environment of 8.5 m × 4.8 m in the field as

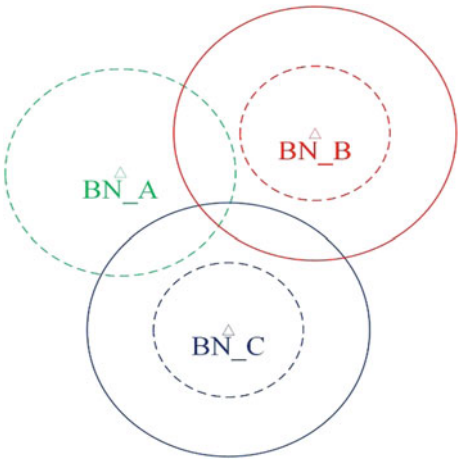


Fig. 5 There is no intersection of three signal ranges

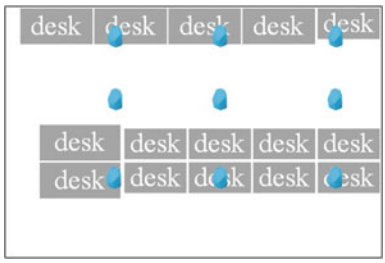


Fig. 6 Laboratory Map of Test bed

shown in Fig. 6. In the indoor environment, 9 BNs are pre-installed and an RPN is used to receive the neighboring beacon node 10 times signal strength to complete the positioning.

4.2 Experimental Results

We randomly selected 10 locations in the experimental area to evaluate the effectiveness of the positioning system. The experimental results show the error distance between the estimate position and the actual position. Compared with the triangulation method, by adding the algorithm of environmental attenuation factor and the proposed positioning method, the differences of positioning accuracy are shown in Fig. 7. In Fig. 7, the red strip only uses the environment attenuation factor algorithm, and position result is more accurate than the traditional triangulation method. The average error of the positioning experiment is about 1 m. In Fig. 7, the blue strip shows the positioning accuracy of the algorithm with ambient attenuation factor and the signal distance compensation mechanism. The results show that the

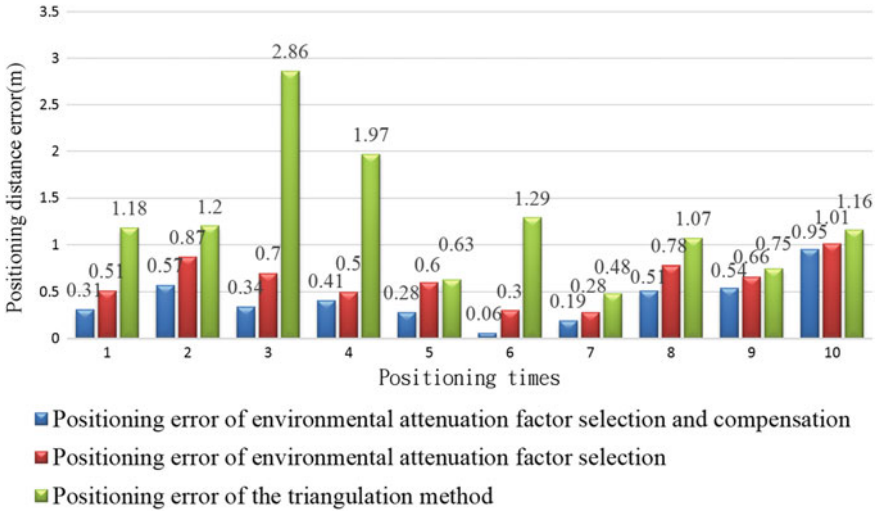


Fig. 7 Comparison of positioning errors

positioning accuracy is further improved, and the average error of the positioning experiment was about 0.5 m. The experimental results show that the environmental attenuation factor has a great influence on the overall indoor positioning accuracy. Therefore, choosing a suitable environment attenuation factor for the environment can effectively improve the positioning accuracy.

4.3 Performance Comparison of Different Algorithms

The experimental results from the localization method proposed by us show that the location error is within 0.5 m on average, compared with the traditional average signal positioning error of 2–3 m; or based on Wi-Fi positioning [16, 17] and Bluetooth positioning [5, 6, 18],the average error of using fingerprint positioning method [19–21] is 1–2 m, and the environmental attenuation factor will be changed according to the environment to provide the positioning performance that best suits the current environment. Therefore, the proposed positioning method is far more accurate than other algorithms.

5 Conclusion

By introducing the environment attenuation factor and distance compensation mechanism, the environmental attenuation factor can be instantly changed according to the current environment when the signal strength is converted into

distance, and the signal distance compensation effectively compensates for the short signal distance caused by unstable signals. Therefore, our proposed method of ambient attenuation factor compensation can provide better positioning accuracy. The experimental results show that the proposed algorithm is better than the traditional positioning algorithm, and the positioning error is smaller than other positioning methods. In the future, this method will be also applied to various scenarios for more experiments to ensure that through this method it can enhance the accuracy of positioning in any environment.

Acknowledgements This study was supported by the Ministry of Science and Technology (No. MOST 105-2221-E-324-009-MY2) of Taiwan.

References

1. Kaiser SA, Christianson AJ, Narayanan RM (2017) Global positioning system processing methods for GPS passive coherent location. IET Radar Sonar Navigation 11(9):1406–1416
2. Santo H, Maekawa T, Matsushita Y (2017) Device-free and privacy preserving indoor positioning using infrared retro-reflection Imaging. In: IEEE international conference on pervasive computing and communications (PerCom), Kona, HI, USA
3. De Angelis A, Moschitta A, Carbone P, Calderini M, Neri S, Borgna R, Peppucci M (2015) Design and characterization of a portable ultrasonic indoor 3-D positioning system. IEEE Trans Instrum Meas 64(10):2616–2625
4. Hybrid WLAN-RFID Indoor Localization Solution Utilizing Textile Tag (2015) Masoumeh Hasani, Jukka Talvitie, Lauri Sydänheimo, Elena-Simona Lohan, Leena Ukkonen. IEEE Antennas Wirel Propag Lett 14:1358–1361
5. Ma Z, Poslad S, Bigham J, Zhang X, Men L (2017) A BLE RSSI ranking based indoor positioning system for generic smartphones. In: Wireless telecommunications symposium (WTS), Chicago, IL, USA
6. Kanzaki R, Fujita S (2016) Indoor positioning based on the iBeacon framework with Gaussian weight functions. In: IEEE international conference on computational science and engineering (CSE) and IEEE international conference on embedded and ubiquitous computing (EUC) and 15th International symposium on distributed computing and applications for business engineering (DCABES), Paris, France
7. Yang C, Shao HR (2015) WiFi-based indoor positioning. IEEE Commun Mag 53(3):150–157
8. Deepika K, Usha J (2017) Design & development of location identification using RFID with WiFi positioning systems. In: Ninth international conference on ubiquitous and future networks (ICUFN), Milan, Italy
9. Alvarez Y, Las HF (2016) ZigBee-based sensor network for indoor location and tracking applications. IEEE Latin America Trans 14(7):3208–3217
10. Bluetooth SIG, https://www.bluetooth.com/markets/smart-industry. Last Accessed 27 Jan 2017
11. Hanssens B, Plets D, Tanghe E, Oestgesy C, Gaillotz DP, Li´enardz M, Martens L, Joseph W (2016) An indoor localization technique based on ultra-wideband AoD/AoA/ToA estimation. In: IEEE international symposium on antennas and propagation (APSURSI), Fajardo, Puerto Rico, pp 1445–1446
12. Keunecke K, Scholl G (2014) IEEE 802.11 n-Based TDOA performance evaluation in an indoor multipath environment. In: 8th European conference on antennas and propagation (EuCAP), Hague, Netherlands, pp 2131–2135

13. Burzacca P, Mircoli M, Mitolo S, Polzonetti A (2014) "iBeacon" technology that will make possible internet of things. In: International conference on software intelligence technologies and applications & international conference on frontiers of internet of things, Hsinchu, Taiwan, pp 159–165
14. Varsamou M, Antonakopoulos T (2014) A bluetooth smart analyzer in iBeacon networks. In: IEEE fourth international conference on consumer electronics, Berlin (ICCE-Berlin), Berlin, Germany, pp 288–292
15. Chen X, Zou S (2017) Improved Wi-Fi indoor positioning based on particle swarm optimization. IEEE Sens J 17(21):7143–7148
16. Kharidia SA, Ye Q, Sampalli S, Cheng J, Du H, Wang L (2014) HILL: a hybrid indoor localization scheme. In: 10th international conference on mobile ad-hoc and sensor networks (MSN), Maui, HI, USA, pp 201–206
17. Zhao F, Luo H, Geng H, Sun Q (2014) An RSSI gradient-based AP localization algorithm. China Commun 11(2):100–108
18. He W, Ho PH, Tapolcai J (2017) Beacon Deployment for Unambiguous Positioning. IEEE Internet Things J 4(5):1370–1379
19. Zhang Y, Zhang S, Li R, Guo D, Wei1 Y, Sun Y (2017) WiFi Fingerprint positioning based on clustering in mobile crowdsourcing system. In: 12th international conference on computer science and education (ICCSE), Houston, TX, USA, pp 252–256
20. Khalajmehrabadi A, Gatsis N, Akopian D (2017) WLAN fingerprinting indoor positioning methods and deployment challenges. IEEE Commun Surveys Tutorials 19(3):1974–2002
21. He S, Chan SHG (2016) Wi-Fi fingerprint-based indoor positioning: Recent advances and comparisons. IEEE Communications Surveys & Tutorials 18(1):466–490

Deblurring Image Using Motion Sensor and SOM Neural Network

Chu-Hui Lee and Yong-Jin Zhuo

Abstract As multimedia image related devices are widely used by the general public, multimedia image processing technology is more and more advanced, however there are still some problems that are worth to be explored and improved. How to deblurring an image without the information of speed and direction of moving objects is still a well-known ill-posed problem. In this paper, we proposed a system to deblurring image that can estimate important parameter advance to reduce the complexity of deblurring process. The data of sensor of moving object is collected. The SOM neural network is used to train to classify the speed and direction of the object from the sensor data. After that, we can estimate the speed and direction of objects without other algorithms. With such important parameters, deblurring processing will more efficient.

Keywords Motion deblurring · SOM neural network · Notion sensor

1 Introduction

In recent years, the science and technology of the age we live in have been changing with each passing day and have a certain level. The technologies available in the field of multimedia image processing are not comparable to those in the past. However, there are still some problems need to be improved. Currently, multimedia-related devices are widely used, but only a few experts actually have the skills and knowledge related to multimedia imaging. People like to keep important moments or scenery shots as their own important memories or to share with friends and family. However, if the objects are captured at high speed or the

C.-H. Lee (✉) · Y.-J. Zhuo
Department of Information Management, Chaoyang University of Technology, Wufong Township, Taipei, Taichung County, Taiwan, Republic of China
e-mail: chlee@cyut.edu.tw

Y.-J. Zhuo
e-mail: s10614603@cyut.edu.tw

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_22

photographer is moving at a high speed, the images taken are easy to blur. This type of blurring is called dynamic blurring [1]. In order to get close to the images taken by people's mean, many scholars continue to pay attention to these issues and strive to improve them.

In today's society, because of the continuous innovation and development of multimedia imaging technology for driving recorders, sports cameras, monitors, and smart phones, the photographic functions are the users' most concern. For example, when police are tracking criminals, the monitor captured the criminal tools or prisoner's face, but the images were not clear and blurry. Therefore, many scholars put forward many ways to improve this issue. Most of the concepts are to try to find out the blur kernel, and then use it to deblurring the image.

With the image blurring problem mentioned above, this paper focuses on the image blurring caused by the speed relationship when the image is captured, because it is in a high-speed moving environment or when the photographer is moving at a high speed. Therefore, the speed and the direction of object movement are two important factors for the restoration of the blurred image. They are the length L and the angle θ [2]. When both factors are known, the blurred image can be restored. The purpose of this paper is to identify the direction and speed of object movement through the data collected by the motion sensor. The 9-axis acceleration obtained by the motion sensor cannot be directly used to get the required parameters for deblurring the image. There are noise and drift problems when using the sensor data always. Generally, the Kalman filter [3] is a Markov processing most frequently used to deal with such problems. Moreover, to get the length L and the angle θ parameters in the deblurring processing from sensor data still a tricky and complicated problem. In this paper, we will use neural network learning to determine the type of object movement speed and angle direction to help to quickly deblurring images.

The structure of this paper is as follows. The first section introduces the research background, research motivation and research purpose. The second section explains and introduces the technologies used in this paper. The third section proposes the research framework for this research. The fourth section is the result of the experiments. The fifth section is the conclusion of the research and future research direction.

2 Literature Review

Nowadays, with the rapid development of multimedia image processing technology, many novel multimedia image processing technologies have been published. There are already gaps in the technology proposed by past scholars. Some technologies focus on the quality of restored image, while others focus on the time complexity of restored image. This type of multimedia image processing is to solve people's trouble shooting images, but also bring unlimited business opportunities. The related previous work will introduce in the following.

2.1 Self-organizing Maps, SOM

Proposed by Kohonen [4] scholars in 1980 and self-organizing maps is an unsupervised learning network in neural networks. It consists of an input layer and an output layer. The SOM network can map any dimension of the input layer in the output layer into a one-dimensional or two-dimensional even low-dimensional extension of the topology map.

The SOM algorithm is described as following steps [5]:

Step 1: Input vector X(t) to all neurons.
Step 2: Initialize the weight vectors W_i.
Step 3: Find the winner neuron C, BMU (Best matching unit) using the following equation:C = min $\{\|X_i - W_i\|\}$.
Step 4: The weight vector is updated using the following equation: $W_i(t+1) = W_i(t) + h_{c,i}(t)[X(t) - W_i(t)]$, where $h_{c,i}(t)$ is a Gaussian neighborhood function.
Step 5: t = t + 1, if t > T^* then stop (T^* is biggest learning echo) else back to step 3 re-start training.

2.2 Point Spread Function, PSF

The point spread function is a function that describes the optical system's ability to resolve point sources. Because the point source will be diffracted after passing through any optical system to form an enlarged image point, the point spread function of the measurement system can more accurately extract the image information [6]. As shown in Fig. 1, since the PSF is a vertically-fuzzy core, the resulting image also appears to be vertically blurred. It follows that different PSFs produce different results, but the current PSF only considers linear blurring.

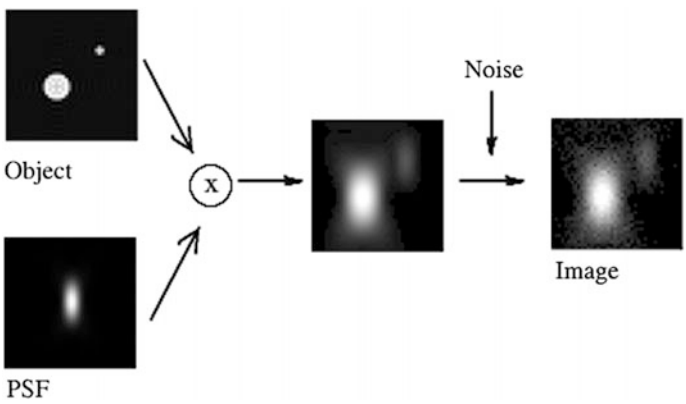


Fig. 1 The blur image model

The estimated PSF has two important parameters, namely angle θ and length L. Figure 2a shows the Fourier transform of the linear motion blurred image. To estimate the angle parameter, we need to find the point A and B in Fig. 2a. They are the nearest points to zero on each axis. Figure 2b shows the relationship of angle parameter with point A and point B. The angle calculation formula is shown in (1) and the length calculation formula is shown in (2). N denotes as image size [7].

$$\theta = \tan^{-1} \frac{\overline{OB}}{\overline{OA}} \tag{1}$$

$$\mathrm{L} = \frac{N}{d}, \mathrm{d} = \overline{OB} \cos\theta \tag{2}$$

2.3 *Motion Blur Model*

The image blur model is defined as follows [8]:

$$g(x,y) = h(x,y) * f(x,y) + n(x,y) \tag{3}$$

where "*"is the convolution operation, $g(x,y)$ is the blur image, $n(x,y)$ denotes noise, and $f(x,y)$ is the original image, $h(x,y)$ is the unknown blur kernel that is called point spread function(PSF).

3 Propose Method

Without the information of length L and angle θ values, the deblurring image process is well known ill-posed problem. It takes time to explore the suitable length L and angle θ values. We proposed a model to reduce the time for deblurring.

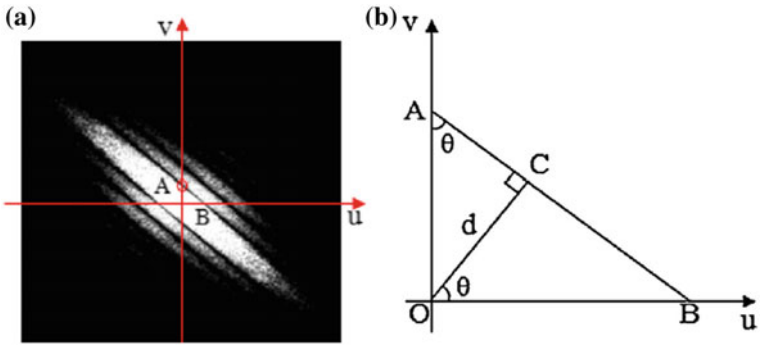


Fig. 2 Calculation of angle θ and length L **a** Fourier spectrum **b** Origin region of spectrum

The motion sensor is bound on the moving object to provide the related information. However, there are noise and drift problems when using the sensor data always. Generally, the Kalman filter is a Markov processing most frequently used to deal with such problems. Moreover, to get the length and angle parameters in the deblurring processing from sensor data still a tricky and complicated problem. We construct a relation table to link the speed groups v_i and suitable deblurring parameter θ_i and Li. Using the neural network helps to classify the speed groups to achieved our goal.

First, we will observed the relationship between speed of object and suitable deblurring parameter θ and L. Assumed that there are k groups of speed values, denoted as $\{v_1, v_2, \ldots, v_k\}$. Each speed group v_i have a specific range of speed values and suitable deblurring parameter θ_i and L_i. The motion sensor is used to detect the acceleration of camera. Each speed group v_i has j cases sensor data $\{m_i^1, m_i^2, \ldots, m_i^j\}$. The input of neural network is all $k \times j$ cases of sensor data. The neural network's automatic learning mechanism will train to classify the sensor data. After that, for a new sensor data input, the neural network will give the corresponding speed group and the suitable parameter can be got as shown in Fig. 3.

4 Experimental Results

The experimental environment is set up as follows. The hardware device includes a motion sensor: Arduino UNO with Arduino 9 Axes Motion Shield, a linear actuator: EASM4XD050ARAC, and a GoPro camera. The GoPro and motion sensor are bound

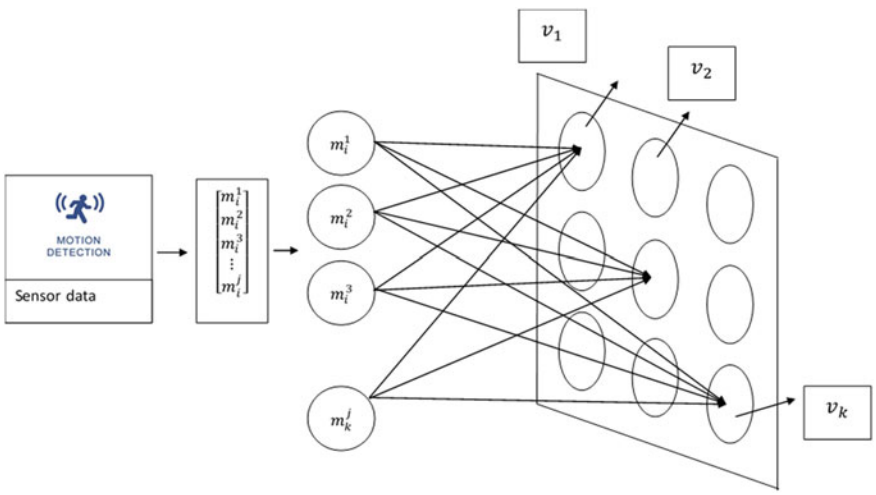


Fig. 3 A general view of SOM

on the linear actuator. There are three groups of speed, 18000 Hz $\leq v_1 \leq$ 22000 Hz, 38000 Hz $\leq v_2 \leq$ 42000 Hz, 58000 Hz $\leq v_3 \leq$ 62000 Hz. Each group contains five cases. There are totally 15 cases in the SOM neural network training process. The sensor data sampling frequency was 0.04 s and 242 packets were taken for experimental analysis for each case. Some acceleration data of one axe in each group are shown in the Fig. 4. Matlab is used to implement the SOM neural network process. The parameter learning rate = 0.2 and neighborhood distance = 0.8 are set in SOM neural network.

The training results of 15 cases in the SOM meet expectations. The data is classified to three correct groups. From case 1 to case 5, data is classified to cluster 1. From case 6 to case 10, data is classified to cluster 2. From case 11 to case 15, data is classified to cluster 3. The results of SOM are shown in Table 1. The 1 represented in the cluster and the 0 represents not in the cluster. After training, another 15 randomized test data are generated and use the already trained SOM

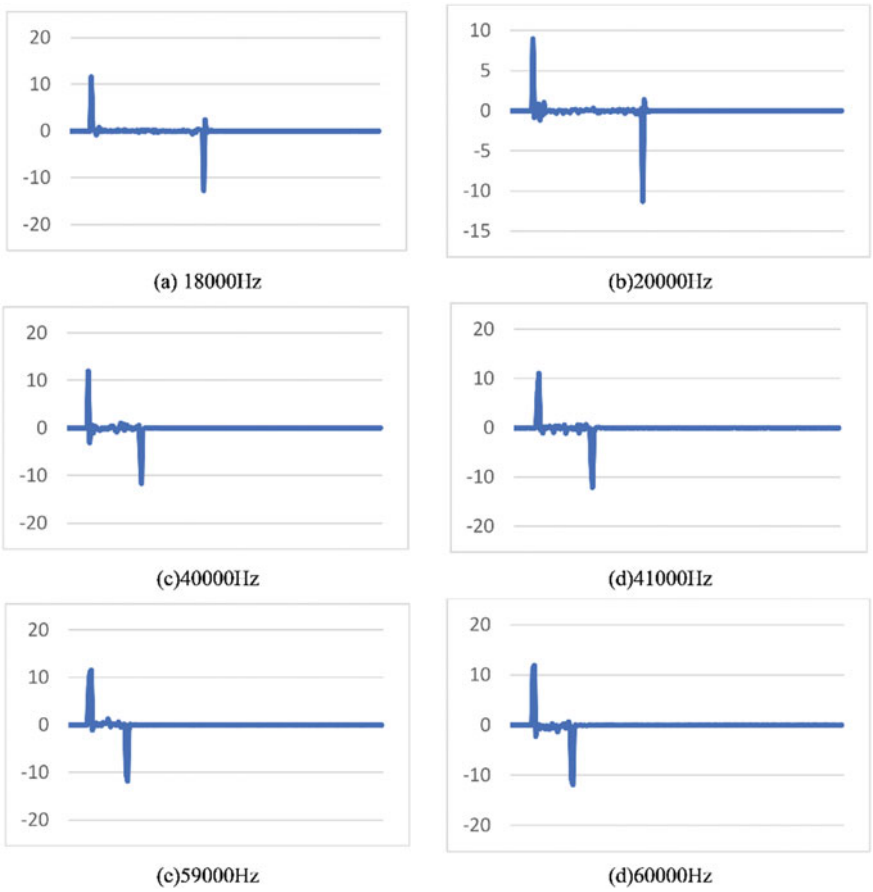


Fig. 4 Some acceleration data of one axe in each group

Table 1 The results of SOM for three groups in training phase

000001111100000
111110000000000
000000000011111

Table 2 The results of SOM for three groups in test phase

000000011100000
111111100000000
000000000011111

neural network to help classify. And test data input can be classified as shown in Table 2. There are two data classified into wrong group. The precision rate is 13/15.

5 Conclusion and Future Work

The main purpose of this paper is to explore how to reduce the complexity of ill-posed problem for deblurring image. The acceleration of motion sensor cannot use directly in the deblurring process. However, a range of speed of moving objects uses the same deblurring parameter. Hence, we established neural networks to automatic groups to estimate suitable parameter for deblurring process. In the experimental process, sensors coupled to linear slides are used to collect the acceleration of camera. The sensor data are used to train SOM neural network at different speeds. After that, the user only needs to feed the collected data into the neural network to know the speed and direction of the movement. It is not necessary to use Kalman filter and other algorithms to get the important parameter of deblurring process, and the efficiency of restoring blurred images can be improved.

In the experiments of this paper, the research framework proposed is only for horizontal movement, so only the factors of length L need to be considered. However, the images taken in real life may occur from different direction and different speed, the angle θ must also be taken into consideration. Therefore, in the future work, the situation in the two-dimensional movement deserves further research and analysis.

Acknowledgements This research was supported by the Ministry of Science and Technology of Taiwan under Grant MOST-106-2221-E-324-023-.

References

1. Couzinié-Devy F, Sun J, Alahari K, Ponce J (2013) Learning to estimate and remove non-uniform image blur. In: IEEE conference on computer vision and pattern recognition, 1075–1082. https://doi.org/10.1109/cvpr.2013.143

2. Oliveira JP, Figueiredo MAT, Bioucas-Dias JM (2014) Parametric blur estimation for blind restoration of natural images: linear motion and out-of-focus. IEEE Trans Image Process 23:466–477. https://doi.org/10.1109/TIP.2013.2286328
3. Arthur J, Attarian A, Hamilton F, Tran H (2018) Nonlinear Kalman filtering for censored observations. Appl Math Comput 316:155–166
4. Kohonen T (1982) Self-organized formation of topologically correct feature maps. Biol Cybern 43:59–69
5. Cheung YM, Law LT (2007) Rival model penalized self-organizing map. IEEE Trans Neural Netw 18:289–295. https://doi.org/10.1109/TNN.2006.885039
6. Wen CY, Lee CH (2002) Point spread functions and their applications to forensic image restoration. Forensic Sci J 1:15–26
7. Lee JM, Lee JH, Park KT, Moon YS (2013) Image deblurring based on the estimation of PSF parameters and the post-processing. Optik 124:2224–2224
8. Dash R, Majhi B (2014) Motion blur parameters estimation for image restoration. Optik 125:1634–1640

The Simulation of the Indoor Positioning by Panoramic Camera and Point Cloud Scanner

Jiun-Jian Liaw, Kun-Leng Chen, Tzu-Cheng Huang and Yu-Huei Cheng

Abstract Positioning system is more and more important but it is not easy to have indoor positioning without wireless signal. In this paper, we proposed a positioning method with digital panorama image and point cloud system. The point cloud system is used to scan the indoor environment, and filter out the feature objects with the coordinates. The panoramic camera is used to shoot the panoramic image and to find the location of feature objects in the environment. The proposed method calculates the position with the image which taken from panoramic camera and the coordinates from the point cloud system. According to the experimental results, the error in the simulation indoor space is less than 30 mm.

Keywords Indoor positioning · Panoramic image · Point cloud system Raspberry Pi · Fish-eye lens

1 Introduction

Positioning system is more and more important, and also developed a lot of new applications. Location-based service allows we have a quick guide when we enter the unfamiliar environment. GPS (Global Positioning System), storage location, traveling are applied to give us good service with positioning schemes [1–4].

J.-J. Liaw (✉) · K.-L. Chen · T.-C. Huang · Y.-H. Cheng
Department of Information and Communication Engineering,
Chaoyang University of Technology, Taichung City, Taiwan, Republic of China
e-mail: jjliaw@cyut.edu.tw

K.-L. Chen
e-mail: s10430621@cyut.edu.tw

T.-C. Huang
e-mail: s10630607@cyut.edu.tw

Y.-H. Cheng
e-mail: yhcheng@cyut.edu.tw

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_23

The main problem of positioning service is how to get the user's position accurately and improve the quality of positioning service. We can easily get good accuracy of position with GPS when we outside the building.

GPS is currently the most mainstream method of outdoor positioning, which is a system developed by the U.S. Department of Defense and used worldwide. It uses the ground equipment to receive time and location signals from the satellite equipment, and then calculated by the ground equipment to obtain the current location of equipment. In bad weather, GPS can also have considerable penetration. The accuracy of the typical GPS is about 10 m. The improved GPS, such as differential GPS, can improve the accuracy to 3–5 m [5]. However, since GPS uses the satellite signals, the accuracy is greatly reduced in areas with high building coverage. When we were indoors, we could not use the GPS because the building obscured the signal [6]. Some scholars proposed the indoor positioning technologies by the using of Wi-Fi, Bluetooth or RFID equipment.

The method of using Wi-Fi is to set up multiple wireless base stations to cover the area of we need to locate. The received RSSI (Received Signal Strength Indicator) of the base stations can be used to calculate the position by triangulation algorithm [7]. With this method, we must know where the wireless base stations are located. Since the interference of obstacle (such as wall) makes the Wi-Fi positioning accuracy is low, we also need enough wireless base stations to make good results of calculation. According to the Binary search method, we can have 1.19 m average error with 4 base stations in a 8 m $\times$ 8 m room [8].

The positioning method with Bluetooth is similar to with Wi-Fi, it deploy the iBeacon nodes to provide the signals of known location. We can calculate the position with the scanned RSSI of the iBeacon nodes. However, the wireless signals may interfere with the wall, obstacle, air temperature or air humidity [9]. According to the previous literature, we got 1.29 m average error with 5 iBeacons in a 15 m $\times$ 15 m room space [10].

RFID (Radio Frequency Identification) system is combined by the tag and reader. The reader can sense the signal from the tag with identification. In the application of positioning, the used tag is passive with no power supply itself [11]. To use the RFID to locate, we use the reader to sense nearby tags and obtain the position by the signal strengths. Since the signal range of passive tag is short, we need to deploy a lot of tags to get a reasonable accuracy. When we use RFID with AOA (angle of arrival) algorithm to be the indoor positioning, we can have 0.7 m average error in a 10 m $\times$ 10 m room space [12].

According to the above, we can see that, whether GPS, Wi-Fi, Bluetooth and RFID, are using wireless signals to locate the position. In the methods of indoor positioning, we have to deploy a lot devices (such as base station, iBecans or tags) in order to cover all the space and make sure we can get enough accuracy. However, the wireless signal will be interfered by indoor environment and it affects the accuracy of the positioning [7–13]. Moreover, we also need to deploy a lot device and use better algorithms to improve the performance [4]. In order to avoid the disadvantages of using wireless devices, in this paper, we proposed a positioning method with digital panorama image and point cloud system. We use the

point clouds to obtain the exact location of the feature objects, and the panorama image can be used to calculate the position with the feature objects.

2 Point Cloud System and Panorama Image

In the positioning system, we need the known objects' positions to be the reference points, such as the satellites, base stations, iBecans and tags in GPS, Wi-Fi, Bluetooth and RFID systems, respectively. In our proposed method, we use the indoor objects to be the known reference objects. We can use the ruler to measure the coordinate of the reference objects, but it will consume a lot of labor. In this paper, in order to obtain the exact coordinate of the feature objects, we use the point cloud system [14]. Point cloud system is a high-precision laser scanner. It uses the laser to scan the environment and records all the scanned points. Scanning refers to projecting energy to an object and calculating three-dimensional spatial information by reflection of energy.

Each scanned point is recorded with three-dimensional coordinates. Some point cloud system can also record the color with RGB [15]. The accuracy of the point cloud system is up to 1 mm. Point cloud system is often used for 3D reconstruction. We can build the digital model of real objects. These digital models have a very wide range of uses, including industrial design, robot guidance, geomorphology, biological information, criminal identification and digital cultural relic collection [16].

The panorama image is also used in the proposed method. The image taken by a common camera is about 30°. In order to capture a wider range of image, the camera derives use the wide-angle lenses from common lenses. In order to get 360-degree panoramic images, we can splice images shot by multiple wide-angle lenses. A single panoramic image has a 360-degree field of view, so all the details of the environment can be captured at the same time. We can take position of the feature objects in a panoramic image and the coordinates of the actual feature objects in real space to achieve the positioning function.

The wide-angle lens which can capture the image with (or more than) 180° angle, we call it as fisheye lens [17]. According to the imaging method, the image from fisheye lens can be divided into circular fisheye and full-frame fisheye [18]. An example of the circular fisheye image is shown in Fig. 1. There are two images taken from two fisheye cameras. In the right image of Fig. 1, edge between the wall and the ceiling (solid line between a and b) should be a straight line. But the line is distorted into a curve by the fisheye lens. Panoramic camera is a device for capturing panoramic images. There are serval cameras on a panoramic camera to capture images with different views at the same time. Panoramic camera splices the captured images to a panoramic image with 360° field of view [19]. Limited to the two-dimensional space on the image, we can't fully present the three-dimensional information. When the images are combined, the greater distortion occurs in the farther from the horizontal center line of the image [20].

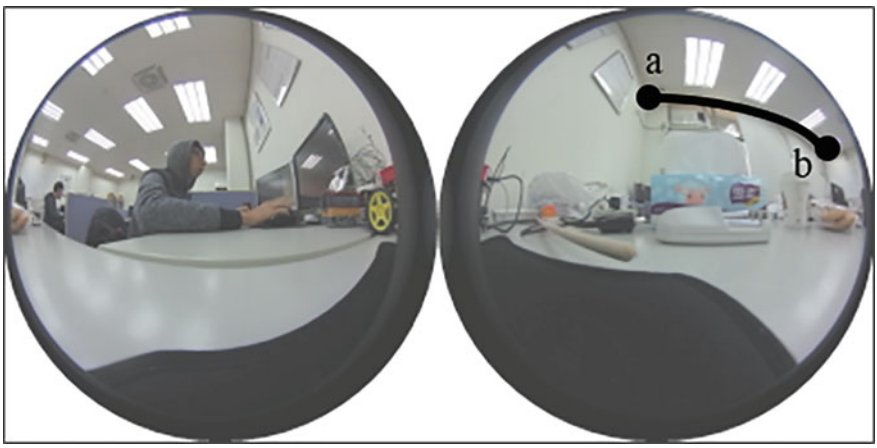


Fig. 1 The example of the circular fisheye images

We use the point cloud system to scan the indoor environment, and filter out the feature objects we want. The feature objects can be the windows, the clocks, paintings or other fixed and easily identifiable items. The three-dimensional coordinates of the feature objects can be obtained from the point cloud system. The panoramic camera is used to take the panoramic image. After the feature objects are extracted from the panoramic image, we can calculate the position with the three-dimensional coordinates of the feature objects.

3 Setup and Proposed Method

In this paper, we propose a method with panoramic image and point cloud system to implement an indoor positioning with a simulation space. We use the characteristic of the panoramic cameras, to shoot the panoramic image and to find the location of feature objects in the environment. The calculation of the feature object in the fisheye lens image, we can find the angle between the feature object and the imaging plane. Since the panoramic image includes the 360-degree view of the environment, we can capture all the feature objects at the same time. Each two feature objects are used to find an angle. The angle can be formed a circle with that two feature objects that the position should be on the circumference. The intersection of circles is the result of the positioning.

We use a handheld laser scanner to obtain the point cloud. The used scanner is FARO SCANNER FREESTYLE 3DX which the 3D point accuracy is less than 1.5 mm and the indoor scanning volume is up to 8 m^3. It is suitable for our simulation space. The scanner is used to obtain the point clouds of the space. And the SCENE Process (the software for the FARO) is used to filter out the feature

objects. The feature objects are filtered with color and the three-dimensional coordinates and saved as text format with ASCII code.

The panoramic camera we used in this paper is RICOH THETA S. It is with two 1200 pixels cameras and provides the panoramic image with 2048 × 1024 size. The panoramic camera connects to Raspberry Pi 3 with the official supported operating system. The Raspberry Pi is used to control the camera the process the captured image. The proposed method is programed with OpenCV (Open Source Computer Vision Library) by C language.

We put the panoramic camera with Raspberry Pi in the simulation space. The schematic diagram of top down view for the simulation space is shown on Fig. 2a. The rectangle in the Fig. 2a is the range of the simulation space. We set four feature objects which are denoted as F_1, F_2, F_3 and F_4. The location of the panoramic camera is denoted as C. Figure 2b shows the diagram of the camera, where y axis is the dire. We denote θ is the angle of the camera shooting direction and the feature object direction. For example, in Fig. 2, θ is the angle of F_1 direction and shooting direction.

The image taken from a fish-eye camera of the panoramic camera is shown in Fig. 3a, and the 3D diagram is shown in Fig. 3b. In Fig. 3, the feature object is denoted as F and projected in the image plane at F'. With this model, we can calculate θ by the geometry with Fig. 3b [21].

We use two feature objects to form a circle which is the possible location. See Fig. 4a, we sum the angles (θ) of F_1 and F_4, we can get the angle of φ. All the position that corresponds the angle of φ and the coordinates of F_1 and F_4 is the circle which is denoted as A in Fig. 4a. We also use F_1 and F_2 is to obtain another circle which is denoted as B in Fig. 4b. The intersection of the circles should be the goal position we want to recognize.

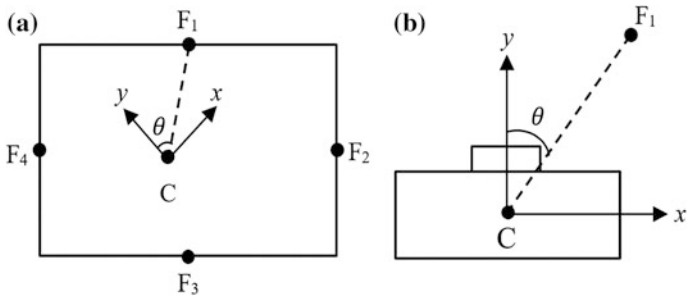


Fig. 2 The space and the camera. **a** the simulation space with four feature objects, and **b** the diagram of the camera

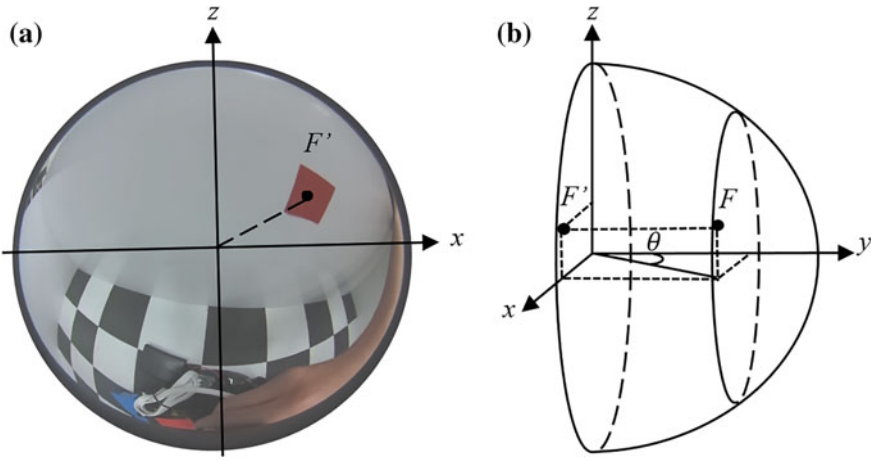


Fig. 3 The shoot image with the camera. **a** the image of the fish-eye camera, and **b** 3D coordinate diagram

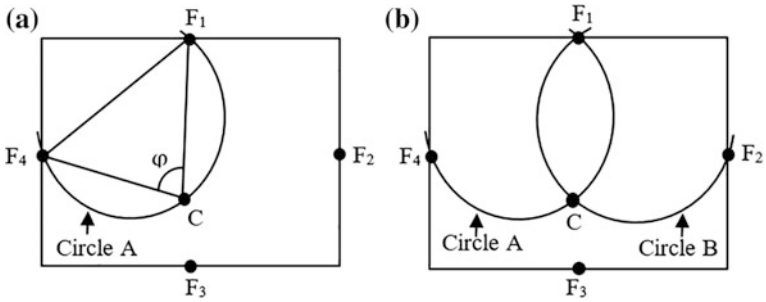


Fig. 4 The diagram of the positioning. **a** two feature objects form a circle, and **b** the intersection of two circles is the position

4 Experiment and Result

The simulate space is shown on Fig. 5. The length, width and height of the space is 90, 60 and 30 cm, respectively. We set four pieces of color paper on the four sides as the feature objects. The colors are red, blue, green and yellow, there are denoted as R, B, G and Y, respectively. We use the handheld laser scanner to obtain the coordinates of four feature objects. The coordinates are recorded ton the Raspberry Pi. The Raspberry Pi and the panoramic camera are mounted on a small car. The car moves in the space and the Raspberry Pi will calculate the position with the image which taken from panoramic camera. We also set another camera above the simulate space to recognize the real position of the car.

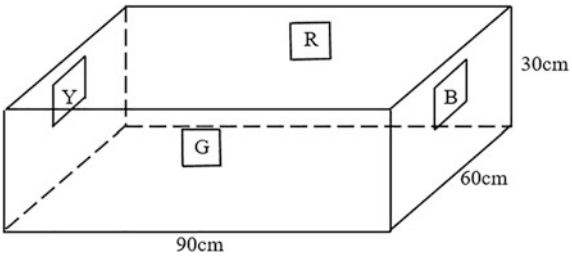


Fig. 5 The simulate space

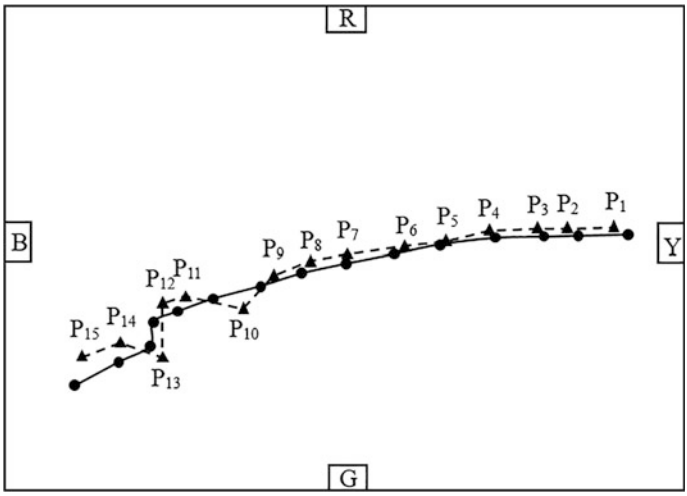


Fig. 6 The records of the moving path with recognized and real positions

We move the car from the right to the left of the space. The recognized position and the real car position are recorded. In the moving path, 15 points are recorded and shown on Fig. 6 with P_1–P_{15}. In Fig. 6, the moving path with recognized position by the proposed method is dotted line, and the moving path with real positon by the above camera is solid line. The error of these 15 points are shown in Table 1.

Table 1 The error of the positioning

Position	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
Error (mm)	16.21	12.65	7.47	6.33	2.72	13.42	8.70	10.73
Position	P_9	P_{10}	P_{11}	P_{12}	P_{13}	P_{14}	P_{15}	
Error (mm)	15.96	29.79	18.03	20.11	19.09	23.80	29.96	

5 Conclusions and Future Works

Positioning-based service allows we have a quick guide when we enter the unfamiliar environment. The common positioning systems are based on wireless signals which may be interfered by indoor environment. In order to avoid the disadvantages of using wireless devices, in this paper, we proposed a positioning method with digital panorama image and point cloud system. We use the point clouds to obtain the exact location of the feature objects, and the panorama image is used to calculate the position with the extracted feature objects. We set a simulation space to test the performance of the proposed method. The proposed method calculates the position with the image which taken from panoramic camera and the coordinates from the point cloud system. According to the experimental results, the error in the simulation indoor space is between 2 and 30 mm. In the coming future, we will try to implement the proposed method to a real indoor room and to compare the performance with other positioning schemes.

Acknowledgements This study is partial supported by "MOST 105-2221-E-324-008-MY2", Taiwan, Republic of China.

References

1. Leu JS, Tzeng HJ (2012) Received signal strength fingerprint and footprint assisted indoor positioning based on ambient Wi-Fi signals. In: 2012 IEEE 75th vehicular technology conference (VTC Spring), Yokohama, Japan, 6–9 May 2012
2. Cherubini A, Spindler F, Chaumette F (2012) A new tentacles-based technique for avoiding obstacles during visual navigation. In: 2012 IEEE international conference on robotics and automation, Saint, Paul, MU, USA, 14–18 May 2012
3. Wilk P, Karciarz J, (2014) Optimization of map matching algorithms for indoor navigation in shopping malls. In: 2014 international conference on indoor positioning and indoor navigation (IPIN), Busan, South Korea, 27–30 Oct 2014
4. Ma L, Li J, Xu Y (2015) Radio map recovery and noise reduction method for green WiFi indoor positioning system based on inexact augmented lagrange multiplier algorithm. In: 2015 IEEE global communications conference (GLOBECOM), San Diego, CA, USA, 6–10 Dec 2015
5. Grisso R, Alley M, Heatwole (2009) Precision farming tools: Global positioning system (GPS). In: Victorian certificate of education (VCE) publications, no 442, pp 442–503
6. Chen Z, Zhu Q, Jiang H (2015) Indoor localization using smartphone sensors and iBeacons. In: 2015 IEEE 10th conference on industrial electronics and applications (ICIEA), Auckland, New Zealand, pp 1723–1728, 15–17 June 2015
7. Abdullah A, Ma Y, Tafazolli R (2014) Survey of state-of-the-art fingerprinting technique for in-door radio positioning. In: Communications surveys and tutorials, pp 97–106, Oct 2014
8. Watcharasukchit S, Triyason T, Krathut W, Tassanaviboon A, Arpnikanondt C (2017) WiFi indoor positioning with binary search method. In: International conference on Ubi-media computing and workshops (Ubi-Media), Pattaya, Thailand, 1–4 Aug 2017
9. Chen Z, Zhu Q, Soh YC (2016) Smartphone inertial sensor-based indoor localization and tracking with iBeacon corrections. IEEE Trans Industr Inf 12(4):1540–1549

10. Li Z, Yang Y, Pahlacan K (2016) Using iBeacon for newborns localization in hospitals. In: International conference on medical information and communication technology, Worcester, MA, USA, 20–23 Mar 2016
11. Wang C, Shi Z, Wu F, Zhang J (2016) An RFID indoor positioning system by using particle swarm optimization-based artificial neural network. In: International conference on audio, language and image processing, Shanghai, China, 11–12 July 2016
12. Wang J, Wang Y, Guan X (2016) An indoor localization system based on backscatter RFID tag. In: IEEE wireless communications and networking conference, Doha, Qatar, 3–6 Apr 2016
13. Liu Y, Qiu Y (2012) An indoor localization of UHF RFID using a hybrid approach. In: International conference on consumer electronics, communications and networks, Yichang, China, 21–23 Apr 2012
14. Xia H, Huang T, Chen G (2015) A method of extracting human facial feature points based on 3D laser scanning point cloud data. In: 2015 23rd international conference on geoinformatics, pp 1–3, Wuhan, China, 19–21 June 2015
15. You H, Zhang S (2006) Reconstructing 3D buildings based on airborne CCD image and laser scanning range-finder data. Opt Precis Eng 14(2):297–302
16. Bu L, Zhang Z (2008) Application of point clouds from terrestrial 3D laser scanner for deformation measurements. In: The international archives of the photogrammetry, remote sensing and spatial information sciences, pp 545–548
17. Bassford M, Painter B (2015) Development of an intelligent Fisheye camera. In: 2015 international conference on IE, pp 160–163, Prague, Czech Republic, 15–17 July 2015
18. Zhang Q, Kamata SI (2015) Fisheye image correction based on straight-line detection and preservation. In: 2015 international conference on SMC, pp 1793–1797, Kowloon, Hong Kong, 9–12 Oct 2015
19. Xu J, Stenger G, Kerola T, Tung T (2017) Pano2CAD: room layout from a single panorama image. In: 2017 IEEE winter conference on WACV, Santa Rosa, CA, USA, 15 May 2017
20. Kweon GI, Choi YH (2010) Image-processing based panoramic camera employing single Fisheye lens. J Opt Soc Korea 14(3):245–259
21. Ho T, Budagavi M (2017) Dual-fisheye lens stitching for 360-degree imaging. In: 2017 IEEE international conference on ICASSP, New Orleans, LA, USA, 19 June 2017

Part III
Applications

A New Data Normalization Method for Multi-criteria Decision Analysis

Zhigao Chen and Renyan Jiang

Abstract Due to the incommensurability between the criteria in the multiple criteria decision-making problem, we need to standardize the scores of alternatives against each criterion. Although the existing normalization method can transform the criterion scores into (0, 1), the averages of the criterion scores of all the alternatives for different criteria may be different. This is equivalent to the case that the criteria weights are revised. In order to solve this problem, a non-linear data normalization method is proposed in this paper. The proposed method can transform each criterion scores into (0, 1) and make the scores for each criterion have the same mean. Two examples are included to illustrate the rationality of the proposed approach.

Keywords Multiple criteria decision-making · Data transformation
Data normalization

1 Introduction

There are many multiple criteria decision making (MCDM) problems in various fields, and typical MCDM problems generally include the following inputs (e.g., see [1–4]): (a) a set of criteria or attributes, (b) a set of alternatives and (c) a decision matrix, whose entries are the criteria scores of alternatives. Then, the decision maker needs to specify the criteria weights, normalize the criteria scores and aggregate the normalized criteria scores into an overall performance score or preference for each alternative. Finally, the alternatives are ranked or/and selected based on their overall

Z. Chen (✉)
School of Mathematics and Statistics, Changsha University of Science and Technology, Changsha 410114, Hunan, China
e-mail: 734316364@qq.com

R. Jiang
School of Automotive and Mechanical Engineering, Changsha University of Science and Technology, Changsha 410114, Hunan, China

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_24

performance scores. Clearly, the key issues are determination of criterion weights, normalization of criterion scores and the aggregation of the normalized criterion scores. In this paper, we focus on the normalization of criterion scores.

Ideally, the normalized criteria scores should meet the following requirements:

(a) They are non-dimensional,
(b) They are larger-the-better or larger-the-more-important,
(c) They fall in the same range, typically (0, 1), and
(d) They have the same "magnitude" in terms of mean.

There are some data transformation methods for multi-criteria decision analysis, such as vector normalized transformation, linear transformation and so on (e.g., see [2]). To my best knowledge, no single method can meet the above four requirements at the same time. Most methods make the criteria scores dimensional but fail to meet the last requirement. When the "magnitudes" of the normalized criteria scores are different, it is equivalent to the case that the criteria weights are revised in an implicit way, and this may result in improper decision.

In order to solve this problem, a non-linear data normalization method is proposed in this paper. The proposed method can transform each criterion scores into (0, 1) and make the scores for each criterion have the same mean. Two examples are included to illustrate the rationality of the proposed approach.

The paper is organized as follows. The proposed approach is presented in Sect. 2 and illustrated in Sect. 3. The paper is concluded in Sect. 4.

2 Proposed Approach

2.1 Decision Matrix

The MCDM problem under consideration is depicted by the following decision matrix of preferences for M alternatives rated on N criteria. Let C_j be the jth criterion ($j = 1, \ldots, N$) and A_i the ith alternative ($i = 1, \ldots, M$) (e.g., see [5]). Suppose r_{ij} stands for the value of a criterion C_j with respect to an alternative A_i ($i = 1, \ldots, M$; $j = 1, \ldots, N$). Then a quantitative MCDM problem of ranking M alternatives based on N criteria may be represented using the following decision matrix, as shown in Table 1.

The data transformation method involved in the following section is carried out in the decision matrix.

2.2 Exiting Data Transformation Methods

We know that because of the incommensurability and ambivalence among the criteria, for every original score of each alternative against each criterion we cannot directly use it for analysis, but must process them first. Data processing includes

Table 1 Decision matrix

Alternative	Criterion					
	C_1	C_2	…	C_j	…	C_N
A_1	r_{11}	r_{12}	…	r_{1j}	…	r_{1N}
A_2	r_{21}	r_{22}	…	r_{2j}	…	r_{2N}
⋮	⋮	⋮		⋮		⋮
A_i	r_{i1}	r_{i2}	…	r_{ij}	…	r_{iN}
⋮	⋮	⋮		⋮		
A_M	r_{M1}	r_{M2}	…	r_{Mj}	…	r_{MN}

three parties including specification of r_{ij} (r_{ij} denotes the original score of each alternative against each criterion), transformation of r_{ij} and normalization of r_{ij} (e.g., see [6]).

There are already perfect methods for the first two parts. In this paper we will focus on the third part. There are many ways to standardize criteria scores. Different methods can be chosen according to the collective situation. Common methods are the following (e.g., see [2]):

(a) Vector normalized transformation method. Let

$$x_{ij} = r_{ij}/(\sum_{i=1}^{M} r_{ij}^2)^{1/2} \tag{1}$$

(b) Linear transformation method. Let $r_j^{\max} = \max_i r_{ij}$, $r_j^{\min} = \min_i r_{ij}$, $1 \le i \le M$. If the attribute of C_j is "larger-the-better" [smaller-the-better], we can normalize criteria scores by Eq. (5) [(6)] below

$$x_{ij} = r_{ij}/r_j^{\max} \tag{2}$$

$$x_{ij} = r_j^{\min}/r_{ij} \tag{3}$$

(c) Range transformation method. Let

$$x_{ij} = (r_{ij} - r_j^{\min})/(r_j^{\max} - r_j^{\min}) \tag{4}$$

$$x_{ij} = (r_j^{\max} - r_{ij})/(r_j^{\max} - r_j^{\min}) \tag{5}$$

(d) Proportional transformation method. Let

$$x_{ij} = r_{ij}/\sum_{i=1}^{M} r_{ij}. \tag{6}$$

The basic purpose of all the above transformations method is to standardize the scores of the criteria so that they can be numerically compared.

In this paper, we use the simple additive weighting (SAW) model to aggregate the normalized criteria scores into an overall performance (e.g., see [7]). To make the SAW model meaningful, the transformation of the criteria scores should meet the following two requirements:

- It is necessary to transform all the criteria scores into (0, 1) and the best criterion score should be 1, and the worst is 0. The purpose of transformation is to put all the criteria under the same standards.
- The means of the transformed criteria scores should be equal. Because the weights of the criteria have been generated, if the average values of the scores of all the alternatives against each criterion are not the same, which in fact shows the difference of the importance of each criterion and also it is equivalent to the case that the criteria weights are revised.

In fact, there is not a very good data transformation method to meet the above two points at the same time. In this paper, we propose a non-linear criteria score transformation method which can meet the above two requirements at the same time.

2.3 New Approach

Assume the decision matrix given is the form of Table 1. The new data transformation method proposed involves the following multi-step procedure:

Step 1: Using linear transformation method or range transformation method to transform the criteria scores into (0, 1), and

Step 2: Calculating the average value of the criteria scores transformed (denoted as $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_M$), then calculating the overall mean (denoted as $\bar{x}$)

Step 3: Making the mean of each criterion score equal to $\bar{x}$ by a power transformation. Let

$$y_{ij} = x_{ij}^{\mu_j}, j = 1, 2, \ldots, N \tag{7}$$

where μ_j can be calculated by the Solver of Excel.

We use a simple example to illustrate how the new data transformation method works. There are ten alternatives and each alternative defines two criteria denoted as C_1, C_2, respectively. The criterion C_1 is the income indicators while the criterion C_2 is the cost indicators. The data for this example come from Ref. [8] and are shown in Table 2.

The raw data are shown in the 2nd, 3rd columns, and the normalized data are shown in the 4th and 5th columns in Table 2, where the 4th column of data are transformed according to Eq. (4), and the 5th column is derived from Eq. (5).

Table 2 Raw data and two transformation data

i	C_1	C_2	C_1	C_2	C_1	C_2
1	126	22000	0.5000	0.5839	0.5628	0.5152
2	105	38000	0.0000	0.0000	0.0000	0.0000
3	138	14000	0.7857	0.8759	0.8187	0.8493
4	140	13000	0.8333	0.9124	0.8597	0.8931
5	147	10600	1.0000	1.0000	1.0000	1.0000
6	116	27000	0.2619	0.4015	0.3292	0.3246
7	112	32000	0.1667	0.2190	0.2263	0.1538
8	132	17000	0.6429	0.7664	0.6932	0.7204
9	122	23500	0.4048	0.5292	0.4723	0.4563
10	135	15000	0.7143	0.8394	0.7565	0.8059

The 6th and 7th columns of data in Table 2 are derived from the new transformation method. We record the average value of data from the 4th column to the 7th column as $\bar{x}_1, \bar{x}_2, \bar{x}_1^*, \bar{x}_2^*$, respectively. Here

$$\bar{x}_1 = 0.5310,\ \bar{x}_2 = 0.6128,\ \bar{x}_1^* = \bar{x}_2^* = \bar{x} = 0.5719.$$

We can see $\bar{x}_1$, $\bar{x}_2$ are unequal, which means that C_1, C_2 have different importance and through a power transformation, they are equal to the total mean 0.5719.

In order to transform normalized data with the same mean, we used two power transformations, the power exponents are 0.8294.1.2328, respectively. In fact, this power transformation is based on standardized data that have fallen into (0, 1), so this new transformation approximates a linear one (as long as the power exponent is not too large). The plot of the two power transformations is shown in Fig. 1. The curves of these two power transformations are very close to the straight lines, so we name the transformation a pseudo linear transformation. The characteristics of this new transformation method are obvious, do not change the order and range of the original data, but change the mean of the original data.

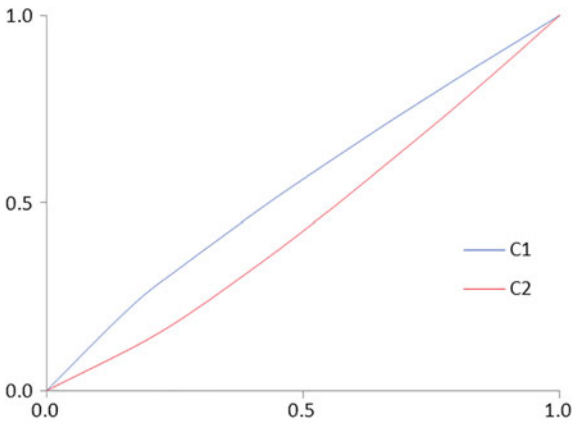


Fig. 1 Plot of power transformation

3 Illustrations

In this section we illustrate the validity and rationality of the proposed approach using two real-world examples published in the literature.

3.1 Example of Flexible Manufacturing System Selection Problem

A set of eight flexible manufacturing systems are evaluated with seven objective attributes (e.g., see [9]): reduced labor costs (%), reduced storage in the product (%), reduced setup costs (%), improved market response, quality improvement, cost and maintenance costs ($1,000), use area ($m^2$), which are denoted as C_1 to C_7, respectively. The attributes of C_1 to C_5 are larger-the-better, while the attributes of the rest are smaller-the-better. Reference [8] uses the AHP to determine the criterion weights, which are 0.1129, 0.1129, 0.0445, 0.1129, 0.2862, 0.2862 and 0.0445, respectively. The raw values of r_{ij} are shown in Table 3; the normalized data are shown in Table 4; the quasi linear transformed data are shown in Table 5; the power exponents are shown in the last row of Table 5. The problem is to select the best alternative.

From Fig. 2, we can see clearly that under the linear transformation the mean of each group of criteria scores vary, which in fact means that each criterion has different importance, that is, they have different criterion weights, while under the pseudo linear transformation, all the means of criteria scores are the same.

As we can see in Table 6, the ranking results for the three different approaches are different, but the differences are small and the same alternative has no fluctuation in rankings over 2 positions. The rankings obtained under the linear transformation and the pseudo linear approaches are broadly the same, only the third and fourth exchange.

Table 3 Raw data for example 3.1

	C_1	C_2	C_3	C_4	C_5	C_6	C_7
1	30	23	5	0.745	0.745	1500	5000
2	18	13	15	0.746	0.745	1300	6000
3	15	12	10	0.5	0.5	950	7000
4	25	20	13	0.746	0.745	1200	4000
5	14	18	14	0.255	0.745	950	3500
6	17	15	9	0.746	0.5	1250	5250
7	23	18	20	0.5	0.745	1100	3000
8	16	8	14	0.255	0.5	1500	3000

Table 4 Normalized data for example 3.1

	C_1	C_2	C_3	C_4	C_5	C_6	C_7
1	1.0000	1.0000	0.2500	0.9987	1.0000	0.6333	0.6000
2	0.6000	0.5652	0.7500	1.0000	1.0000	0.7308	0.5000
3	0.5000	0.5217	0.5000	0.6702	0.6711	1.0000	0.4286
4	0.8333	0.8696	0.6500	1.0000	1.0000	0.7917	0.7500
5	0.4667	0.7826	0.7000	0.3418	1.0000	1.0000	0.8571
6	0.5667	0.6522	0.4500	1.0000	0.6711	0.7600	0.5714
7	0.7667	0.7826	1.0000	0.6702	1.0000	0.8636	1.0000
8	0.5333	0.3478	0.7000	0.3418	0.6711	0.6333	1.0000

Table 5 Pseudo linear transformed data for example 3.1

	C_1	C_2	C_3	C_4	C_5	C_6	C_7
1	1.0000	1.0000	0.4147	0.9985	1.0000	0.5137	0.6258
2	0.6876	0.6235	0.8330	1.0000	1.0000	0.6329	0.5294
3	0.6016	0.5835	0.6439	0.6326	0.2831	1.0000	0.4596
4	0.8749	0.8907	0.7607	1.0000	1.0000	0.7113	0.7680
5	0.5719	0.8163	0.7973	0.2928	1.0000	1.0000	0.8681
6	0.6594	0.7019	0.6023	1.0000	0.2831	0.6702	0.5984
7	0.8230	0.8163	1.0000	0.6326	1.0000	0.8075	1.0000
8	0.6307	0.4171	0.7973	0.2928	0.2831	0.5137	1.0000
μ_j	0.7331	0.8281	0.6350	1.1444	3.1648	1.4584	0.9176

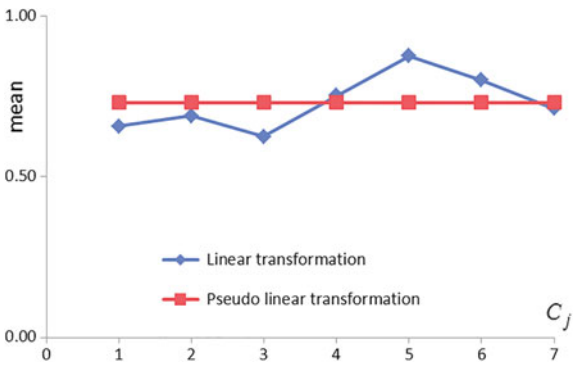


Fig. 2 Mean plot of two transformation methods for example 3.1

Table 6 Comparison of the ranking of the three methods for example 3.1

Alternative	Linear transformation	TOPSIS	New method
1	3	3	4
2	5	4	5
3	6	8	6
4	1	2	1
5	4	5	3
6	7	6	7
7	2	1	2
8	8	7	8

3.2 Example of Rapid Prototyping System Selection Problem

The data shown in Table 5 come from Ref. [10] and deal with the criteria scores of 6 alternative systems. There, criterion C_1 is accuracy, criterion C_2 is surface roughness (μm), criterion C_3 is tensile strength (mpa), criterion C_4 is elongation (%), criterion C_5 is production costs, criterion C_6 is production time. Obviously, The attributes of C_3 to C_4 are larger-the-better, while the attributes of the rest are smaller-the-better. Reference [9] uses the AHP to determine the criterion weights, which are 0.3185, 0.3185, 0.1291, 0.1291, 0.0524, and 0.0524, respectively.

The raw values of r_{ij} are shown in the 2nd to 7th columns of Table 7; the quasi-linear transformed data based on linear transformation are shown in Table 8; the pseudo linear transformed data based on range transformation are shown in Table 9; the power exponents corresponding to two quasi-linear transformations are shown in last rows of Tables 8 and 9, respectively. The problem is to select the best alternative.

The means of the criterion scores obtained under the four different transformation methods are shown in Fig. 3, which are linear transformation (denoted as L), range transformation (denoted as R), pseudo linear transformed data based on linear transformation (denoted as P-L), pseudo linear transformed data based on range transformation (denoted as P-R). We can see that the fluctuations of the means of the criteria scores obtained under the linear and range transformation methods are relatively large and reflect the different importance among the criteria. Two kinds of quasi-linear transformations can eliminate these differences.

Table 7 Raw data for example 3.2

	C_1	C_2	C_3	C_4	C_5	C_6
1	120	6.5	6.5	5	0.745	0.5
2	150	12.5	40	8.5	0.745	0.5
3	125	21	30	10	0.665	0.745
4	185	20	25	10	0.59	0.41
5	95	3.5	30	6	0.745	0.41
6	600	15.5	5	1	0.135	0.255

Table 8 Pseudo linear transformed data I for example 3.2

	C_1	C_2	C_3	C_4	C_5	C_6
1	0.6888	0.6832	0.1197	0.2062	0.4274	0.4281
2	0.4824	0.4569	1.0000	0.6906	0.4274	0.4281
3	0.6453	0.3320	0.7146	1.0000	0.4523	0.2590
4	0.3452	0.3421	0.5775	1.0000	0.4800	0.5497
5	1.0000	1.0000	0.7146	0.3124	0.4274	0.5497
6	0.0528	0.4002	0.0881	0.0053	1.0000	1.0000
μ_{1j}	1.5690	0.6154	0.1682	2.1278	0.4976	1.2601

Table 9 Pseudo linear transformed data I I for example 3.2

	C_1	C_2	C_3	C_4	C_5	C_6
1	0.7107	0.8811	0.0973	0.1643	0.0000	0.4249
2	0.4604	0.6150	0.2370	0.6662	0.0000	0.4249
3	0.6623	0.0000	0.8922	1.0000	0.2935	0.0000
4	0.2671	0.1456	0.7787	1.0000	0.5687	0.6253
5	1.0000	1.0000	0.8891	0.2700	0.0000	0.6253
6	0.0000	0.4588	0.2062	0.0000	2.2382	1.0000
μ_{2j}	6.7266	0.6732	0.7398	2.2275	2.2382	1.2347

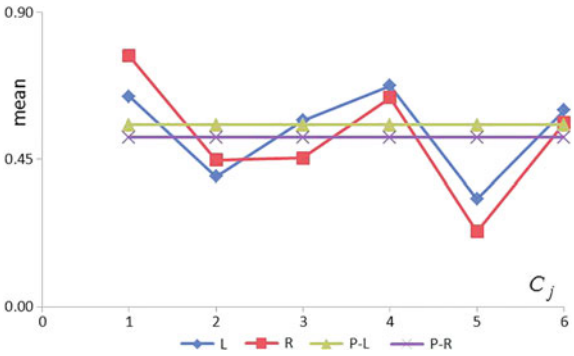


Fig. 3 Mean plot of four transformation methods for example 3.2

Table 10 shows that the rankings obtained under the five methods are different, and the rankings obtained by the two quasi-linear transformations are not significantly different. If the ranking obtained by TOPSIS method is used as a reference, the quasi-linear transformation method based on range transformation (denoted as pseudo linear2) has the closest rankings, indicating that pseudo linear2 method is a better method.

Table 10 Comparison of the ranking of the five methods for example 3.2

Alternative	Linear	Range	TOPSIS	Pseudo linear1	Pseudo linear2
1	4	3	1	4	2
2	2	2	3	3	3
3	3	5	4	2	4
4	5	4	5	5	5
5	1	1	2	1	1
6	6	6	6	6	6

4 Conclusions

In this paper, we have proposed a quasi-linear transformation method for the standardization of the normalized criteria scores. With the new method, the data can be transformed into (0, 1), and the best criterion scores is 1. If the quasi-linear transformation is based on range transformation, then the best criterion score is transformed to 1 and the worst is transformed to 0. Obviously, all the criteria scores have the same mean. It can eliminate the different importance of the criterion scores, which truly standardize the criteria scores so that the decision-making results obtained are more objective and reasonable.

The pseudo linear transformation is a good method in MCDM problem, both in theory and practice. When make multi-criteria decision-making, we can jointly use the quasi-linear transformation technique and other methods. Through comparison and analysis of several different results we can get more scientific decision-making.

The results obtained by the new non-linear approach are roughly the same as those obtained by the TOPSIS method. There is some kind of internal connection between the two methods, which deserves further research.

Acknowledgements The research was supported by the National Natural Science Foundation of China (No. 71771029).

References

1. Hwang CL (1981) Multiple attribute decision making. Springer Verlag
2. Maccrimmon KR (1968) Decision making among multiple attribute alternatives: a survey and consolidated approach. Rand
3. Stewart TJ (1992) A critical survey on the status of multiple criteria decision making theory and practice. OMEGA 20(5–6):569–586
4. Saaty TL (1994) Fundamentals of decision making and priority theory with the analytic hierarchy process. RWS Publications, Pittsburgh, PA
5. Rao RV (2007) Decision making in the manufacturing environment. Springer, London
6. Jiang R (2015) Introduction to quality and reliability engineering. Springer, London & Science Press, Beijing

7. Fisher MR, Churchman CW, Ackoff RL, Arnoff EL (1957) Introduction to Operations Research. Econ J 69(276):785
8. Besharati B, Azarm S, Kannan PK (2006) A decision support system for product design selection: a generalized purchase modeling approach. Decis Support Syst 42(1):333–350
9. Karsak EE, Kuzgunkaya O (2002) A fuzzy multiple objective programming approach for the selection of a flexible manufacturing system. Int J Prod Econ 79(2):101–111
10. Byun HS, Lee LH (2005) A decision support system for the selection of a rapid prototyping process using the modified TOPSIS method. Int J Adv Manuf Technol 26(11):1338–1347

A Signal-to-Noise Ratio Based Optimization Approach for Data Cluster Analysis

Renyan Jiang and Chaoqun Huang

Abstract There are many cluster analysis problems in the context of multi-criteria decision analysis. These problems often need to simultaneously determine the number of clusters and their boundaries. There is no good method available to automatically determine the number of clusters. In this paper, we propose a simple and intuitive approach to address this issue. The proposed approach first aggregates a set of multi-criteria or multi-attribute data into a one-dimensional data set. Then, we consider an arbitrary data point, which divides the dataset into two groups. The between-groups distance and within-group variances are combined into a clustering quality measure called the signal-to-noise ratio (SNR). The plot of SNR versus each data point provides the clue about the number of clusters and their boundaries. Specifically, the cluster boundaries are at the local maxima of the plot; and this also simultaneously determines the number of clusters. The proposed approach can be conveniently implemented using an Excel spreadsheet program. Two real-world examples are included to illustrate the appropriateness of the proposed approach. The results are also validated through comparing them with the results obtained from the Gaussian kernel density estimation.

Keywords Multi-criteria decision · Cluster analysis · Signal-to-noise ratio Kernel density estimation

R. Jiang (✉)
Faculty of Automotive and Mechanical Engineering, Changsha University of Science and Technology, Changsha 410114, Hunan, China
e-mail: jiang@csust.edu.cn

C. Huang
Xinjiang Goldwind Science & Technology Co., Ltd., No. 8 Bo Xing 1st Road, Economic & Technological Development Zone, Beijing 100176, China

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_25

1 Introduction

There are many cluster analysis problems in the context of multi-criteria decision analysis (e.g., see [1, 2]). These problems usually involve a set of multi-dimensional data, which can be further aggregated into a one-dimensional data set. In this paper, we focus on clustering for such data.

Many clustering problems often need to simultaneously determine the number of clusters (denoted as K) and their boundaries. Most of existing clustering methods cannot automatically determine the value of K. For example, the well-known K-means algorithm is implemented only for a pre-specified value of K (e.g., see [1]). The mean shift algorithm or kernel density estimation can determine the value of K by identifying the modes of kernel density estimation of data but its clustering quality strongly depends on the smoothing parameter or bandwidth (e.g., see [3, 4]). As such, a simple and intuitive approach is desired to automatically determine the number of clusters. This paper addresses this issue for a one-dimensional dataset aggregated from a set of multi-criteria or multi-attribute data.

The proposed approach is based on the definition of Jain [1] for an ideal cluster, which is "a set of points that is compact and isolated". Consider an arbitrary data point of a one-dimensional dataset. It divides the dataset into two groups. The sum of within-group variances reflects "*compactness*" and the between-groups distance reflects "*isolation*". We define the ratio of the between-groups distance and the square root of sum of within-group variances as a clustering quality measure called the signal-to-noise ratio (SNR). If the SNR associated with a certain data point is a local maximum, this point can be viewed as a cluster boundary. Specifically, we draw and examine the plot of SNR versus each data point. The cluster boundaries are at the local maxima of the plot. When all the cluster boundaries are identified, the number of clusters is simultaneously determined.

The proposed approach can be conveniently implemented using an Excel spreadsheet program. Two real-world examples are included to illustrate the appropriateness of the proposed approach. The appropriateness is further confirmed by the results obtained from the Gaussian kernel density estimation.

2 Proposed Approach and Its Validation

2.1 Proposed Clustering Approach

Consider n items, whose performances are measured against m criteria. Let x_{ij} denote the score or value of the ith item under the jth criterion. As such, the item is characterized by a multi-dimensional data set given by

$$(x_{ij}, 1 \le i \le n, 1 \le j \le m). \tag{1}$$

For a specific problem, we can combine all the criteria scores or criteria values into an overall performance score or overall performance value

$$S_i = g(x_{ij}, 1 \le j \le m), 1 \le i \le n \tag{2}$$

where $g(.)$ is a non-negative aggregation operator. The problem is to classify the items into K (which is unknown) groups or classes based on dataset

$$(S_i, 1 \le i \le n). \tag{3}$$

We sort the data in Eq. (3) in an ascending order and denote the reordered data as

$$s_1 \le s_2 \le \cdots \le s_n. \tag{4}$$

Consider an arbitrary point s_p (called the bisecting point), $1 \le p \le n-1$. It divides data set $(s_i,\ 1 \le i \le n)$ into two groups:

$$(s_1, \ldots, s_p), (s_{p+1}, \ldots, s_n). \tag{5}$$

The group means are defined as

$$\mu_1(p) = \frac{1}{p}\sum_{i=1}^{p} s_i, \quad \mu_2(p) = \frac{1}{n-p}\sum_{i=p+1}^{n} s_i. \tag{6}$$

The group variances are defined as

$$v_1(p) = \frac{1}{p}\sum_{i=1}^{p} [s_i - \mu_1(p)]^2, \quad v_2(p) = \frac{1}{n-p}\sum_{i=p+1}^{n} [s_i - \mu_2(p)]^2. \tag{7}$$

We define SNR as

$$SNR(p) = [\mu_2(p) - \mu_1(p)]/\sqrt{v_1(p) + v_2(p)}. \tag{8}$$

For a good bisecting point, between-group distance, $\mu_2(p) - \mu_1(p)$, should be as large as possible and the sum of within-group variances, $v_1(p) + v_2(p)$, should be as small as possible. As a result, SNR should be as large as possible. We draw the plot of SNR versus p. Generally, the SNR plot can have one maximum or more local maxima. For example, the SNR plot shown in Fig. 1 has two local maxima. As such, through examining the SNR plot we can determine: (a) the number of clusters, which equals to the number of local maxima plus one, and (b) the items in each group.

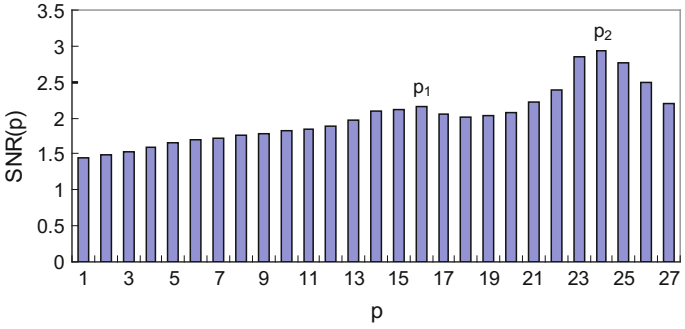


Fig. 1 Plot of SNR(p) versus p for example 1

2.2 *Validation of the Results*

Assume that the data given by Eq. (4) come from a well-separated K-fold normal mixture model. For data point s_i, we use the normal probability density function (pdf), $\phi(x; s_i, h)$, to approximate it. Here, h is a smoothing parameter called the bandwidth. The non-parametric estimation (i.e., kernel density estimation, [5, 6]) of the normal mixture pdf is given by

$$f(x;h) = \frac{1}{n}\sum_{i=1}^{n}\phi(x; s_i, h). \tag{9}$$

Equation (9) depends on the value of h. A small [large] value of h will lead to a significantly fluctuated [over-smoothed] empirical pdf. Therefore, the value of h needs to be carefully selected. Silverman [5] presents a rule-of-thumb bandwidth estimator, given by

$$h = \sigma/(0.75n)^{0.2}, \tag{10}$$

where σ is the sample standard deviation. According to Wikipedia [6], Eq. (10) can lead to a significantly over-smoothed kernel density estimation. Therefore, the h obtained from (10) can be viewed as an upper limit of h, and we denote it as h_u.

Let h_l denote the lower limit of h. It can be approximately estimated as follows. Consider the range of $(s_1 - h_l, s_n + h_l)$. We take the mean inter-point distance of the data set as h_l, i.e.,

$$h_l \approx [(s_n + h_l) - (s_1 - h_l)]/n. \tag{11}$$

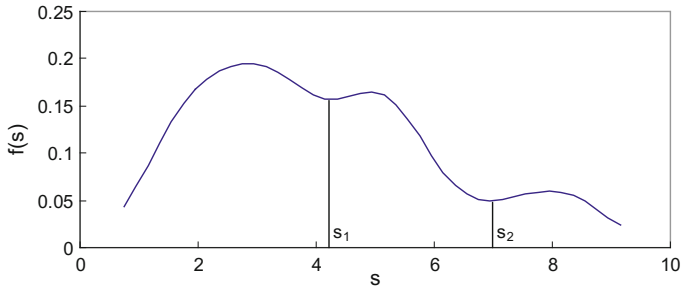


Fig. 2 Normal kernel density estimation for example 1

From Eq. (11), we have

$$h_l \approx (s_n - s_1)/(n-2). \tag{12}$$

It appears reasonable to take h as the arithmetic average of the upper and lower limits of h, i.e.,

$$h \approx (h_l + h_u)/2. \tag{13}$$

For a given dataset, using Eqs. (13) to (9) we can obtain the plot of the normal kernel density estimation (e.g., see Fig. 2). Through examining the kernel density estimation plot, we can determine: (a) the number of clusters, which equals to the number of local maxima, and (b) the boundary points between groups, which correspond to the local minima. As such, the proposed approach can be validated if the classification obtained from the kernel density estimation is roughly consistent with the classification obtained from the proposed approach.

3 Illustrations

In this section we illustrate the appropriateness and usefulness of the proposed approach using two real-world examples published in the literature.

3.1 Example 1

The data shown in Table 1 deal with 27 robots (RT), which are evaluated by repeatability (REP) in mm, cost in $\$10^4$, capacity (CAP) in kg and velocity (VEL) in m/s.

The data have been studied by several authors (e.g., [7, 8]) using different methods (e.g., the utility model, TOPSIS and DEA). In this paper, we focus on their classification. Since the first [latter] two attributes are smaller-the-better [larger-the-better], we view them as inputs [outputs]. We use the geometric mean

Table 1 Performance data of robots

	Inputs		Outputs			Inputs		Outputs	
RT	REP	Cost	CAP	VEL	RT	REP	Cost	CAP	VEL
1	0.15	7.2	60.0	1.35	15	1.00	3.68	47.0	1.0
2	0.05	4.8	6.0	1.1	16	1.00	6.88	80.0	1.0
3	1.27	5.0	45.0	1.27	17	2.00	8.0	15.0	2.0
4	0.025	7.2	1.5	0.66	18	0.20	6.3	10.0	1.0
5	0.25	9.6	50.0	0.05	19	0.05	0.94	10.0	0.3
6	0.10	1.07	1.0	0.3	20	2.00	0.16	1.5	0.8
7	0.10	1.76	5.0	1.0	21	2.00	2.81	27.0	1.7
8	0.10	3.2	15.0	1.0	22	0.05	3.8	0.9	1.0
9	0.20	6.72	10.0	1.1	23	0.10	1.25	2.5	0.5
10	0.05	2.4	6.0	1.0	24	0.10	1.37	2.5	0.5
11	0.50	2.88	30.0	0.9	25	0.20	3.63	10.0	1.0
12	1.00	6.9	13.6	0.15	26	1.27	5.3	70.0	1.25
13	0.05	3.2	10.0	1.2	27	2.03	4.0	205.0	0.75
14	0.05	4.0	30.0	1.2					

to combine two inputs [outputs] variables into a composite input [output] variable. The performance is evaluated by the ratio between the composite output and composite input (OIR). The results are shown in the 2nd and 6th column of Table 2.

We classify the robots based on the OIR values using the proposed approach. The plot of SNR has been displayed in Fig. 1, which indicates that the robots can be divided into three classes: A, B and C. The robots in class A are Robots 27, 14 and 1.

Table 2 Parameters of the transformations for criteria scores

RT	OIR	SNR	Class	RT	OIR	SNR	Class
27	8.482	2.489	A	7	3.452	1.879	C
14	8.190	2.760	A	25	3.426	1.839	C
1	7.598	2.933	A	21	3.375	1.833	C
19	6.792	2.854	B	18	2.985	1.779	C
5	5.681	2.400	B	9	2.937	1.758	C
16	5.523	2.215	B	23	2.659	1.717	C
26	5.194	2.071	B	24	2.599	1.706	C
8	5.149	2.024	B	12	2.275	1.665	C
11	5.000	2.009	B	17	1.936	1.584	C
13	5.000	2.059	B	4	1.880	1.529	C
15	4.950	2.162	B	6	1.748	1.481	C
3	4.226	2.107	C	20	1.628	1.451	C
10	4.162	2.099	C	22	1.437		C
2	3.500	1.963	C				

We now validate this classification using the normal kernel density estimation. The upper and lower of the bandwidth are 1.1006 and 0.2818, respectively. Their average is 0.6912. Using this value to Eq. (9), we obtained the empirical pdf shown in Fig. 2. As seen from the figure, the empirical pdf is smooth and has three modes, implying that $K = 3$. The local minima are at $s = 4.28$ and 6.97. According to these two boundary points, we can obtain fully the same classification as the one obtained from the SNR-based approach and hence confirm the appropriateness of the proposed approach.

3.2 Example 2

The data shown in Table 3 come from Ref. [9] and deal with the criteria scores of 47 inventory items. Criterion C_1 is annual dollar usage, criterion C_2 is average unit cost, criterion C_3 is critical factor, and criterion C_4 is lead time. All these criteria are positively related to the importance of inventory items.

Table 3 Data for Example 2

i	C_1	C_2	C_3	C_4	i	C_1	C_2	C_3	C_4
1	5840.64	49.92	1	2	25	370.5	37.05	0.01	1
2	5670	210	1	5	26	338.4	33.84	0.01	3
3	5037.12	23.76	1	4	27	336.12	84.03	0.01	1
4	4769.56	27.73	0.01	1	28	313.6	78.4	0.01	6
5	3478.8	57.98	0.5	3	29	268.68	134.34	0.01	7
6	2936.67	31.24	0.5	3	30	224	56	0.01	1
7	2820	28.2	0.5	3	31	216	72	0.5	5
8	2640	55	0.01	4	32	212.08	53.02	1	2
9	2423.52	73.44	1	6	33	197.92	49.48	0.01	5
10	2407.5	160.5	0.5	4	34	190.89	7.07	0.01	7
11	1075.2	5.12	1	2	35	181.8	60.6	0.01	3
12	1043.5	20.87	0.5	5	36	163.28	40.82	0.5	3
13	1038	86.5	1	7	37	150	30	0.01	5
14	883.2	110.4	0.5	5	38	134.8	67.4	0.5	3
15	854.4	71.2	1	3	39	119.2	59.6	0.01	5
16	810	45	0.5	3	40	103.36	51.68	0.01	6
17	703.68	14.66	0.5	4	41	79.2	19.8	0.01	2
18	594	49.5	0.5	6	42	75.4	37.7	0.01	2
19	570	47.5	0.5	5	43	59.78	29.89	0.01	5
20	467.6	58.45	0.5	4	44	48.3	48.3	0.01	3
21	463.6	24.4	1	4	45	34.4	34.4	0.01	7
22	455	65	0.5	4	46	28.8	28.8	0.01	3
23	432.5	86.5	1	4	47	25.38	8.46	0.01	5
24	398.4	33.2	1	3					

For a given inventory item, there are many approaches to combine its criteria scores into an overall performance score (OPS). Discussing these approaches is beyond the scope of this paper since our focus is on the natural classification of the items based on their OPS. We use the weighted arithmetic mean to combine the criteria scores

$$s_i = \sum_{j=1}^{4} w_j \frac{x_{ij} - m_j}{M_j - m_j}, m_j = \min(x_{ij}, 1 \le i \le n), M_j = \max(x_{ij}, 1 \le i \le n) \qquad (14)$$

where w_j is the weight of the jth criterion. Reference [9] uses the AHP to determine the criterion weights, which are 0.091, 0.079, 0.420 and 0.410, respectively. Using this set of weights to Eq. (14) we obtained OPS of the items shown in the 2nd and 6th columns of Table 4.

Using the proposed clustering approach we obtained the SNRs shown in the 3rd and 7th columns of Table 4. Figure 3 shows the plot of the SNR. According to the

Table 4 Parameters of the transformations for criteria scores for Example 2

Item	OPS	SNR	Class	Item	OPS	SNR	Class
13	0.8772	2.3969	A	6	0.4002	1.8048	C
2	0.8607	2.5350	A	7	0.3972	1.8313	C
9	0.8255	2.6352	A	28	0.3744	1.8450	C
3	0.7106	2.4213	B	16	0.3722	1.8719	C
23	0.6628	2.2460	B	38	0.3703	1.9147	C
21	0.6393	2.1393	B	40	0.3608	1.9657	C
1	0.5966	2.0179	B	36	0.3605	2.0429	C
15	0.5951	1.9632	B	39	0.2958	2.0409	C
18	0.5756	1.9166	B	33	0.2931	2.0575	C
24	0.5733	1.9008	B	37	0.2849	2.0862	C
14	0.5352	1.8532	C	43	0.2834	2.1444	C
10	0.5101	1.7974	C	47	0.2746	2.2271	C
31	0.5100	1.7672	C	8	0.2651	2.3458	C
32	0.5097	1.7562	C	35	0.1605	2.2687	D
19	0.5061	1.7571	C	44	0.1537	2.2128	D
11	0.5048	1.7714	C	26	0.1526	2.1923	D
12	0.5032	1.7983	C	46	0.1459	2.2028	D
29	0.4636	1.7917	C	4	0.0830	2.1144	D
22	0.4427	1.7747	C	42	0.0817	2.0604	D
20	0.4404	1.7694	C	41	0.0748	2.0393	D
17	0.4272	1.7624	C	27	0.0353	1.9604	D
45	0.4214	1.7621	C	30	0.0227	1.8761	D
5	0.4190	1.7722	C	25	0.0177		D
34	0.4133	1.7893	C				

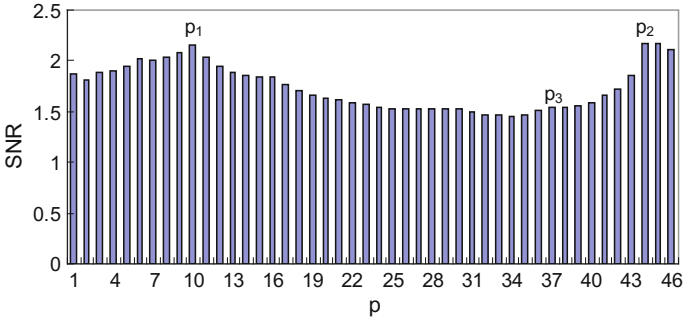


Fig. 3 Plot of SNR(p) versus p for example 2

figure, the items can be roughly classified into three classes with boundary points $p_1 = 10$ and $p_2 = 44$. This implies that the three items with $s > 0.77$ can be classified as class A and the ten items with $s < 0.21$ can be classified as class D. The other 34 items need to be further classified so that the final classification roughly follows the Pareto principle. From Fig. 3, we can find that another local maximum is at $p_3 = 37$. As such, we have $K = 4$ and the final classification of the inventory items is shown in the 4th and last columns of Table 4.

We now use the normal kernel density estimation to confirm the above classification. The upper and lower of the bandwidth are 0.1058 and 0.0191, respectively. Their average is 0.0624. Using this value to Eq. (9), we obtained the empirical pdf shown in Fig. 4. As seen from the figure, the empirical pdf is smooth and has three modes, implying that $K = 3$. The local minima are at $s = 0.1885$ and 0.7692. According to these two boundary points, we can obtain fully the same classification obtained before the fine tuning of classification. This confirms, once more, the appropriateness of the proposed approach. In the meantime, it also confirms the appropriateness of the proposed bandwidth selection approach.

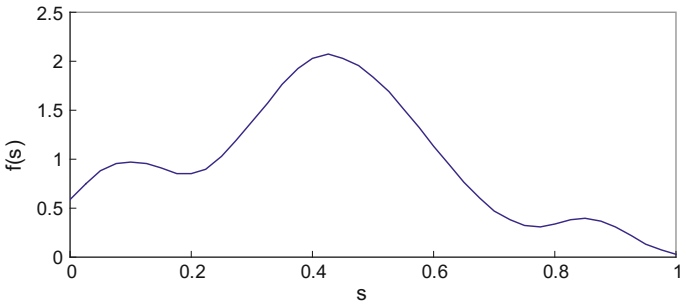


Fig. 4 Normal kernel density estimation for example 2

Compared with the kernel density estimation, the proposed approach does not need to select a smoothing parameter and provides richer information about the number of clusters and their boundaries.

4 Conclusions

In this paper, we have proposed a SNR-based approach for one-dimensional data cluster analysis. The proposed approach combines between-group distance and within-group variances into the SNR, whose plot provides intuitive information about the number of clusters and their boundaries. It does not need an iterative process and can be conveniently implemented using an Excel spreadsheet program. The proposed approach has been illustrated by two real-world examples, and the results have been validated through comparing them with the results from the kernel density estimation.

In the meantime, a simple method for selecting the bandwidth parameter in the kernel density estimation has been proposed. Its appropriateness has also been confirmed by the two examples. It has been found that the SNR-based approach provides richer information about the number of clusters than the kernel density estimation. It can be advantageous to jointly use the SNR-based approach and the kernel density estimation for the one-dimensional data cluster analysis problems.

Acknowledgements The research was supported by the National Natural Science Foundation of China (No. 71771029).

References

1. Jain AK (2008) Data clustering: 50 years beyond K-means. Pattern Recogn Lett 31(8):651–666
2. Jiang R (2009) Cluster analysis of maintenance management problems. In: 2009 IEEM. Hong Kong, pp 1150–1154
3. Mandelli D, Yilmaz A, Aldemir T, Metzroth K, Denning R (2013) Scenario clustering and dynamic probabilistic risk assessment. Reliab Eng Syst Saf 115(1):146–160
4. Murthy CA (2016) On bandwidth selection using minimal spanning tree for kernel density estimation. Comput Stat Data Anal 102:67–84
5. Silverman BW (1986) Density estimation for statistics and data analysis. Chapman & Hall, London
6. Wikipedia. Kernel density estimation. https://en.wikipedia.org/wiki/Kernel_density_estimation, last accessed 14 Dec 2015
7. Celik P, Wu ML (1999) Decision-making and performance measurement models with applications to robot selection. Comput Ind Eng 36(3):503–523
8. Mecit ED, Alp I (2013) A new proposed model of restricted data envelopment analysis by correlation coefficients. Appl Math Model 37(5):3407–3425
9. Flores BE, Olson DL, Dorai VK (1992) Management of multi-criteria inventory classification. Math Comput Model 16(12):71–82

Multi-AGVs Conflict-Free Routing and Dynamic Dispatching Strategies for Automated Warehouses

Xiaowen Li, Canrong Zhang, Wenming Yang and Mingyao Qi

Abstract Automated Guided Vehicle System (AGVS) has been playing a vital role in many applications such as flexible manufacturing systems and automated warehouses. Efficient vehicle routing and dispatching are two important aspects when designing an AGVS. Based on zone control mode, this paper proposes a conflict-free routing strategy to improve the system efficiency with the ability to avoid loops. A dynamic dispatching strategy is proposed which can find a suitable AGV to conduct the task to decrease the total traveling distance and the waiting time. Simulations are carried out to verify the effectiveness of the strategies.

Keywords Conflict-free · Routing strategy · Dynamic dispatching strategy

1 Introduction

Recently, with rapid development of e-commerce, efficiency and accuracy in storage and retrieval of goods in the warehouses have been becoming more and more important. Various forms of automation have been used in warehouses for several years, but the demands for e-commerce are growing such that automation is now an absolute prerequisite and Automated Guided Vehicles (AGVs) are starting to play a vital role [1].

AGVs are advanced material-handling robots used for repeating transportation tasks in many areas, such as warehousing and manufacturing system.

X. Li
Department of Industrial Engineering, Tsinghua University, Beijing, China

C. Zhang · M. Qi (✉)
Modern Logistics Research Center, Graduate School at Shenzhen, Tsinghua University, Shenzhen, China
e-mail: qimy@sz.tsinghua.edu.cn

W. Yang
Shenzhen Key Laboratory of Information Sci&Tech, Graduate School at Shenzhen, Tsinghua University, Shenzhen, China

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_26

Conflict avoidance must be addressed during the operation of AGVS. In this paper, we focus on designing a conflict-free routing strategy based on a zone control mode. A route indicates a sequence of zones, which should be taken by the AGV to make a pickup or delivery [2]. Zone control refers to dividing the traffic area into non-overlapping areas, which can resolve the conflict directly by only allowing one vehicle to occupy a zone at each time.

Typically, there are two ways to deal with the conflict-free routing problem in AGVS: pre-planning method and dynamic method [3]. The AGV's route of the former category is pre-planned assuming that there will be no interruption on route and all the AGVs have a same and equal speed [4]. A primary disadvantage of such method is that it is based on static Gantt charts with ideal assumptions. Compared to the pre-planning method, the dynamic method avoids conflict during the operation and does not plan the AGV route [5]. Nevertheless, the drawback of the dynamic routing method is clear that the performance of the system is not optimized.

In this paper, we study the conflict-free vehicle routing problem with the following characteristics. First, we adopt a strategy to efficiently find a route which avoids conflict and loop which runs in polynomial time. Second, multiple route choices are provided to improve the performance of the system. Third, we ensure that the vehicle is non-preemptive by varying task release time, namely, once the vehicle starts to work, it cannot stop until finishing its task. Furthermore, a dispatching strategy is proposed to select a suitable AGV to conduct the upcoming task which promotes the system performance. The strategy proposed here attributes itself to the pre-planning method and is suitable for automated warehouses with multiple AGVs, which can move smoothly and rarely have interactions with humans to obtain higher efficiency.

The rest of the paper is organized as follows. Section 2 describes and models the system. Sections 3 and 4 propose the routing and dynamic dispatching strategies. Section 5 presents the simulation results. Section 6 concludes the paper.

2 System Description and Modeling

2.1 Central Control System

Similar to [6], we adopt a central control system, where there is a central controller to supervise all the vehicles. However, the traffic area is divided into non-overlapping zones rather than lanes and points. Instead of getting the permission from the central controller before every movement, the AGV only needs to get the permission of a route before starting to execute the task. It can move without stopping until it reaches the ending zone, which largely reduces the communication overhead between the controller and the AGV and the delay caused by deadlock recovery or conflict avoidance.

The control system consists of three modules: communication, dispatching and routing. As a new task arrives, the dispatching module selects a suitable AGV to

conduct the task based on the dispatching strategy and on the states of the AGVs provided by the communication module. Then, the dispatching module sends the information of the task and the code of the selected AGV to routing module. The routing module gets multiple route choices for the AGV and then picks the best route which avoids undesired situations (conflict or loop) according to the routing strategy, the AGVs states and the state of the warehouse. The communication module receives the route, converts it to a sequence of zones and then sends the sequence to the selected AGV. The AGV moves according to the sequence and occupies each zone at a time in strict order. We assume that in this kind of system, all vehicles have the same and equal speed, and the time for each vehicle to move from one zone to another is constant and unchanged.

2.2 Warehouse Modeling

We divide the warehouse into non-overlapping zones, and one zone can only accommodate one vehicle at a time. To exclude the influence of idle AGV to traffic management, we assume that there is a parking area for idle AGVs with infinite capacity.

Definition 1 Zone z. The smallest unit of area in the warehouse. z_0 represents the parking area in the warehouse. Z represents the zone set of the warehouse.

Definition 2 Capacity c_i. The number of AGV that a zone z_i can accommodate at each time. Since one zone except for the parking area can only accommodate one vehicle at each time, $c_i = 1$, $i \neq 0$ and the parking area has an infinite capacity.

Definition 3 Zone direction $d_i = (up, down, right, left)$. The direction by which the vehicle can move in zone z_i can be used to identify the border of the warehouse. If $up = 1$, then an AGV can travel to the upper zone, while $up = 0$ means that no AGV can travel that way.

Definition 4 Zone state s_i^z. The state of zone z_i, which could be occupied or vacant. The state of zone z_i is vacant if $s_i^z = 0$, and $s_i^z = 1$ refers to occupied. Once the AGV enters the zone, the zone state becomes occupied immediately. After the AGV leaves the zone, the state changes from occupied to vacant.

Definition 5 AGV v_i. Each AGV has a unique AGV code i.

Definition 6 Route procedure rp_i. A route procedure consists of a sequence of zones in a strict order generated by the route scheduler. $rp_i[n]$ represents the (n − 1)th element in the sequence. For example, for $rp_i = (z_1, z_3, z_2)$, $rp_i[0] = z_1$, $rp_i[1] = z_3$ and $rp_i[2] = z_2$.

Definition 7 AGV state s_i^V. The AGV state takes the value of 0, 1, and 2 respectively for waiting, working and idle state. AGV is in the idle state if it has no task to conduct. If the AGV is moving from one zone to another according to its

route procedure, it is in the working state. If the AGV receives the route procedure, but it has to wait due to some restrictions, it is in the waiting state. Finally, the state of the warehouse can be represented as $S = \{(s_i^z, s_i^V) | s_i^z \in S^Z, s_i^V \in S^V\}$, where S^Z and S^V are sets of states of zones and AGVs in the system.

Definition 8 A task t can be represented by a pair of zones in order: $t = (z_i, z_j)$. The former zone z_i is the starting zone of task t and the latter zone z_j is the ending zone. When AGV v_i is selected to conduct task t, the path scheduler would first generate a route from the current location of v_i to starting zone z_i of task t, and then generate a route from z_i to the ending zone z_j.

3 Routing Strategy

In this section, we present a routing strategy that can efficiently find a route without conflict or loop by following some restrictions and rules. To get into the strategy, first we introduce some related notations.

Definition 8 Residual route procedure rrp_i. rrp_i represents the residual sequence of zones that v_i has still to visit before finishing the task. $rrp_i[n]$ is the (n − 1)th element of rrp_i. It should be noted that $rrp_i[0]$ is the current location of v_i, and if $s_i^V = 2$ we have $rrp_i = rp_i$.

Definition 9 Waiting time w_i. The waiting time of a vehicle in the working or idle state is 0. Waiting time represents the time that the vehicle has to wait at its current location before it starts to move according to rp_i in order to avoid conflict and loop.

3.1 Conflict and Loop Detection

In our system, the conflict means that two vehicles occupy the same zone at a time. Formally, a newly entered vehicle v_i is in conflict with existing AGVs if rp_i satisfies one of the following conditions:

1. $\exists \mathrm{j}, n \leq \min(card(rp_i), card(rrp_j))\text{s.t.}\, s_j^V = 1, rp_i[n] = rrp_j[n]$
2. $\exists \mathrm{j}, n \leq \min(card(rp_i), card(rrp_j))\text{s.t.}\, s_j^V = 0, rp_i[n] = rp_j[\max(0, n - w_j)]$

The former condition is to test whether v_i will crash with a vehicle v_j in the working state by judging if there exists a zone that v_i and v_j are occupying at the same time. Condition 2 indicates that v_i cannot pass through the current location of v_j when v_j is in the waiting state $(n \leq w_j)$.

Further, we need to avoid another situation called loop. A loop is a situation that for a set of AGVs, each AGV is going to occupy the location currently taken by another vehicle. In Fig. 1, $rrp_1[0] = z_1$, $rrp_2[0] = z_2$, $rrp_3[0] = z_3$, $rrp_4[0] = z_4$. After one unit of time $rrp_1[0] = z_2$, $rrp_2[0] = z_3$, $rrp_3[0] = z_4$, $rrp_4[0] = z_1$. This kind of situation could also be referred as deadlock [6]. Unfortunately, loop cannot be automatically avoided by the aforementioned conflict identification method. To address this problem, we adopt the deadlock detection algorithm proposed in [6].

We keep the route valid and avoid conflict and loop by allowing or postponing the moving of the AGV while a new route is assigned to it. It is evident that the rp_i of v_i could always be valid as long as the AGV is postponed for a long enough time since all the other AGVs would leave the system in a finite time. Here we propose a strategy to calculate w_i, the waiting time of v_i, based on rp_i to make the route valid.

```
Function: delay_time(v,rp)
int w = 0
while (judge_conflict(v,rp,w) || judge_loop(v,rp,w))
  w++
return w
```

Function `judge_conflict` and `judge_loop` are based on the conflict detection and loop detection rules. We emit their implementations for brevity. The calculated waiting time of the task makes sure that the vehicle is non-preemptive, namely, once the vehicle starts to work, it cannot stop until finishing its task.

3.2 Routing Strategy

A straightforward approach to choose rp_i of v_i is to calculate the shortest path from the beginning zone to the ending zone. However, it may take a long time for v_i to wait until the shortest path become valid, i.e., conflict-free and loop-free. In this section, we propose a routing strategy which adopts the k-th shortest path algorithm [7] to get an optimal route and the corresponding waiting time.

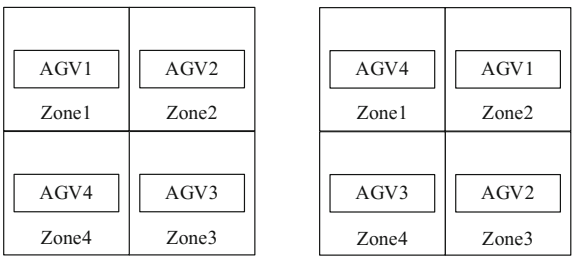


Fig. 1 An example of loop

We define RP_i as the set of routes generated by the k-th shortest path algorithm for v_i. $RP_i[n]$ refers to the (n − 1)th element of RP_i, which is also the (n − 1)th shortest path.

```
Function: Routing_stratrgy(v,RP)
w=delay_time(v,RP[0])
rp=RP[0]
for n from 1 to k-2
    if card(RP[n])+delay_time(v,RP[n])<w+card(rp)
    w=delay_time(v,RP[n])
    rp=RP[n]
return w and rp
```

The routing strategy takes into account the influence of both the length of the route (rp_i) and the waiting time (w_i), and multiple route choices are provided to improve the performance of the system.

4 Dynamic Dispatching Strategy

A naïve dispatching strategy immediately assigns a new task to the AGV which is the first one to become idle [6]. We refer to it as the immediate dispatching strategy. However, this kind of strategy ignores the distance between the current location of the AGV and the starting zone of the task. Sometimes, an idle AGV needs to go through the whole warehouse to get to the starting zone of the task, which is time-consuming. Based on the static pre-planning method, we assign the new task to either an AGV in the idle state or an AGV which nearly finishes its current task and is nearer to the starting zone of the new task. This dynamic dispatching strategy largely decreases the total distance that the AGVs need to travel to finish all the tasks. The algorithm of this dispatching strategy is explained as follows. First we select an idle state AGV which has the shortest distance from its current zone and the starting zone of the task. We record the distance as the *mindistance* and the corresponding AGV as the *dotask_AGV*. Then we select a working state AGV which has the minimum sum of its corresponding card(*rp*) and the distance from the current ending zone to the starting zone of the new task. It should be noted that the corresponding card(*rp*) and the number of the rest zones that the AGV still needs to travel describe how soon this AGV can finish its current task. We compare the sum to the *mindistance*, select the smaller one as the *mindistance* and update the *dotask_AGV*.

5 Numerical Experiments

To evaluate the performance of different routing and dispatching strategies, we implement a simulation platform to simulate the behaviors of AGVs and analyze how various factors influence the system performance.

5.1 Simulation Setup and Performance Metrics

To eliminate the impact of idle AGVs on the traffic system, when an AGV is not executing any tasks, the controller sends instructions guiding it to the parking area. In each experiment, a series of picking tasks are generated one by one sequentially and sent to the central controller. A task is considered to be completed if the AGV arrives at the ending zone of the task. The experiment completes after AGVs have finished all the tasks. The simulation is conducted on a Java-based platform with the input of warehouse topology, the number of task, the starting zone and ending zone of each task and the number of AGVs.

The topology of the warehouse is shown in Fig. 2. Each zone of the warehouse is designed as a square with a length of 2 m and the length of AGV is set as 1.5 m. The red line represents the direction of the zone.

During the simulation, we use three metrics to evaluate the system performance. (1) Total traveling distance $\mathrm{L} = \sum_{i=1}^{T_n} l_i$, where l_i is the AGV travel distance for task i and T_n is the number of tasks. (2) Total traveling time $\tau = T_s - T_e$, where T_s is the starting time of the first task and T_e is the ending time of the last task. (3) Total delay $\mathrm{D} = \sum_{i=1}^{T_n} w_i$, where w_i is the AGV's waiting time for task i.

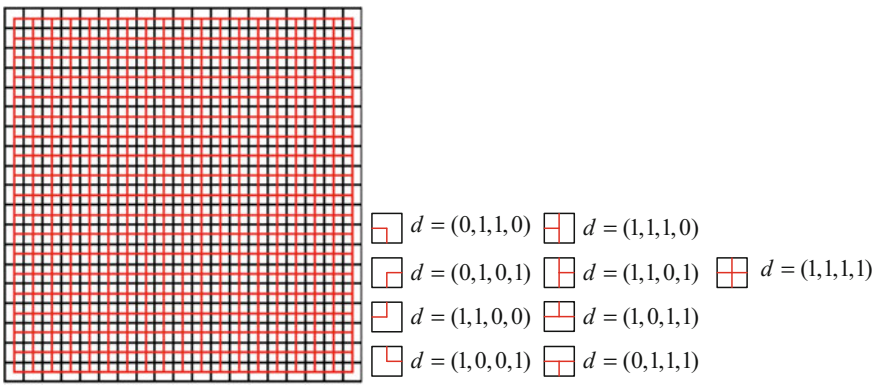


Fig. 2 Topology of the warehouse

5.2 Results

We adopt the k-th shortest path algorithm to provide multiple route choices; therefore the choice of k would affect the system performance significantly. In this section, we aim to set a reasonable value of k. A too large k increases the complexity and running time of the algorithm, resulting in unnecessary waste. But if k is too small, more AGVs have to wait for a long time, and the results may not be optimal. Intuitively, a heavy-loaded system would require more route choices since many AGVs may compete for a zone at the same time. Therefore, we first expose the system to maximum load conditions, with 160 tasks and 40 vehicles to get a reasonable k for all the vehicles.

Figure 3a shows that the increase of k would considerably decrease the total delay time of the system, and when k comes to 10, the system performance is almost stable. Besides, compared to the immediate dispatching strategy, the dynamic dispatching strategy we proposed shows better system performance and becomes stable when k equals to 10. It should be noted that, as k increases, the total delay time of the system may have a small range of fluctuations, and the reason is that the different route choices of the former vehicles could affect the waiting time of the latter vehicles. In the following experiments, we set k as 10 for all the vehicles, which means RP is the set of routes generated by finding the top 10^{th} shortest paths.

Figure 3b, c display the relationship between the total traveling time and the number of tasks using two dispatching strategies. In these experiments, we set the number of vehicles as 20. When the number of AGVs equals to the number of tasks in the system, two strategies have the same system performances, which is obvious since one AGV only has to execute one task. However, as the number of task increases, the total traveling time and the total traveling distance increase for both

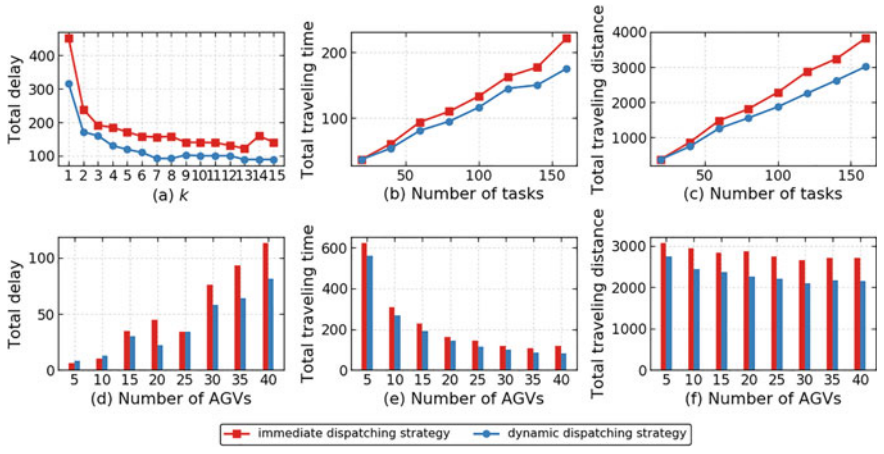


Fig. 3 System performance with different settings and strategies

dispatching strategies, whereas the dynamic dispatching strategy shows better system performance with less total traveling time distance.

To illustrate how the number of AGVs affects system performance, we fix the number of tasks as 120 and vary the number of AGVs from 5 to 40. Figure 3e shows that increasing the number of AGVs will considerably decrease the total traveling time. Nevertheless, once the number of AGVs reaches a specific value, the total traveling time will not change rapidly. The reason is that there are abundant AGVs to complete all the tasks, and adding more vehicles to the system is a waste of resources and even leads to large-scale congestion. Figure 3d shows that in general, the total delay time would increase if we increase the number of AGVs with the same tasks, since more AGVs in the system would make the system more congested and the waiting time of AGV becomes longer. Figure 3f displays that the total traveling distance is mainly related to the number of tasks, which does not change largely with the number of AGVs in the system.

6 Conclusion

In this paper, we focus on the vehicle routing and dispatching strategy in the AGVS of automated warehouse. We propose a conflict-free routing strategy, which can find a route for the AGV without conflict or loop. We propose a dynamic dispatching strategy to select a more suitable AGV to conduct the task to decrease the total AGV traveling distance. Both algorithms run in polynomial time and can be applied in large-scale systems. Simulations verify the effectiveness of the strategies. Since the proposed strategies are based on the static Gantt charts with ideal assumptions, further works will add real control as a second stage to improve the robustness of the system.

Acknowledgements This work is supported by the National Natural Science Foundation of China under grants 71772100 and 71472108, and Shenzhen Science and Technology Project under grants JCYJ20170412171044606 and JCYJ20160531195231085. We would like to thank the anonymous referees for their helpful comments.

References

1. Bogue R (2016) Growth in e-commerce boosts innovation in the warehouse robot market. Ind Rob Int J 43:583–587
2. Vis IFA (2006) Survey of research in the design and control of automated guided vehicle systems. Eur J Oper Res 170:677–709
3. Maza S, Castagna P (2005) A performance-based structural policy for conflict-free routing of bi-directional automated guided vehicles. Comput Ind 56:719–733
4. Kim CW, Tanchocoj JMA (1993) Operational control of a bidirectional automated guided vehicle system. Int J Prod Res 31:2123–2138

5. Li Q, Pogromsky A, Adriaansen T, Udding JT (2016) A control of collision and deadlock avoidance for automated guided vehicles with a fault-tolerance capability. Int J Adv Rob Syst 13:64
6. Yan X, Zhang C, Qi M (2017) Multi-AGVs collision-avoidance and deadlock-control for item-to-human automated warehouse. In: 2017 international conference on industrial engineering, management science and application. IEEE
7. Eppstein D (1994) Finding the k shortest paths. In: Proceeding of IEEE symposium on foundations of computer science

Patent Technology Networks and Technology Development Trends of Neuromorphic Systems

Shu-Hao Chang and Chin-Yuan Fan

Abstract Neuromorphic systems have been recognized by developed countries as the most promising research area in AI computing. However, previous studies on neuromorphic systems have mostly focused on technical details or specific devices or products, failing to actively indicate the technological focus and recent development trends. Therefore, this study used neuromorphic system patents to construct a technology network through patent technology network analysis. The results show that the technological focuses of neuromorphic systems are biological models; specific functions and applications of digital computing; and detection, measurement, and recording for diagnostic purposes. In addition, the development of medical diagnosis and measurement technology as well as equipment such as speech recognition and optical apparatuses has flourished in recent years. This study proposed a technological map of neuromorphic systems that can provide the government with valuable information for exploring development trends in this field.

Keywords Neuromorphic system · Technology network · Network analysis Technology trend · Patent analysis

1 Introduction

The emergence of artificial intelligence (AI) and the Internet of Things has increased the demand for mass data processing in recent years; smart machines will be closely integrated into our daily lives in the future. For mass data processing,

S.-H. Chang · C.-Y. Fan (✉)
Science & Technology Policy Research and Information Center, National Applied Research Laboratories, 14F., No. 106, Sec. 2, Heping E. Rd., Taipei 10636, Da'an District, Taiwan, China
e-mail: cyfan@narlabs.org.tw

S.-H. Chang
e-mail: shchang@narlabs.org.tw

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_27

the operational rules and transmission speed between a single central processing unit and the random access memory limit the overall efficiency and operating time and therefore cannot satisfy the actual need for instant application. Neuromorphic systems can overcome this problem. They use distributed artificial neural networks for parallel computing and learning to achieve AI computation. According to MarketsandMarkets [10], the neuromorphic computing market is expected to grow from USD 6.6 million in 2016 to USD 272.9 million by 2022, at a CAGR of 86.0%, demonstrating that this technology still has considerable room for improvement; its development has been studied by numerous researchers [11, 27].This study focused on the technology network of neuromorphic systems and constructed a technology network model through patent analysis. Patents are direct indicators of technology development trends [6, 26]. They have been used to identify emerging technological fields [9, 13] and track technological evolution [23]. Thus, patent data are a highly direct means of measuring technology development, and were used in this study to observe the direction of technology development and identify key technological fields. In contrast to previous studies that have investigated the technical aspects of neuromorphic systems and their specific applications, this study examined technological deployment and overall neuromorphic systems from a macro perspective. By focusing on the construction of the technology network model and technology development trends, this study can serve as a reference for the government, academia, and industries.

2 Literature Review

2.1 Current Status of Neuromorphic System Development

Neuromorphic systems are electronic implementations of neural systems [20]; they perform circuit simulation, circuit design, and engineering simulation of operation architecture, algorithms, and systems according to the composition of biological neural systems and its functions of signal transmission and operational memory, including synapses, neurons, spiking neural networks, and visual, auditory, and olfactory systems. Problems related to heat dissipation and power consumption have prevented improvement of the operating speed of the traditional central processing unit since 2004. Thus, development has focused on multicore processing. However, multiple cores cannot be started simultaneously because of the dark silicon problem related to heat dissipation. The brain-inspired neuromorphic engineering design provides an innovative direction for computing architecture (e.g., memristor) that can authentically simulate the function of neuron synapses and improve the spiking-neuron circuit. This architecture can be applied to highly complex computing problems and features low energy consumption. Biological neurons and synapses can provide a blueprint for inference and learning machines that are potentially 1000-fold more energy efficient than mainstream computers [12]. Currently, the problem of simulating simple functions of the human brain is

being successfully solved by programming tools through the formation of artificial neural networks, leading to the accumulation of experience and success in the use of artificial neural networks in different fields of human activity, such as pattern recognition (e.g., recognition of texts, images, speech, and music) [4].

The applications of neuromorphic systems might diversify in the future. Neuromorphic systems arguably offer significant advantages over conventional computers in several areas, such as sensory processing and autonomous systems. Other application domains where dedicated neural processing hardware systems are being used to complement or even replace conventional computers are in custom accelerated simulation engines for large-scale neural modeling and in very-large-scale spiking neural networks applied to machine learning problems. The output value of neuromorphic systems is promising [10]. Developed countries have paid considerable attention to the development potential of neuromorphic systems. Therefore, this study focused on these systems, using technology network analysis integrated with patent analysis to identify the technologies for applying them. The analysis method is introduced in the following section.

2.2 Patent Technology Network Analysis

Recent studies have performed network analysis to explore the structure and scope of the knowledge domain [23], track the development path of specific technologies [8, 28], and identify the trend and direction of technology development through patent analysis [6, 19, 26]. Network analysis can accurately present technological knowledge networks and information transmission paths [3, 7]. In addition, objective and practical information can be obtained through patent analysis, including number of patents, patent year, and technology category [6]. Therefore, using patent data to analyze technological domains is meaningful because patents are the sources of technological and business knowledge. The technologies described in patents are generally applied in specific industries. By identifying patents of specific technologies, researchers can understand the development of these technologies and identify potential areas for application [15]. Therefore, this study adopted technology network analysis to understand the key players in patents of neuromorphic systems and investigate the connections between various types of technology.

3 Research Design

3.1 Data Source and Retrieval Strategy

In this study, the data source was the United States Patent and Trademark Office (USPTO). The search criteria were as follows: [(TTL/neuromorphic and system) or (ABST/neuromorphic and system) or (ACLM/neuromorphic and system) or

(SPEC/neuromorphic and system)]. A total of 407 patents were retrieved. Regarding the technology classification, the USPTO and European Patent Office have implemented cooperative patent classification (CPC) since 2013. Thus, this study adopted CPC as the analysis framework.

3.2 *Key Player Analysis*

Through network analysis, researchers can define participants as stars or gatekeepers to determine key players [1]. Previous studies have used the network-centric analytic approach to identify key players in networks [21, 22]. This study likewise employed this approach to identify the key application technology in patent technology networks.

3.2.1 Closeness Centrality

The distance between nodes can be measured by closeness centrality, namely the sum of reciprocal distance between a node and all other nodes. In a technology network, closeness centrality can be used to measure the global centrality of a technology and determine its connections with other technologies.

$$\mathrm{C}_C(i) = \left[\sum_{j=1}^{n} d(i,j)\right]^{-1}$$

where $d(i,j)$ indicates the distance between node i and node j.

3.2.2 Eigenvector Centrality

Eigenvector centrality is another method of determining whether a node is a central hub. It considers not only the distance between the target node and other nodes but also whether the nodes connecting with the target node simultaneously connect with other nodes [24]. Adjacent nodes do not have equal connections with the target node. A high score in eigenvector centrality indicates that a technology is related to numerous technologies in the network, and the connected technologies are relatively important in the overall technology network. The contribution of adjacent technologies to the importance of the target technology is considered.

$$\mathrm{C}_e(i) = \lambda^{-1} \sum_{j=1}^{n} a_{ij} \mathrm{C}_e(j)$$

where $C_e(i)$ and $C_e(j)$ are the eigenvector centralities of node i and j, respectively; a_{ij} is the node in the adjacency matrix A; λ is a constant and is the maximum eigenvalue of the adjacency matrix A.

In this equation, the eigenvector centrality is a linear function obtained when the centrality of a single node is considered to be the linear combination of the centrality of all the other nodes [2].

3.2.3 Betweenness Centrality

Betweenness centrality refers to how some nodes depend on the target node (i.e., the mediator) to connect with other nodes in the network. It is a measure of a node's importance in transferring information across the broader context of the network. A technology with high betweenness centrality is located at a key position in the network; other technologies rely on it for connection and communication.

$$C_B(i) = \sum_{i \neq j \neq k}^{n} d_{jk}(i)/d_{jk}$$

where d_{jk} indicates the number of shortest paths from node j to node k; $d_{jk}(i)$ is the number of these shortest paths that passes through node i.

3.2.4 Fragmentation Centrality

Fragmentation centrality, which measures the extent to which a network is fragmented after a node is removed from the network, can also be used to measure the betweenness of nodes [1]. It examines the proportion of the nodes that cannot connect with each other in the network when the target node is removed. A small value indicates that the network remains stable after the node is removed; therefore, the node has low importance in the network. This study used distance-weighted fragmentation to measure fragmentation centrality according to Borgatti [1].

$$F_c(i) = 1 - \frac{2\sum_{i>j}\frac{1}{d(i,j)}}{n(n-1)}$$

where $d(i, j)$ is the distance between node i and j; n is the total number of nodes.

4 Results

4.1 Key Technology of Neuromorphic Systems

The patent retrieval results were analyzed before the technology network was analyzed, to obtain a preliminary understanding of the status of technology

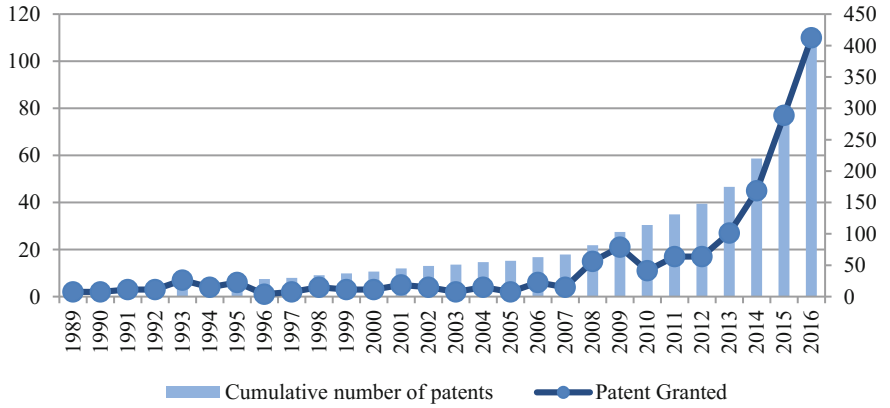


Fig. 1 Yearly change in the number of neuromorphic system patents granted

development. Figure 1 shows the yearly change in the number of neuromorphic system patents granted.

Figure 1 shows that the number of patents began to surge in 2013, indicating that neuromorphic system-related technologies have been developed extensively in recent years. According to the four-level CPC, the studied patents involved 202

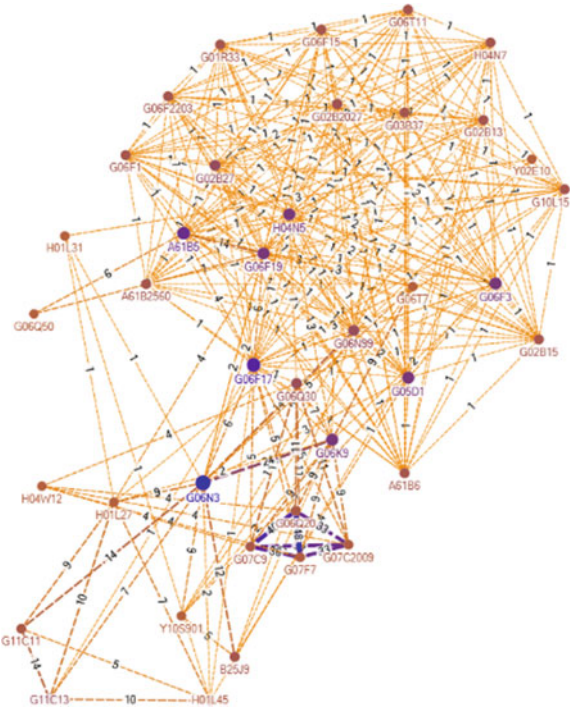


Fig. 2 Four-level CPC network of neuromorphic system patents. *Note* The size of a node indicates the number of other nodes it connects to. The thickness of a line indicates the strength of the connection. Only the nodes that connect to more than 15 other nodes are shown

Table 1 Top five CPC codes of neuromorphic system patents

CPC code	Closeness centrality	CPC code	Eigenvector centrality	CPC code	Betweenness centrality	CPC code	Fragmentation centrality
G06N3	115.333	G06F17	0.245	G06N3	7451.112	G06N3	0.71
G06F17	107.000	H04N5	0.227	G06F19	3588.018	A61B5	0.696
G06F19	101.833	G06F19	0.221	G06F17	2862.613	G06Q50	0.695
H04N5	101.667	G06F3	0.221	A61B5	2438.003	G06F19	0.692
G05D1	99.333	A61B5	0.219	G06Q50	2139.067	G01N33	0.685

CPC codes, on the basis of which a network model was constructed (Fig. 2). The key CPC codes are shown in Table 1.

According to Fig. 2 and Table 1, G06N3, G06F17, G06F19, and A61B5 were ranked in the top five in at least three of four indicators (i.e., closeness, eigenvector, betweenness, and fragmentation centrality). This signifies that the key players in the neuromorphic system patents were computer systems based on biological models (G06N3); digital computing or data processing equipment or methods, specially adapted for specific functions and applications (G06F17 G06F19); and detecting, measuring, or recording for diagnostic purposes (A61B5). Notably, G06Q50 was ranked in the top five for only betweenness centrality and fragmentation centrality. Because these two indicators denote the level of betweenness of a node in the network, G06Q50 can be regarded as the bridging technology between various types of technologies. According to the definition provided by CPC, G06Q50 is "systems or methods specially adapted for a specific business sector." This implies that the neuromorphic system applied in the business sector serves as the bridging technology in the entire technology network. In other words, between technology clusters, the application of neuromorphic systems to business operation, administration, and management is crucial.

4.2 Direction and Trends of Neuromorphic System Development

This study examined the patent technology network in different periods to understand the change in development focus over the years. The studied time was divided into earlier (2014 and before) and later (post–2014) periods. The technology development directions and trends were delineated according to the change in the centrality value for the CPC codes in the earlier and later periods (Table 2).

According to Table 2, the neuromorphic system patents after 2015 mostly sought to concretize neuromorphic systems, developing equipment and hardware such as input arrangements for transferring data to be processed into a form that computers were capable of handling (G06F3), relevant control or modulation systems (e.g., G05D1), and data processing equipment (e.g., G06F17). In addition,

Table 2 Development trends of neuromorphic system technology

Centrality indicators	Top five development focuses in the earlier period	Top five development focuses in the later period	New additions in technology field	Eliminations in technology field
Closeness centrality	G06N3, G06F17, G06F19, H04N5, G05D1	G06N3, G06F17, G05D1, H04N5, G06F3	G06F3	G06F19
Eigenvector centrality	G06F17, H04N5, G06F19, G06F3, A61B5	G06F17, G05D1, H04N5, G06F3, G06F19	G05D1	A61B5
Betweenness centrality	G06N3, G06F19, G06F17, A61B5, G06Q50	G06N3, G06F17, G05D1, G10L25, G06F3	G05D1, G10L25, G06F3	G06F19, A61B5, G06Q50
Fragmentation centrality	G06N3, A61B5, G06Q50, G06F19, G01N33	G06F17, G06N3, G10L25, A61B5, G02B27	G06F17, G10L25, G02B27	G06Q50, G06F19, G01N33

Note The earlier period (1989–2014) involved 220 patents and 390 CPC codes; the later period (2015–2016) involved 187 patents and 378 CPC codes

related technologies began to be applied in speech recognition (G10L25) and optical apparatuses (G02B27).

Postanalysis: yearly change in the key technology of neuromorphic systems

This study analyzed the yearly change in the number of patents that involved G06N3, G06F17, G06F19, and A61B5 to understand the technology trends of neuromorphic systems. The results are shown in Fig. 3.

According to Fig. 2, the development of computer systems based on biological models (G06N3) rapidly increased, and technology for detecting, measuring, or recording for diagnostic purposes (A61B5) also received much attention. By comparison, other application technologies were developed relatively slowly; the number of related patents in recent years was low.

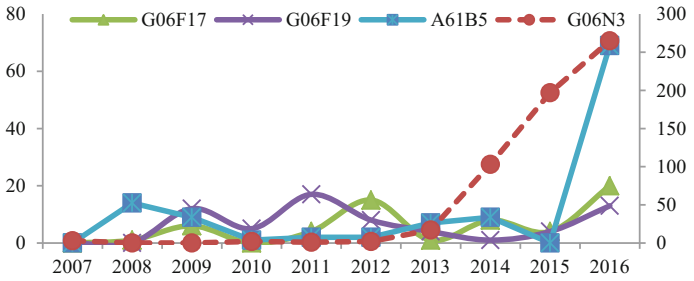


Fig. 3 Yearly change in the key technologies of neuromorphic systems. *Note* Because the number of G06N3 (the dotted line) is relatively large, it is indicated on the right-side axis

5 Conclusion

This study explored the key technologies of neuromorphic systems through network analysis. According to the empirical results, the key technologies are biological models; specific functions and applications of digital computing; and detection, measurement, and recording for diagnostic purposes. The third group, diagnostic technologies, showed particularly fast development in recent years and can be regarded as an active research area. Regarding recent technology trends, this study found it to be the concretization of neuromorphic systems, in forms such as neuromorphic chips, which will play an important role in signal processing, data processing, and image recognition in the future. In the post-CMOS era, neuromorphic chips may change the method of pattern recognition for big data and become a game changer, especially in pharmaceutical and bioinformatics applications [17].

This study also found that neuromorphic system-related technologies are being applied in equipment such as speech recognition and optical apparatuses. Previous studies have noted that signal recognition applications, which have been developed based on neuromorphic systems, are expected to hold the largest market share. For example, speech recognition is used in the field of healthcare by radiologists conducting ultrasonography [14]. Therefore, neuromorphic system technologies can be applied in image recognition, speech recognition, and data mining. Neuromorphic hardware is currently in the technology trigger stage and is expected to be adopted by mainstream industries after 10 years [5].

From the perspective of theoretical contributions, previous studies on neuromorphic systems have mostly investigated technical problems such as performance improvement [16, 25, 27] and focused on specific products or devices [11, 17, 18]. However, they have failed to identify the major focus of neuromorphic systems in the technological field, technology development trends, and technology distribution and networks. The present study filled this research gap, adopting a new perspective to observe technological fields.

As for policy recommendations, this study created a technological map of neuromorphic systems that can provide the government with valuable information. The technology network will enable researchers to understand the focus and trends of technology development and provides the government with information related to AI and smart life, answering questions about application directions and technological fields for neuromorphic systems. This study found that the primary technology hotspots in recent years are related to diagnosis and measurement in the medical field. Moreover, brain science has become one of the most crucial science frontiers; brain computing and brain intelligence simulation are policy priorities in developed countries. Studying the computing mode and application of neuromorphic systems is critical to achieving the goal of smart living. However, because research on neuromorphic systems is still in the preliminary stage, government should play a crucial role in R&D of this technology.

References

1. Borgatti SP (2006) Identifying sets of key players in a social network. Comput Math Org Theory 12(1):21–34
2. Borgatti SP, Everett MG (2006) A graph-theoretic perspective on centrality. Soc Netw 28(4): 466–484
3. Chen Z, Guan J (2016) Measuring knowledge persistence: a genetic approach to patent citation networks. R&D Manage 46(1):62–79
4. Demin V, Emelyanov A, Lapkin D, Erokhin V, Kashkarov P, Kovalchuk M (2016) Neuromorphic elements and systems as the basis for the physical implementation of artificial intelligence technologies. Crystallogr Rep 61(6):992–1001
5. Garnter (2016) Gartner's 2016 hype cycle for emerging technologies identifies three key trends that organizations must track to gain competitive advantage. Garnter, Stamford, CT
6. Gwak JH, Sohn SY (2017) Identifying the trends in wound-healing patents for successful investment strategies. PLoS One 12(3):1–19
7. Huenteler J, Ossenbrink J, Schmidt TS, Hoffmann VH (2016) How a product's design hierarchy shapes the evolution of technological knowledge-evidence from patent-citation networks in wind power. Res Policy 45(6):1195–1217
8. Kim M, Park Y, Yoon J (2016) Generating patent development maps for technology monitoring using semantic patent-topic analysis. Comput Ind Eng 98:289–299
9. Kreuchauff F, Korzinov V (2017) A patent search strategy based on machine learning for the emerging field of service robotics. Scientometrics 111(2):743–772
10. MarketsandMarkets (2016) Neuromorphic computing market by offering (hardware, software), application (image recognition, signal recognition, data mining), industry (aerospace & defense, IT & telecom, automotive, medical & industrial) and geography-global forecast to 2022. MarketsandMarkets, Seattle, WA
11. Moon K, Kwak M, Park J, Lee D, Hwang H (2017) Improved conductance linearity and conductance ratio of 1T2R synapse device for neuromorphic systems. IEEE Electron Device Lett 38(8):1023–1026
12. Neftci EO, Augustine C, Paul S, Detorakis G (2017) Event-driven random back-propagation: enabling neuromorphic deep learning machines. Frontiers Neurosci 11:1–18
13. Noh H, Song YK, Lee S (2016) Identifying emerging core technologies for the future: case study of patents published by leading telecommunication organizations. Telecommun Policy 40(10–11):956–970
14. OBRC (2017) Global neuromorphic chip market insights, opportunity analysis, market shares and forecast, 2017–2023. Occams Business Research & Consultancy, Mumbai
15. Park H, Yoon J, Kim K (2013) Using function-based patent analysis to identify potential application areas of technology for technology transfer. Expert Syst Appl 40(13):5260–5265
16. Partzsch J, Schüffny R (2015) Network-driven design principles for neuromorphic systems. Frontiers Neurosci 9:1–14
17. Pastur-Romay LA, Cedrón F, Pazos A, Porto-Pazos AB (2016) Deep artificial neural networks and neuromorphic chips for big data analysis: pharmaceutical and bioinformatics applications. Int J Mol Sci 17(8):1–26
18. Rafiue MA, Lee BG, Jeon M (2016) Hybrid neuromorphic system for automatic speech recognition. Electron Lett 52(17):1428–1429
19. Shin J, Lee CY, Kim H (2016) Technology and demand forecasting for carbon capture and storage technology in South Korea. Energy Policy 98:1–11
20. Smith LS (2010) Neuromorphic systems: past, present and future. Adv Exp Med Biol 657:167–182
21. Soon C, Cho H (2011) Flows of relations and communication among singapore political bloggers and organizations: the networked public sphere approach. J Inf Technol Politics 8(1):93–109

22. Swar B, Khan GF (2013) An analysis of the information technology outsourcing domain: a social network and triple helix approach. J Am Soc Inform Sci Technol 64(11):2366–2378
23. Trappey AJC, Trappey CV, Lee KLC (2017) Tracing the evolution of biomedical 3D printing technology using ontology-based patent concept analysis. Technol Anal Strateg Manag 29(4): 339–352
24. Wang C, Rodan S, Fruin M, Xu X (2014) Knowledge networks, collaboration networks, and exploratory innovation. Acad Manag J 57(2):454–514
25. Woo J, Moon K, Song J, Kwak M, Park J, Hwang H (2016) Optimized programming scheme enabling linear potentiation in filamentary HfO_2 RRAM synapse for neuromorphic systems. IEEE Trans Electron Devices 63(12):5064–5067
26. You H, Li M, Hipel K, Jiang J, Ge B, Duan H (2017) Development trend forecasting for coherent light generator technology based on patent citation network analysis. Scientometrics 111(1):297–315
27. Zhang P, Li C, Huang T, Chen L, Chen Y (2017) Forgetting memristor based neuromorphic system for pattern training and recognition. Neurocomputing 222:47–53
28. Zhou X, Zhang Y, Porter A, Guo Y, Zhu D (2014) A patent analysis method to trace technology evolutionary pathways. Scientometrics 100(3):705–721

Future Oriented Planning of Product-Service Systems

Dominik Weidmann, Marcel Stenger and Markus Mörtl

Abstract Currently, innovations in product development are often implemented through a shift towards product-service systems. Thereby, products are combined with additional services, whereas this integration of both disciplines causes a new complexity. This complexity combined with a future oriented perspective requires a systematic support for planning these product-service systems. Core aspects are the reduction of uncertainty and a better preparation for future evolvements. This contribution presents a procedure that connects a systematic analysis of the future environment with the product-service system. The impact of future contextual influences on the product-service system is assessed by a future-oriented requirements specification. Consequently, the procedure transfers planning information into an engineering perspective. In order to ensure the procedure's applicability, a case study was performed in the innovative sector of autonomous driving in cooperation with a company from the German automotive industry.

Keywords Product-service system planning · Product development
Requirements engineering · Contextual influences · Future orientation
Case study

1 Introduction

Customer acceptance is a success factor of innovative products and product-service systems (PSS) can positively influence customer acceptance of innovations [1]. Consequently, many innovations in technical products are implemented through the integration of services. A shift from traditional products to PSS can be observed. However, since each upcoming product is influenced by a dynamic corporate context [2, 3], it is not only the shift of the business model that influences planning and specification of future PSS. PSS are developed under current conditions for

D. Weidmann (✉) · M. Stenger · M. Mörtl
Technical University of Munich, Boltzmannstr. 15, 85748 Garching, Germany
e-mail: weidmann@tum.de

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_28

future situations, hence, the dynamic corporate context needs integration into the planning of PSS. The future context could be described by changed legal issues, technology evolvement, changed usage behavior or a reorientation in the corporate strategy. A core issue is the comparison of various planning-relevant information, for instance, from marketing, but also from research and development, to derive coherent, future-oriented development proposals [4].

Current developments are mostly based on existing products, which means that new development shares are typically limited to shape variation or variation of solution principle [5]. These product generation developments use a reference product to give the basic structure. Albers et al. [5] point out that important decisions, which will largely determine the innovation potential of the product on the market, have to be made in a very early planning phase. Consequently, a systematic support using reference product information and connect them with future contextual situations can support the planning of PSS. Therefore, this contribution presents a systematic procedure for future oriented planning of PSS based on existing product structures and simultaneously with respect to future situation.

2 Research Methodology

In context of future oriented planning of product-service systems, the question how a potential future will affect the system is addressed. We ask the following research questions: (1) How can relevant influence factors that affect the future PSS be forecasted and combined, (2) how can these forecasted factors be connected with the PSS, and (3) which impact will these factors have on the PSS and its development strategy.

This research follows the four major steps of the Design Research Methodology by Blessing and Chakrabarti [6]: (1) In the research clarification, the problem statement is defined based on literature and observations of industrial practice in different sectors (automotive, public transport, and electric charging infrastructure). (2) As descriptive study I a literature analysis was conducted. On the one hand holistic approaches and on the other hand stand-alone approaches were analyzed according to the single aspects of the problem statement. (3) In the prescriptive study the solution concept was designed. Here, a procedure and adjusted methods from literature were combined to forecast the influence factors and connect them with the PSS. (4) In the descriptive study II an in-depth case study was conducted in context of autonomous driving to evaluate the concepts feasibility. This empirical industrial evaluation transfers the methodology from academia into industry and ensures its applicability.

3 Theoretical Background

3.1 Planning of Product-Service Systems

A PSS basically consists of a tangible product, a mostly commercial service adding economic value, and a system as collection of elements and their relations [7]. Three major types of PSS depending on the level of interaction between product and service can be distinguished [8]: (1) Product-oriented PSS are customer owned products, which are extended with additional services during the product lifecycle. (2) Use-oriented PSS are usually not owned by consumers, but companies offer the availability in order to maximize the utilization of the product. (3) Result-oriented PSS sell capabilities instead of traditional products. Consequently, customers pay for results, without owning the product. The concept of PSS implies a paradigm shift from selling technical products to fulfilling customer demands [9].

The strategic planning itself is increasingly anchored in companies and with it an understanding about the dynamic of numerous influencing factors and their associated complexity [10]. Nevertheless, Hepperle [10] points out the great need for improvement in the strategic planning, i.e. increasing transparency, planning reliability and plausibility. In terms of future oriented PSS planning, he especially asks for additional methods that include factors projected into future while analyzing planning-relevant relations.

In literature exists a number of planning approaches, however, only a few of them explicitly focus PSS. Previously conducted literature reviews can be found in contributions by Hepperle [10] and Kammerl et al. [11]. In context of systematic product planning Hepperle [10] figures out that methodological approaches addressing acquisition and projection of product-related, planning-relevant information are described in detail, but the analysis of the associated system context (i.e. market or technology) is only partially addressed. Kammerl et al. [11] show that, in contrast to sole product- or service modeling, there are few approaches of model-based support for product-service system planning that can be applied for a consistent representation and management of the complete planning-relevant information. Furthermore, the previously identified PSS design methodologies in literature are typically adoptions of methodologies developed for traditional products or services [12].

Core aspects of this concept are contextual influences and their forecasting. Literature provides various collections and understanding of influences on future product developments: technological, social and ecological changes increase complexity and dynamics [13]; changing customer demands, markets, legal and social systems result in a dynamic of competition and continuous development of applied technologies [14–16]. However, the focus is solely on influences on the PSS itself and not on its development process. In the field of forecasting previously identified and existing methods are integrated [3]. Exemplary forecasting methods are qualitative methods like scenarios, interviews, use cases, and roadmaps and quantitative methods like time series forecasting, cycles, and historical analysis [3].

3.2 Systems Engineering and Requirements Traceability as Underlying Idea

Systems engineering builds a general setting of this research. In previous research various systems engineering approaches were identified that deal with processes and procedures to systematically develop interdisciplinary products [17]. They typically start with requirements that are fulfilled by functions that are again fulfilled by product elements. Hence, a reasonable starting point for the concept are requirements as interdisciplinary basis of the development. Our previous analysis focused the dependencies of the system elements: requirements, product architecture (connection of functions and elements) and organizational units. Existing systems engineering concepts connect these elements at least partly and, consequently, focus a consistent system.

Since requirements are the basis of each development, requirements are used as intersection between PSS planning and development. With requirements as development basis, requirements traceability plays a major role to plan future PSS. The presented graph based approach that uses structural dependencies [18] is applied to identify the propagation of requirements in the system. Based on this traceability, adaptions of systems caused by requirements changes can be identified. A PSS specific model to visualize the elements and their connections is presented by Kammerl et al. [11]. This model is integrated in the presented approach.

4 Future Oriented Planning of Product-Service Systems

The brief literature review motivates to develop a holistic planning approach. Therefore, aspects of innovation management, requirements management and systems engineering are combined. Planning relevant information of various sources are combined and transferred from a planning into an engineering design perspective. Previous studies [11, 17] have shown that requirements serve as interdisciplinary foundation for the PSS concretization. According to the product generation development [5], the presented approach is based on a reference product and consequently, its dataset.

From a given set of contextual information (e.g. technology, legal issues, or corporate strategy) a problem-, company- and product-specific collection of influences is selected. This selection is forecasted to defined time horizons. Depending on type and availability of data, a combination of quantitative and qualitative forecasting methods is applied. The forecasted influences are connected with the PSS. Here, requirements serve as an interdisciplinary basis for connecting the prognosis with the actual PSS. Building on that, requirements traceability is used for an impact analysis on the PSS to further detail the development perspective. This impact analysis can cover various factors to either set development priorities or prepare development decisions.

This concept is transferred into a systematic procedure (Fig. 1) which is aligned to the single aspects of the concept. Even it is focused on one planning horizon, however, it could easily extended to multiple horizons.

In the first phase (I) product information are acquired according to the PSS planning model [11]. Core information are the product architecture and connected requirements. Additionally, contextual information are collected according to the considered system boundary. The second phase (II) consists of three steps that must not follow a stringent order. In the prioritization step (II-A) relevant information are selected on product and context level. Depending on the task, influences can be summarized to handle complexity and variety. Fitting prediction methods (II-B) are selected according to the given data. The now forecasted influences are connected with the PSS's requirements (II-C). Therefore, a matrix (Fig. 2) is used, which has on the one side forecasted and prioritized contextual information (left), and on the

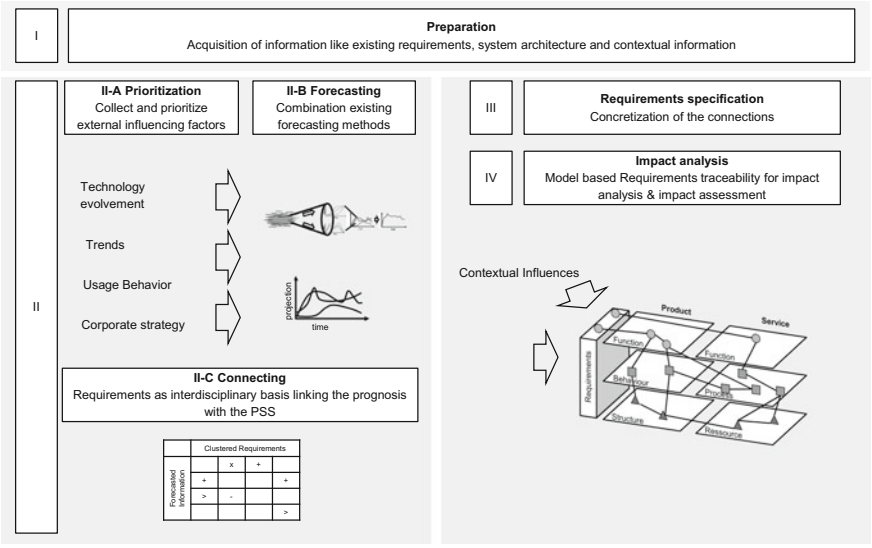


Fig. 1 Methodology for future oriented planning of product-service systems

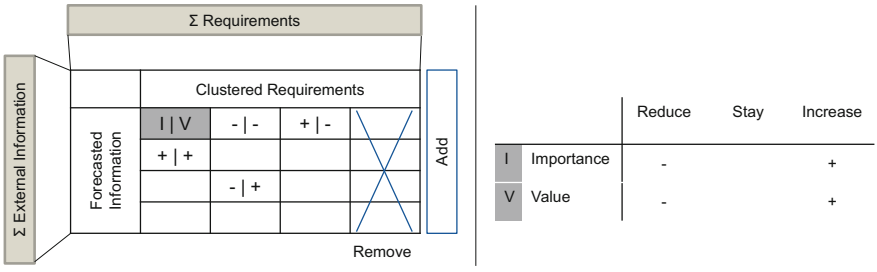


Fig. 2 Connection matrix to summarize information, connect elements and assess requirements

other side PSS requirements (top). The mostly extensive amount of requirements can be clustered into categories to handle complexity. At the same time the linking of contextual influences and the PSS can simultaneously be a prioritization. If there are no connections to the PSS, the influence might not be relevant. The connection of both domains follows two aspects. On the one hand, the changed future importance and on the other hand the changed future value of the requirement. Consequently, the connection follows one the following options:

- There is no influence (importance and value) on the requirement
- The requirement's importance: stays/reduces/increases
- The requirement's value: stays/reduces/increases
- There is need for a new requirement (add)
- The requirements is unnecessary (delete).

With a concretization of the connections, the requirements are specified (III) and subsequently, a future requirement set that describe future PSS on requirement level is generated. The connection matrix (Fig. 2) summarizes influences and shows how requirements change under various influences. This procedure could be performed for various time horizons, so requirements can be considered under both aspects, time and influence robustness.

Based on these information a more detailed impact analysis can be conducted (IV). According to changes in the requirement set specific parts of the PSS are analyzed in more detail. Alternatively, defined decision criteria can be used for prioritization. The PSS planning model [11] and requirements traceability [17] are applied for the impact analyses. Therefore, the background to identify more detailed effects on the PSS is given by the structural complexity management [18]. Depending on the abstraction level, affected system elements included in the PSS planning model can be identified. Nevertheless, the structural analysis is only the basis for a reasonable expert evaluation, i.e. to assess the effects of changed requirements. The challenge to identify relevant stakeholder could be solved with consequent traceability throughout the system.

5 Case Study-Based Evaluation in the Automotive Industry

5.1 Case Study Description and Implementation Overview

Due to confidentiality reasons, we need to make a compromise between a detailed as possible and an abstract as necessary description of the following methodology's evaluation. The procedure and methods are developed in cooperation with an industry partner from the German automotive industry. The case study's technical goal was to develop a concept for the inner space of a car in the year 2025 in context of an urban and autonomous vehicle concept. New characteristics of an

autonomous vehicle are derived and compared with existing requirements to identify how future application and situation affects the properties of the vehicle. From a scientific point of view, the case study's goal was to ensure the methodology's applicability and efficiency.

The case study's implementation basically follows the proposed procedure, however, it is adopted to the task-specific situation. The applied procedure is shown in Fig. 3. It starts with the information acquisition including a reference requirements list and collected influence factors, followed by the forecasting and prioritization step. Here, requirements are clustered and influence factors are combined and forecasted in a higher level application scenario. Within this scenario different use cases are developed. The connection matrix serves as tool to connect the use cases with the requirements cluster. Finally, an assessment of future requirements is conducted by experts, whereas the impact analysis was only conducted incidentally.

Preparation and Acquisition In order to manage the huge number of reference requirements in this study, we used property clusters, which are commonly used in automotive industry to show potentials and deficits in the development of new vehicles in comparison to the competition. The clusters can be divided into several levels and include both, internal and customer requirements. We considered aspects of different areas: environment, society, market, state, traffic, and company. In concrete terms, information from different departments are considered, i.e. strategy, marketing, laws to collect contextual factors.

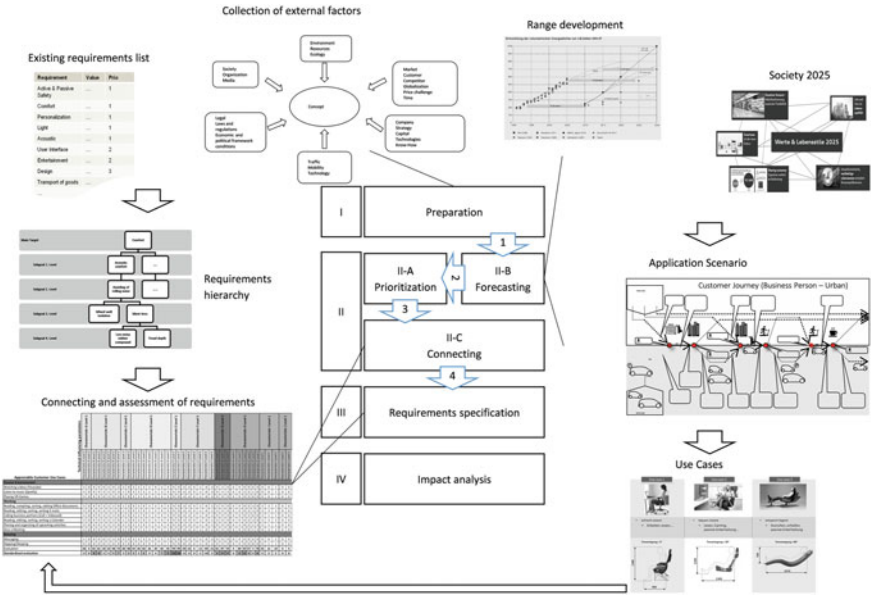
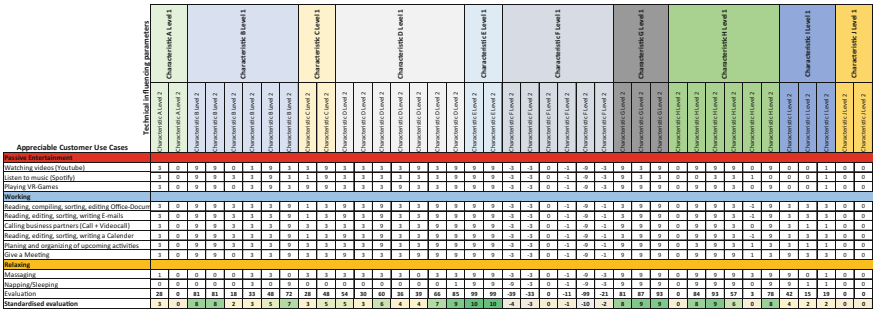


Fig. 3 Methodology (I–IV) and case study specific adjusted procedure (1–4), methods and results

Prioritization and Forecasting Both steps, prioritization and forecasting were conducted in an iterative manner. First, the change in customer needs for 2025 was estimated based on the collected information. Therefore, aspect of three areas were considered: customer view, area of application and mobility concept. From the customer's point of view, using time and space during the journey are particularly important. Furthermore, based on defined parameters cities were clustered to identify potential suitable areas of operation. In addition, climatic conditions and the influence of new technologies, such as fast-charging options were analyzed in order to reasonably predict the mobility concept (i.e. charging infrastructure). Therefore, quantitative (i.e. battery capacity) and qualitative (i.e. society) forecasting methods are applied. Based on this forecasted factors a high level application scenario is derived and refined with use cases, which represent a set of customer needs based on a role or task. Based on a study by Dungs et al. [19], nine main use cases were developed. Each main use case is defined by a more precise description to limit the room for interpretation during the connection phase. Exemplary use cases are: private communication, eating/drinking, passive entertainment, relaxing, and working.

In a next step three use cases were prioritized with a pairwise comparison: working, relaxing and passive entertainment. A major criteria was their estimated relevance according to important adjustments on the hardware side. This gets particularly clear while looking at the seating position, i.e. for the use case work, the optimal seating position is an upright position, however, for maximum comfort (relaxing) a zero gravity position is preferred. As final result of this step, the three main use cases are refined by more specific use cases (a total of 11, i.e. passive entertainment: watching videos).

Connecting Use Cases and PSS Requirements Forecasted information are condensed in prioritized use cases and the large amount of requirements is managed with a hierarchy according to the property clusters. To connect both domains, the previously presented connection matrix is applied while solely focusing on the changed importance of requirements caused by the use cases. A specific rating for the relationship of each technical requirement and use case is conducted in the connection matrix. The rating is executed by different departments of the company, precisely Overall Vehicle Properties, Predevelopment Vehicle Concepts and User Experience. It is particularly important that involved persons have a good understanding of the use cases. The question how future use cases affect the vehicle characteristics is depicted in a manageable but nevertheless sufficient detail level. The strength of the relationship between the use cases and technical requirements is measured with a progressive scale, while it was extended it to negative values (−9 to +9) due to their relative character. We tried to balance the subjective character of the assessment with the interdisciplinary perspectives of the departments. The major use cases and their refined use cases are on the left side of the matrix and the requirements cluster are on the upper side (Fig. 4). The assessment of future importance of each requirement in context of a specific use case can be seen in each cell.



Technical influencing parameters	Characteristic A Level 1		Characteristic B Level 1					Characteristic C Level 1		Characteristic D Level 1						Characteristic E Level 1		Characteristic F Level 1				Characteristic G Level 1		Characteristic H Level 1						Characteristic I Level 1		Characteristic J Level 1	
Appreciable Customer Use Cases	Characteristic A Level 2	Characteristic A Level 2	Characteristic B Level 2	Characteristic B Level 2	Characteristic B Level 2	Characteristic B Level 2	Characteristic B Level 2	Characteristic C Level 2	Characteristic C Level 2	Characteristic D Level 2	Characteristic D Level 2	Characteristic D Level 2	Characteristic D Level 2	Characteristic D Level 2	Characteristic D Level 2	Characteristic E Level 2	Characteristic E Level 2	Characteristic F Level 2	Characteristic F Level 2	Characteristic F Level 2	Characteristic F Level 2	Characteristic G Level 2	Characteristic G Level 2	Characteristic H Level 2	Characteristic H Level 2	Characteristic H Level 2	Characteristic H Level 2	Characteristic H Level 2	Characteristic I Level 2	Characteristic I Level 2	Characteristic J Level 2	Characteristic J Level 2
Passive Entertainment																																	
Watching videos (Youtube)	3	0	9	9	0	3	9	3	3	9	3	3	3	3	9	3	0	9	9	-3	-3	0	-1	-9	-3	1	9	0	9	9	9	0	9	0	0	1	0	0
Listen to music (Spotify)	1	0	9	9	3	3	9	1	1	9	3	3	3	3	3	3	0	9	9	-3	-3	0	-1	-9	-3	1	9	0	9	9	9	1	1	0	0	1	0	0
Playing VR-Games	3	0	9	9	0	3	9	3	9	9	3	3	3	3	9	3	0	9	9	-3	-3	0	-1	-9	-3	1	9	0	9	9	9	3	3	0	0	1	0	0
Working																																	
Reading, compiling, sorting, editing Office-Docum	3	0	9	9	3	3	3	9	1	3	9	3	9	3	3	9	9	9	9	-3	-3	0	-1	-9	-3	3	9	0	9	9	9	3	-1	9	3	3	0	0
Reading, editing, sorting, writing E-mails	3	0	9	9	3	3	3	9	1	3	9	3	9	3	3	9	9	9	9	-3	-3	0	-1	-9	-3	3	9	0	9	9	9	3	-1	9	3	3	0	0
Calling business partners (Call + Videocall)	3	0	9	9	3	3	3	9	3	3	3	3	3	9	3	9	9	9	9	-3	-3	0	-1	-9	-3	3	9	0	9	9	9	3	0	9	3	1	1	0	0
Reading, editing, sorting, writing a Calendar	3	0	9	9	3	3	3	9	1	3	9	3	9	3	3	9	9	9	9	-3	-3	0	-1	-9	-3	3	9	0	9	9	9	3	-1	9	3	3	0	0
Planing and organizing of upcoming activities	3	0	9	9	3	3	3	9	3	3	3	3	9	3	3	9	9	9	9	-3	-3	0	-1	-9	-3	3	9	0	9	9	9	3	-1	3	3	1	1	0	0
Give a Meeting	3	0	9	9	0	3	3	9	3	3	9	3	3	3	9	9	9	9	9	-3	-3	0	-1	-9	-3	3	9	0	9	9	9	3	1	3	3	3	0	0
Relaxing																																	
Massaging	-1	0	0	0	0	3	3	0	3	3	3	3	3	0	3	3	9	9	9	-3	-3	0	-1	-9	-3	3	9	0	9	9	9	3	3	9	9	0	1	0	0
Napping/Sleeping	-9	0	0	0	0	3	3	0	9	0	0	0	0	0	0	0	1	9	9	-9	-3	0	-1	-9	-3	0	9	0	9	9	9	9	9	0	9	9	1	1	0	0
Evaluation	**28**	**0**	**81**	**81**	**18**	**33**	**48**	**72**	**28**	**48**	**54**	**30**	**60**	**36**	**39**	**66**	**85**	**93**	**99**	**-39**	**-33**	**0**	**-11**	**-99**	**-21**	**81**	**87**	**93**	**0**	**84**	**93**	**57**	**3**	**78**	**42**	**15**	**19**	**0**	**0**
Standardised evaluation	**7**	**0**	**8**	**8**	**2**	**3**	**5**	**7**	**3**	**5**	**5**	**3**	**6**	**4**	**4**	**7**	**9**	**10**	**10**	**-4**	**-3**	**0**	**-1**	**-10**	**-2**	**8**	**9**	**9**	**0**	**8**	**9**	**6**	**0**	**8**	**4**	**2**	**2**	**0**	**0**

Fig. 4 Connection matrix of the case study

Requirements Specification and Impact Analysis Based on the results of the connection matrix and thus the changed importance of requirements, future development priorities were derived. Consequently, the question how and to what extent the use cases influence the reference property framework were answered. This points out the gain or loss in importance of each requirements. Therefore, a column-wise and row-by-column comparison is carried out. The sum of each column in the matrix represents the importance change of each requirement in context of the future oriented use cases. To make the values comparable, the summarized importance values are standardized and ranked. Thus, the requirements can be categorized into different prioritization groups that serve as guideline for the development to derive development priorities for future PSS. Based on this information a more detailed impact analysis according to the model of Kammerl et al. [11] and the methodology of Weidmann et al. [17] can be conducted, but nevertheless, the future importance of each requirement was assessed, i.e. an increased importance for active and passive safety, unchanged importance of entry & exit, and a reduced meaning for sound character.

5.2 Discussion

This holistic methodology combines planning relevant information from various sources and transfers them from a future-oriented planning perspective into an engineering design perspective. It shows how to handle the variety and complexity of various influences and integrates several stand-alone concepts in a continuous procedure. The procedure efficiently connects forecasted information with the PSS via requirements. The influence on the requirements is assessed by experts. This serves as basis for an impact analysis on the PSS with reasonable effort. However, this case study could not provide a systematic proof of the impact analysis, but at least a starting point based on the planner's information. The contribution shows the feasibility and applicability of the procedure and its methods in an industrial context. We gained the knowledge that there was no systematic way to identify cross

system boundary relations. For instance, the focus of the case study was on the interior, however, effects on the overall vehicle concept exists, like the interior use case “relaxation” has effects on overall vehicle characteristics (for example, longitudinal and lateral dynamics).

The methodology is limited in its scope, hence, system adaptions that are not based on requirements are not considered, i.e. cost reductions strategies or production aspects on component level. Furthermore, a number of complexity drivers were identified during the case study, i.e.: known/unknown influences, number of considered influences, number of considered time horizons, number of alternative solutions, or granularity of the requirements specification. All of them increase both complexity and effort.

6 Summary and Outlook

In industry a shift from traditional products towards product-service systems (PSS) can be observed. These PSS are developed under current conditions for dynamic future situations. Therefore, a future oriented PSS planning procedure is presented. This contribution shows how to prioritize and summarize forecasted contextual information in use cases and connect them efficiently with requirements of a reference product to derive development priorities or support development decisions. The procedure and methods are applied in the German automotive sector to ensure its practicability and efficiency. Nevertheless, more case studies are needed to ensure the methodology’s generic character. Based on the assessed requirements there is potential in the field of an impact analysis to identify concrete PSS elements that are critically affected by future adjustments. Another open issues is a decision support to assess the criticality of adaptions in requirements importance. Therefore, a model based support is proposed to conduct analyzes on the system with structural complexity management methods [18].

Acknowledgements We thank the German Research Foundation (Deutsche Forschungsgemeinschaft —DFG) for funding this project as part of the collaborative research center „Sonderforschungsbereich 768—Managing cycles in innovation processes—Integrated development of product-service-systems based on technical products .

References

1. Schmidt DM, Braun F, Schenkl SA, Mörtl M (2016) Interview study: how can product-service systems increase customer acceptance of innovations? CIRP J Manufact Sci Technol 15:82–93
2. Browning TR, Honour EC (2008) Measuring the life-cycle value of enduring systems. Syst Eng 11(3):187–202

3. Bauer W (2016) Planung und Entwicklung änderungsrobuster Plattformarchitekturen. Dissertation, Technical University of Munich, Munich
4. Ulrich KT, Eppinger SD (2004) Product design and development, 3rd ed. Irwin McGraw-Hill, Boston, Mass
5. Albers A, Bursac N, Wintergerst E (2015) Product generation development-importance and challenges from a design research perspective. New Develop Mech Mech Eng
6. Blessing LT, Chakrabarti A (2009) DRM, a design research methodology. Springer, London, London
7. Goedkoop MJ, van Halen CJG, Te Riele HRM, Rommens PJM et al (1999) Product service systems, ecological and economic basics. In: Report for Dutch Ministries of environment (VROM) and economic affairs (EZ), vol 36, no 1, pp 1–122
8. Tukker A (2004) Eight types of product–service system: eight ways to sustainability? Experiences from SusProNet. Bus Strat Env 13(4):246–260
9. Baines TS et al (2007) State-of-the-art in product-service systems. J Eng Manuf 221 (10):1543–1552
10. Hepperle C (2016) Planung lebenszyklusgerechter Leistungsbündel. Dissertation, Technical University of Munich, Munich
11. Kammerl D, Winkler S, Schmidt D, Mörtl M (2016) Model-based support for product-service system planning. In: Design 2016 14th international design conference
12. Weidmann D, Maisenbacher S, Kasperek D, Maurer M (2015) Product-service system development with discrete event simulation. In: IEEE systems conference (SysCon)
13. Bleicher K (2011) Das Konzept Integriertes Management, 7th edn. Campus Verlag, Frankfurt
14. Guiltinan JP (1999) Launch strategy, launch tactics, and demand outcomes. J Prod Innov Manag 16(6):509–529
15. Schulz AP, Fricke E, Igenbergs E (2000) Enabling changes in systems throughout the entire life-cycle-key to success? In: INCOSE international symposium 2000, vol 10, pp 565–573
16. Cardin MA (2014) Enabling flexibility in engineering systems: a taxonomy of procedures and a design framework. J Mech Des (136)
17. Weidmann D, Becerril L, Hollauer C, Kattner N, Lindemann U (2017) A network-based approach to identify lacking coordination using higher order links. In: 21st international conference on engineering design (ICED17)
18. Lindemann U, Maurer M, Braun T (2009) Structural complexity management: an approach for the field of product design. Springer, Berlin
19. Dungs J, Herrmann F, Duwe D, Schmidt A, Stegmüller S, Gaydoul R, Sohl M (2016) The value of time. Fraunhofer IAO, Stuttgart

The Berkeley Innovation Index: A Quantitative Approach to Measure, Track and Forecast Innovation Capability Within Individuals and Organizations

Alexander Fred-Ojala, Ikhlaq Sidhu, Charlotta Johnsson and Mari Suoranta

Abstract Innovation and entrepreneurship are essential processes for human development, market growth, and technological breakthroughs, and it is vital for economic growth. Despite its importance, innovation is inherently difficult to measure and hence it is almost impossible for an individual or organization to know how they can improve their innovation output or claim that they are great at innovation. This paper presents a novel approach to measure and quantify innovation, called the Berkeley Innovation Index (BII). The BII characterizes and measures innovation capability of an individual or an organization. It builds on the hypothesis that innovation performance depends on the people, culture, and attitudes in an organization. With the BII, the individual mindset and the culture of a larger organization, can be measured, analyzed, evaluated, and tracked. It also enables organizations and individuals to adopt a forward-looking approach to forecast and predict future innovation capabilities. This paper presents the approach and its fundamentals, and it includes examples of how the BII has been used to track changes in innovative mindsets.

Keyword Innovation assessment · Quantitative measurement · Entrepreneurship research · Forecasting · Innovation capability

A. Fred-Ojala (✉) · I. Sidhu
UC Berkeley, Berkeley, USA
e-mail: afo@berkeley.edu

I. Sidhu
e-mail: sidhu@berkeley.edu

C. Johnsson
Lund University, Lund, Sweden
e-mail: charlotta.johnsson@lth.lu.se

M. Suoranta
Jyvaskyla University, Jyvaskyla, Finland
e-mail: mari.suoranta@jyu.fi

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_29

1 Introduction

Innovation and entrepreneurship are essential processes for human development, market growth, and technological breakthroughs. Entrepreneurship is often thought of as the act of commercializing an innovation. In modern open economies, entrepreneurship is one of the key aspects for economic growth. However, innovation is quite a vague and abstract concept, and therefore it is inherently difficult to measure. Today it is almost impossible for an individual or organization to know how they can quantify innovation capability, improve their innovation output, or claim that they are great at innovation.

Measurements and performance tracking matter, because it is difficult to improve what you cannot measure. Measurements and indices enable comparisons and the ability to track progress by the use of e.g. a performance score. Being able to measure and characterize innovation capability, of an individual or organization, is therefore crucial for an entity to know how they perform today and how they should act in order to improve in the future. A performance index enables insight into current capabilities as well as tracking of the development and improvement of specific traits.

Berkeley Method of Entrepreneurship is a holistic teaching and learning approach that is hypothesized to enable engineers to be more entrepreneurial. It stresses the fact that in order to be innovative, knowledge in the STEM subjects is not enough, an entrepreneurial mindset is also needed. If the majority of the members in a community has an entrepreneurial mindset, the result will be a vibrant entrepreneurial culture.

The Berkeley Innovation Index (BII) is a measurement and performance index that focuses on individual and organizational innovation capability, and related innovative mindsets. It builds on the hypothesis that innovation performance depends on the people, culture, and attitudes in an organization. With the Berkeley Innovation Index, the individual mindset and the culture of a larger organization, can be measured, analyzed, evaluated, and tracked. The BII also enables organizations and individuals to adopt a forward-looking approach to forecast and predict future innovation capabilities.

Empirical results from applying the BII demonstrate how it can diagnose the state of an organization as well as the traits and characteristics among the teams and the individual employees. The BII provides an important framework for granular insight into company culture, team alignment, mindset characteristics, and values, all of which can be used to inform strategic decisions and shape company objectives.

Innovation is one of the top three most important aspects of business success according to over 70% of executives surveyed in a recent study by McKinsey, and 94% of the same senior executives say that people and corporate culture are the most important drivers of innovation. However, only 27% of the same executives believe that their company has the right people and leaders to innovate [1]. This gap is what the BII aims to bridge, it is a tool designed to quantify innovation capability

and give managers, leaders, and executives a data driven approach to increasing their innovation performance. The BII also provides the ability to benchmark an organization's current innovative state and compare it to other market players in the same geographical region or similar industry, as well as to track improvements over time and predict performance in the future.

This paper presents BII and how it can be used to track innovation performance for groups over time, how it can be used to measure innovation culture and operational focus of workgroups, and how to evaluate and improve overall innovation performance. The paper starts by presenting our view on Innovation and Entrepreneurship (Sect. 2), as well as why Measuring, Tracking and Forecasting is important (Sect. 3). Thereafter follows a description of Berkeley Method of Entrepreneurship (Sect. 4) and the Berkeley Innovation Index (Sect. 5). Empirical results from applying the BII on a group are presented (Sect. 6), this is followed by the workgroup index extension of the BII (Sect. 7), and finally the conclusions (Sect. 8) are presented.

2 Innovation and Entrepreneurship

Entrepreneurship is often thought of as the act of commercializing an innovation. In modern open economies, entrepreneurship is one of the key aspects for economic growth [2, 3]. The economic engine over time has, in the US and other countries, been technology [4], as technology constitutes a fundamental part in many of today's innovations. Clearly, a good understanding of the fundamentals in technology is a prerequisite for innovation and entrepreneurship, and hence technical companies are investing in R&D activities. In parallel engineering education is focusing on providing a thorough understanding of science, technology, engineering, and mathematics (STEM), where the hypothesis seems to be, the better understanding of you have of STEM, the higher the innovation potential.

Innovation is the ability to come up with original ideas, improve a process or a product, be creative, and to act as a pioneer. The current market landscape is characterized by rapid technological progress and disruptive forces, an organization today needs to present solutions to new and old problems by applying new methods and emerging technologies. Without the ability to be innovative an organization or an individual experience stagnation, and in a world that encourages constant progress the lack of innovation capability becomes a major disadvantage [5].

Companies and organizations have identified the importance of innovation, and this is clearly reflected in the increase of global spending on Research and Development (R&D). Global R&D investments have increased significantly from 2005 to 2014, see Fig. 1.

Innovation is evidently deemed to be important, but it is also a vague and abstract concept. In a study conducted to define innovation the authors found that the foundation for all innovation processes in an organization is brilliant, motivated, experienced, and creative employees. They found that innovation processes are

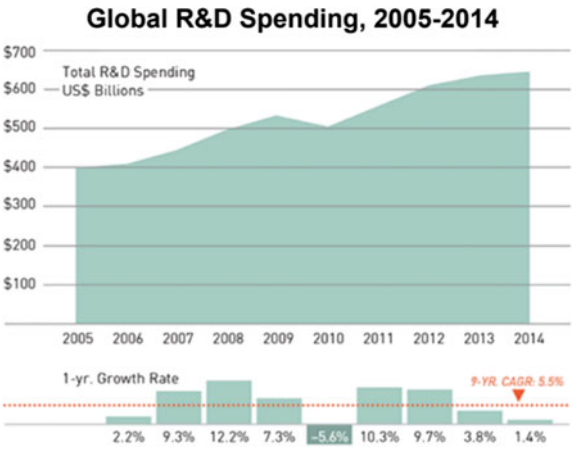


Fig. 1 Global R&D spending from 2005 to 2014. *Source* Bloomberg data, capital IQ data, strategy and analysis

generally a collective achievement of the organization's members. An organization with a thriving innovation culture and people with the right mindsets are critical for the firm to pursue technological advancements and constitutes the best incentive towards obtaining new knowledge and innovating [6].

3 Measuring, Tracking, and Forecasting

The importance of innovation for organizations and individuals has created a need for a tool that can track and measure innovation capability. This in order to benchmark performance and discover how to optimize innovation output. A measurement of innovation capability can help identify behaviors and strategies that can be implemented to develop an innovative culture and innovation mindsets. In the article *Why Measuring Innovation Matters* Brian Quinn (Senior Contributor, Forbes) eloquently expressed why it is important to quantify innovation capability:

> What gets measured gets done [..] Without measuring these things [innovation], we're effectively driving without headlights — faintly hoping once again that innovation will deliver something useful rather than demanding it, and holding ourselves accountable for achieving it.

I.e., if you can set a target and measure where you are as well as tracking your progress, then any type of goal will be attainable. If a trait cannot be measured, it is inherently difficult, for any person or organization, to improve. Most past measures quantifying innovation capability and output have looked at hard financial metrics that have proved to have little to no correlation with innovation output. E.g., the numbers of patents filed within a year or the amount R&D spending have not shown any significant relationship with an organization's ability to be innovative nor to make profits [5].

4 Berkeley Method of Entrepreneurship

The Berkeley Method of Entrepreneurship (BMoE) is a holistic teaching framework and a learning approach that is hypothesized to enable engineers to be more entrepreneurial. It stresses the fact that in order to be innovative, knowledge in technical STEM subjects is not enough—an entrepreneurial mindset is also needed. Generally, the mindset is a way of thinking that influences the way someone sees and acts upon a situation; the mindset is reflected in the person's behavioral patterns [7]. BMoE has identified ten behavioral patterns of successful entrepreneurs and, by using inductive and game-based teaching, the mindset behind the behaviors can be explored. The ten behaviors are: Pay It Forward, Story Telling, Friend or Foe, Seek Fairness, Plan to Fail, Diversify, Role Model, Believe, Good Enough, and Collaboration [5, 8]. BMoE also presents a teaching and learning method used to get engineers to explore their own mindset in comparison to these ten behaviors [9]. The novelty of BMoE lies in its strong focus on mindset and the belief that an entrepreneurial mindset can be created and improve over time.

5 Berkeley Innovation Index

The Berkeley Innovation index is a psychometrics tool that quantifies and measures an individual's or organization's overall innovation capabilities [10]. The index assesses overall innovation capability as well as six sub-traits. The sub-traits constitute a subset of the ten entrepreneurial mindsets identified in BMoE that are linked to innovation performance. The individual sub-traits of the BII are defined in Table 1.

An entrepreneurial mindset constitutes of three sub-dimensions: innovativeness, risk-taking, and proactiveness [11]. The BII is based on a combination of entrepreneurial mindset characteristics and research findings from the fields of psychology, business, and mathematical statistics. The BII deconstructs six of the ten traits identified with an entrepreneurial mindset in the Berkeley Method of Entrepreneurship [5]. It is these six personality traits that form the basis of the BII. The approach is also intended to cover layers of innovation that range from the following fields: (1) Strategy and Leadership, (2) Innovation Culture from an Organization's Viewpoint, (3) Organizational Operations and Measures across functions, (4) Mindset: The Innovation DNA of the People, and (5) Tactical measures [5].

To assess the BII scores of an individual or organization a two-part survey is distributed among the participants of the assessment. Each question belongs to a category, and each category is related to a trait that has been linked to innovation capability. The BII algorithm, that calculates the scores for the traits, has been developed utilizing a Higher-Order Item Response Theory approach. This makes the inferred index scores statistically relevant according to contemporary algorithmic theory and survey design [12].

Table 1 List of Berkeley Innovation Index traits and characteristics linked to individual innovation capability

Mindset and description	Psychological construct	Trait
Friend or foe Learn to trust others without expecting anything in return	Social cohesion, honest behaviour	Trust
Plan to fail It is necessary to be wrong sometimes. Plan to experiment. plan to fail (fail fast). Analyze, adapt and repeat. The smarter you think you are. The harder this is going to be	Grit, resilience, entrepreneurial failure	Resilience
Diversity Diversify your networks. Connect to people you would not normally, then go and listen, open up, and connect them to others	Social capital	Diversity
Believe Believe that what you do can change the world	Self-efficacy	Belief
Good enough Perfection is not good, but good enough is perfect	Perfectionism	Perfection
Collaboration Individual versus team and competitors versus partners	Cooperation	Collaboration

The BII is also based on the assumption that innovation capability can be improved, i.e., it is not a static trait, but something that can change if the right tools are employed. Therefore, the BII can be used to measure an individual's degree of innovation capability over time and track if any improvements have been made. The novelty of the BII lies in its forward-looking approach (i.e. the ability of being innovative), rather than focusing on past-oriented metrics such as investments (e.g. R&D spending) or actions (e.g. patents filed).

Other similar attempts to measure innovation has been made, for example the ISO certified Innovation360 assessments or InnoQuotient [13]. These two other approaches that aim to quantify innovation capability and performance are different from the BII in that they look at innovation from a structural and organizational viewpoint, while the BII is emphasizing the individual mindsets and personality characteristics that are linked to creating an organizational culture where innovation can thrive.

6 Track Individual Innovation Performance Increase

As of today (Feb 2018), over 6000 individuals from over 50 countries have completed the BII survey. The data has been collected by hosting the surveys on the research project's website [14]. The samples include professionals from many different industries (startups, mid-sized firms, and Fortune 500 companies) as well

as academics, scholars, and students. This data shows a trend where people exposed to mindset training can increase their innovation capability scores, as measured by the BII. Below is a sample case from the data set.

In January of 2016, the Berkeley Innovation Index survey was offered as an instructive aid to the participants of a four-day intensive class at UC Berkeley called The Berkeley Method of Entrepreneurship Bootcamp. In a controlled manner, the entire class of 106 students were offered the survey instrument before the first session of class. Note that these results were shown with linearly weighted scores, but not as number of standard deviations from a mean. The results of the pre-test show a mean of 7.5 for the total BII score with a standard deviation of just under 1.0 and are presented in Fig. 2.

Then, on day 3 of the bootcamp, the instrument was offered again but only to 35 of the students. At this stage of the course, there was no direct instruction to explain innovation mindset or entrepreneurial culture. However, the mean score increased approximately by 0.5, see Fig. 3. The changes in scores at this point were only due to indirect exposures of growth mindset training.

38 of the remaining students retook the BII survey after the course completed on day 4. During day 4, instruction included an explanation of innovation mindset,

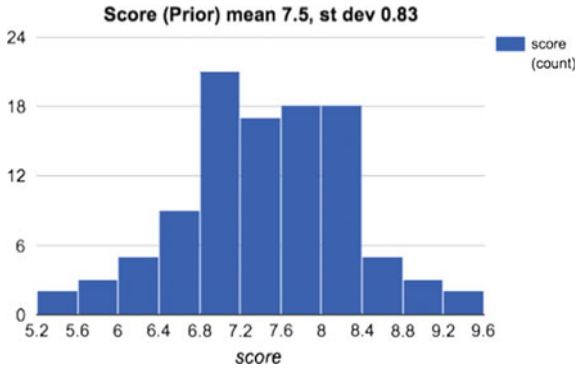


Fig. 2 BII score results prior to by exposed to entrepreneurial mindset training

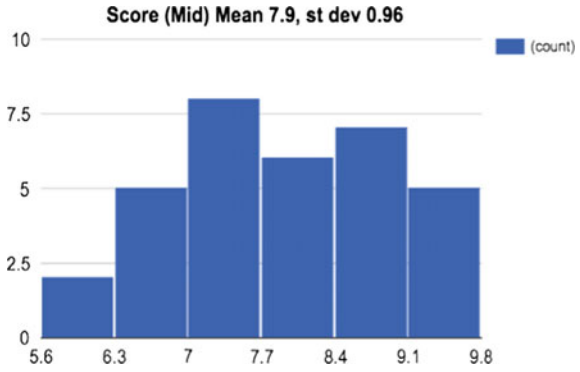


Fig. 3 BII score results for a third of the class on day 3 of the Bootcamp (prior to mindset training)

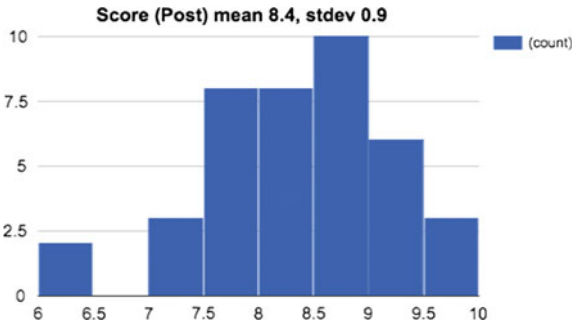


Fig. 4 BII score results for 38 students in the class on day 4 of the Bootcamp. Note that this was not the same group as the one assessed on day 3 (see Fig. 3)

innovation mindset training through exercises, and how to reinforce resulting behaviors common to entrepreneurs who possess a growth mindset. The mean score for the participants on that day increased by over one standard deviation, i.e., an absolute increase of 0.9 in Berkeley Innovation Index score for the group (Fig. 4).

It is important to understand that the test questions are a psychometrics tool and that they do not measure knowledge learned, but instead they measure what students believe about themselves at a psychological level. This result shows that the mindset training influenced the students' psychological beliefs and behaviors, not simply the logical understanding of the materials. We believe that the students' actual, real-life behaviors will be driven by this change in mindset and psychology more than if they remember or have a logical understanding of the material introduced.

Teaching a person to be entrepreneurial and innovative must include the behaviors needed to adapt in a volatile environment and to be able to go forward in risky and uncertain situations. This is what is taught at the BMoE bootcamp. The use of the BII survey instrument indicates that entrepreneurial behavior can be measured as well as learned. Innovation in large corporations also requires the right mindset and culture for the employees, or the correct balance of characteristics and profiles among workgroups. The BII makes it possible to measure the collective mindset of groups, as well as to track progress and improvement generated by training programs.

7 Extending the Berkeley Innovation Index to Workgroups

In order to add additional components needed for the BII to provide insight on innovation within a larger organization, and not only aggregated results for individuals, a second psychometrics instrument has been designed and implemented in the Berkeley Innovation Index suite. This tool focuses on the culture of the workgroup as measured by the individual's perception of the organization, management, company values, and the team they are part of.

Areas of measurement that are included in the Workgroup assessment are:

1. Where ideas originate
2. Transparency in decision making
3. Organizational comfort with ambiguity and learning
4. Responses to organizational failures
5. Cultural understanding about operating measures such as product/service quality, customer happiness, product cost, and market share
6. Culture of execution.

The organizational characteristics above are measured with statement questions on an ordinal Likert-scale related to the categories. All of these perspectives and aspects are critical components for an organization to recognize its strengths and weaknesses, form alignment with its team members, effectively allocating resources, and optimizing the organizational workflow according to project needs. The results of these questions are grouped to form a score and measure that quantifies a company's *innovation culture* (questions 1–3) and *operational focus* (questions 4–6). It is hypothesized that a balance between these two skills is essential for any organization to be successful in the long run.

8 Conclusions

In this paper we presented a novel way to utilize the Berkeley Innovation Index as a measurement to track innovation performance, and capability when a group of people are exposed to entrepreneurial mindset training. We also suggested an extended feature to the BII that incorporates a workgroup score with emphasis on organizational effectiveness (in the form of Innovation Culture and Operational Focus). This research leads the way for companies to setup processes where the BII is used to track and improve their innovation performance. The empirical results present a tool that can be useful for individuals as well as companies on a global basis. Over 6000 individuals have already taken the BII surveys. The BII data set in combination with the theoretical frameworks provide a unique tool that can be used to quantify innovation capability and produce recommendations on how innovation performance can be improved.

References

1. Barsh J, Capozzi M, Davidson J (2008) Leadership and innovation. McKinsey Q 1(2008):37–47
2. Thurik R, Audretsch D (1998) The Knowledge society, entrepreneurship and unemployment. Scales Research Reports H199801, EIM Business and Policy Research, 1998
3. Casson M (1995) Entrepreneurship and business culture. Edward Elgar, Aldershot

4. Today's engineer (2007) Teaching today's engineering students to be tomorrow's entrepreneurs. In: IEEE-USA Today's Eningeer Online, http://www.todaysengineer.org/2007/jul/entrepreneurship.asp
5. Sidhu I, Singer S, Johnsson C, Suoranta M (2015) Introducing the Berkeley method of entrepreneurship—a game-based teaching approach. In: American society for engineering education (ASEE) annual conference 2015, Seattle, WA, USA, 14–17 June, 2015
6. Castro M, Delgado-Verde M, Navas-López JE, Cruz-González J (2013) The moderating role of innovation culture in the relationship between knowledge assets and product innovation. Technol Forecast Soc Chang 80:351–363
7. Dweck C (2006) Mindset: the new psychology of success. Random House, New York
8. Johnsson C, Sidhu I, Souranta M, Singer K (2017) Guiding students towards an entrepreneurial mindset. In: Teaching and learning entrepreneurship in higher education, July 2017. Libri Publishing
9. Johnsson C, Loeffler R, Sidhu I, Nilsson CH (2016) A student-centered approach and mindset-focused pedagogical approach for entrepreneurship and leadership. Appl Innov Rev (2):57–63, June 2016
10. Sidhu I, Goubet JE, Weber H, Fred-Ojala A, Johnsson C, Pries JC (2016) Berkeley innovation index: an approach for measuring and diagnosing individual's and organizations' innovation capabilities. Berkeley Concept Paper, Feb 2016
11. Menold J, Purzer S, Ferguson DM, Ohland MW (2014) A critical review of measures of innovativeness. In: 121st ASEE annual conference & exposition
12. Fred-Ojala A, Eng Larsson J (2016) Construction of the Berkeley innovation index: a higher-order item response theory model approach. In: FMS820 20162 mathematical statistics, Lund University
13. Rao J, Weintraub J (2013) How innovative is your company's culture? MIT Sloan Manage Rev 54(3):29
14. Berkeley Innovation Index Survey (2018) URL: https://berkeleyinnovationindex.org/mindset/

Retrospective Analysis of Engineering Change Iterations—A Case Study on Reducing Engineering Changes by Defining Target Values for Engineering Design

Niklas Kattner, Aleksej Trepatschko, Gert Assmann, Lucia Becerril and Udo Lindemann

Abstract In the development of technical systems, engineering changes nowadays are more a rule than an exception and can take a big share of the overall capacity of a design project. Thus, reducing avoidable changes can have a huge impact the overall project performance. Hence, knowledge of companies about change iterations can reveal performance lacks in development projects. This paper therefore presents an industrial case study using past change information to reduce engineering change iterations. Hence, it briefly introduces the applied method and describes the procedure to reduce change iterations as well as the findings within the company's environment in detail.

Keywords Engineering change management · Engineering design
Engineering performance

1 Introduction

The raising global competition in the market as well as the shift from a seller's to a buyer's market result in more and more market oriented and customized solutions that have to go through an even shorter innovation cycle [1]. In addition, the performance and the complexity of the products continue to increase due to the integration of mechatronic systems, new safety and environmental requirements as well as the increased cost awareness [2]. These trends also effect the engineering

N. Kattner (✉) · A. Trepatschko · L. Becerril · U. Lindemann
Technical University of Munich, Munich, Germany
e-mail: niklas.kattner@tum.de

G. Assmann
Kelvion Holding GmbH, Bochum, Germany

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_30

change management. Since the quality, complexity and interconnectivity of technical systems increase, the quantity of engineering changes mirrors this trend. For Example, the complexity of modern technical systems results in an enormous effort to identify possible change propagation [3]. The handling of engineering changes can demand up to 40% of the resources and capacity of the overall design efforts [4]. Therefore, engineering changes are more the norm then the exception and the effect of engineering changes on the overall performance of the design process is crucial [5]. In addition, almost a quarter of the overall amount of changes are rated as unwanted or unnecessary. Hence, the reduction of engineering changes offers huge potential to increase the performance in the engineering design as well as decline the resources needed to conduct these changes [6].

2 Background

A big part of all technical product changes can be attributed to product design imperfections [1, 7]. This underlines the importance of the topic of technical product changes as part of product development. It further emphasizes the need for a closer look at the support of product developers with regard to avoiding errors and dealing with emerging product changes [8]. Using past change data to analyze the performance of design projects is promising for improving the development of technical systems. Since a documentation of change information is often mandatory due to regulations and law, there is a high chance these information is available within many companies [9]. Furthermore, the amount and type of occurring technical changes can be feedback to the quality of the development project and can therefore be a viable compass for the derivation of potential for reducing engineering changes and thus improve future development. Missing review loops within the design process, uncertain and undefined requirements and the lack of communication often lead to a number of technical changes [4]. Leveraging change information to identify key shortcomings of the design process causing many change iterations can therefore be crucial for optimizing the design performance. The avoidance of emerging critical changes provides great potential for improvement to free up resources and further reduce change propagation [10]. Research activities about analyzing engineering change information increases recently, nevertheless these approaches mainly focus on the data-mining techniques itself. However, literature about engineering change management lacks a methodical approach to analyze engineering change information retrospectively—even without change data available—with the objective to improve future developments [11]. To address the objective to reduce engineering changes, this paper gives a brief introduction into an approach to use the knowledge of past change iterations to reduce engineering changes in future engineering design situations. The application of the method itself in an industrial context as well as the generated results of the application is then discussed in detail.

3 Methodology

To identify potential to reduce engineering change iterations by analyzing information about past changes, the following procedure was applied in the use case (Fig. 1). The method for the retrospective analysis of change iterations essentially consists of four major steps. Within this case study, the method was applied in the industry to evaluate the applicability and success of the method. The company is a supplier in the mobility sector and has more than 10,000 employees. The system in focus was a mechatronic system the company sells as a product of their portfolio. Due to reasons of confidentiality, the result of the research and the use case is anonymized. The case however describes the actual activities and the procedure taken within the company to apply the method. Hence, within this specific use case, the steps of the method contain the following tasks.

Step 1 Preparing the current change situation and gather relevant information. The prevailing development process with its phases and reviews was analyzed and modeled based on Cooper [12]. In addition, the product structure was investigated and essential product areas were determined. We used the distinction of Pahl et al. [13] to separate the product assemblies into standard and variant areas. Hence, the focus of step 1 was the collection of information about the change information.

Step 2 Assessing the information and identify potential for reducing change iterations. The second step of the method addresses the assessment of the previously gathered information. To reveal potential for reducing change iterations, the information about the existing changes and the product structure were linked. As a result, a portfolio of the most critical iterations with promising characteristics pictured the change situation and change causes were identified.

Step 3 Implementing measures to address the previously revealed potential. The change causes were then investigated to define target values triggering and effecting the change causes. The relevant target values were

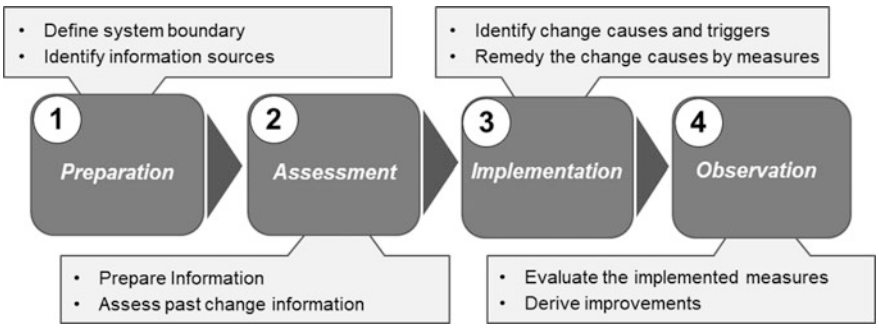


Fig. 1 Method for assessing past engineering change information

defined based on the changes made to the product areas. For this purpose, properties and characteristics of critical variant components are determined. Target values are defined from the selected properties. The target values monitoring was supported by a tool and review meetings for the evaluation of the target values were defined.

Step 4 Observing the implemented activities to reduce change iterations. After the implementation of a Tool addressing the reasons for the change iterations of the critical assembly, the effect of the tool is evaluated.

4 Applying the Method in an Industrial Use Case

The methodical procedure was applied using the example of a complex product development project in the supplier industry. In the following sections, the actual application of the method, the findings as well as the results of the application in the case study is therefore described in detail.

4.1 Preparation

The first step of the method creates the foundation of the following assessment. It includes gathering all relevant information to assess the situation about change iterations in the case. Since the available data set about engineer changes was not extensive, the lack of relevant change information was remedied by creating a big picture of the change situation, the design process, the product structure and interviewing the employees.

Analyzing the engineering design process. Investigating the actual process of designing the product helps to identify phases and milestones where changes occur. Hence, with experienced development engineers and experts, the entire project process was mapped (Fig. 2). Therefore, an image of the development process was created by leveraging the knowledge of experienced engineers and project managers as well as available internal process documents. The result was the documentation of the overall project plan as a stage-gate model according to Cooper [12]. To gather change information, the bottom level of the process (Development) was of particular importance.

The activities of the development process (bottom level) can be divided into three phases—design phase 1, design phase 2 and design phase 3. The level of detail of the product increases with each phase. The three phases contain decision points or milestones (DR 0-6) to observe the properties, i.e. the maturity of the product. Some of the reviews were not explicitly documented in the available process sheets (see in Fig. 2 the DR1 & DR2) and were revealed by interviews with the involved engineers (orange milestones). Thus, the identified review meetings,

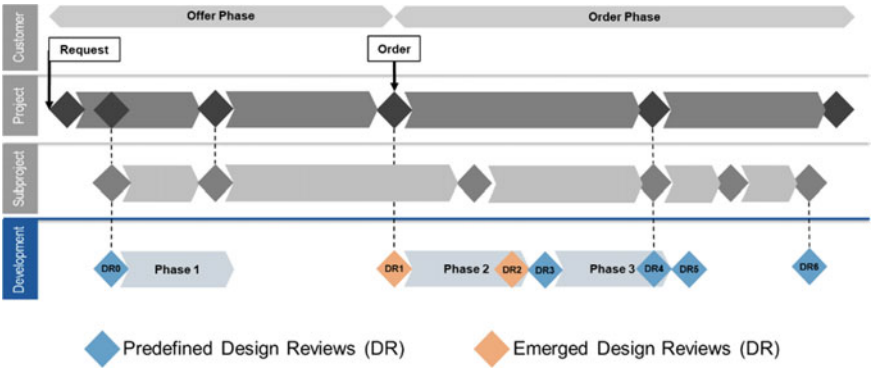


Fig. 2 Development process (bottom) embedded into the overall project schedule

its participants as well as the documents all contributed to the identification of the current change situation.

Analyzing the product structure. The product structure was examined for the selected product. The goal is to determine the components of the product that can be directly influenced by the engineers. The product structure in the case study consists of the following six assemblies: Assembly 1 (Combination of joined metal parts building the frame), Assembly 2 (Unit that provides the desired output to the customer), Assembly 3 (Unit that supports the output of Assembly 2), Assembly 4 (Assembly to meet special requirements), Assembly 5 (Combination of pipes, hoses and connections), Assembly 6 (Combination of cables and electrical units). According to Pahl et al. [13] product structure areas can be divided into variant and standard areas. Since the design of assemblies 1, 5 and 6 is project dependent, e.g. due to customer-specific installation space, and can be directly influenced by the engineers, these belong to the variance area of the product. Assemblies 2, 3, and 4 are used in other products of the company, i.e. are assemblies allocated to standard area, and therefore are excluded from the investigation.

Analyzing past engineering change information. After the evaluation of the product structure and the development process, the general change situation during engineering design is surveyed to identify changes iterations and eventually define target values. The identified development process, with the individual phases and milestones, contributed to identify the relevant people, documents and additional sources for the data collection of changes information. In detail, the analyzed product structure with the standard and variant areas is needed for determining and assigning the changes to the areas affected by changes. Thus, it is necessary to determine which variant areas are most frequently affected by change iterations and its causes. Change data, review protocols, project sheets and change instructions can be used as sources for collecting the quantitative change situation. Due to lack of extensive change documentation, a questionnaire was used. First, a qualitative overview of the change situation was provided, which should help to narrow down the issues. The aforementioned development process as a stage-gate model and the

specific product structure are used as a guide and to concretize the questions of the questionnaire. The questionnaire was prepared according to Stier [14] with both open and closed questions. Open questions are used to understand the change process and discover further room for improvement for the development situation. Nine persons (from engineers to divisional directors) participated in the questionnaire. The survey was conducted anonymously.

4.2 Assessment

After identifying and gathering information about change iterations, the information was processed by a detailed analysis of the change survey and the product structure. The result is a matrix comparing each object, i.e. assemblies, regarding its potential for reducing engineering changes. Therefore, attributes where defined in advance by the project responsible tailored to the specific environment of the company. Table 1 introduces an excerpt of the actual assessment matrix and shows the frequency of the changes as well as the resulting delay on the project. In addition, the area of the product structure, the influence of the engineer and the initiator are shown. Since some criteria are confidential, the portfolio is simplified.

Nevertheless, the portfolio emphasizes the procedure to reveal critical assemblies regarding the overall change situation. In the case study changes to assembly 1 have the biggest effect on the schedule and a high frequency of occurrence. Hence, according to the assessment, assembly 1 has the biggest potential to improve development performance by reducing engineering changes and is therefore selected for further investigation and the definition of the target values. The next step was the identification of the change cause manifesting in the symptom of recurring change iterations. A questionnaire revealed that the reduction of the property "cost" was the most common internal initiator for changes (as a boundary condition, only internal reasons for changes were considered). Thus, the critical changes to assembly 1 result from the target deviation of the property "costs". However, early tracking and controlling of this deviation can avoid changes

Table 1 Change iteration portfolio

Assembly	Area	Frequency of changes	Delay through changes	Initiator of the changes
Assembly 1	Variant	Often	Very strong	External and internal
Assembly 5	Variant	Often	Medium	External and internal
Assembly 6	Variant	Often	Medium	External and internal
Assembly 2	Standard	Occasional	Strong	Internal
Assembly 3 and 4	Standard	Occasional	Medium	External and internal

iterations. The property "cost" can be defined as a target-capable property whose characteristic values are scalable and clearly describable. This property "cost" is related to several characteristics that can influence the property. According to observations and discussions with experts, the values of the following characteristics are of major importance for the property "cost" of assembly 1 and can be determined and measured by the engineers (Table 2).

Since even moderately complex products can have several hundred characteristics (even for a single property), it is necessary to identify characteristics for the current situation with a big impact on the product property. Thus, the engineers can use the identified important characteristics to control the extent of the target value.

4.3 Implementation

Prepare measures to address the identified shortcomings. Since the consideration and collection of all characteristics would represent a very large effort for the engineers, only those are selected which have the greatest influence on the extent of the target value. To further reduce the number of characteristics, the supplier for certain components of assembly 1 was interviewed to define crucial characteristics and its value range on the property. The following characteristics are of particular importance to the supplier regarding the target values and therefore are key to reduce engineering changes: Number of single sheet metal parts, Number of laser cutouts, Number of bends, Length of fillet welds, and Length of HV welds. However, additional required characteristics that affect the property can be added on demand.

Apply measures. To utilize the previously defined characteristics, target values for these characteristics will be defined and tracked during the product development stages. Furthermore, the values will be used in design reviews to select between alternative solutions of concepts. This ensures that the selected property "costs" is monitored during the whole development process. If there is a possible deviation of the target values, the engineer can initiate a corrective action by regulating the values of the characteristics. In comparison to a single criterion assessment, the

Table 2 Characteristics

Metal sheet	Single parts	Jointing methods
Surfaces Thickness Materials	Sheet metal parts Equal parts Attachment parts	Screws Welding
Cut-outs	Geometry	Coatings
Drills Openings Contours	Bends Tolerances	Masking Powder coating Painting

target values allow multi criteria assessment that results in a better consideration of the system complexity. In order to be able to correlate the specific characteristics of the target values, they must be converted into dimensionless values. This enables the engineers to compare their designs either with the available alternatives and/or with previously defined goals for the characteristics. In addition, the individual characteristic values are normalized between 0 and 10 in relation to the respective maximum value.

For the implementation, a better handling and an easier tracking of the target values, a tool was created. The tool depicts the previously introduced target tracking and allows the user to assess its own concepts of assembly 1 (cf. Fig. 3).

Figure 3 introduces the GUI of the tool to support engineers in their development tasks. The visualization of the results facilitates the choice of the alternatives and the target tracking. The radar chart is a viable way of graphically visualizing the normalized values of the characteristics and the possible correlations between the characteristics. In this example, the alternative with the smallest area is the ideal. To further improve the tracking of the characteristics, the earliest time within the engineering design process, where the target values could be reviewed, was evaluated. Each target value then was allocated to one of the design review meetings. Therefore, remaining change iterations could be identified as early as possible with the design phase.

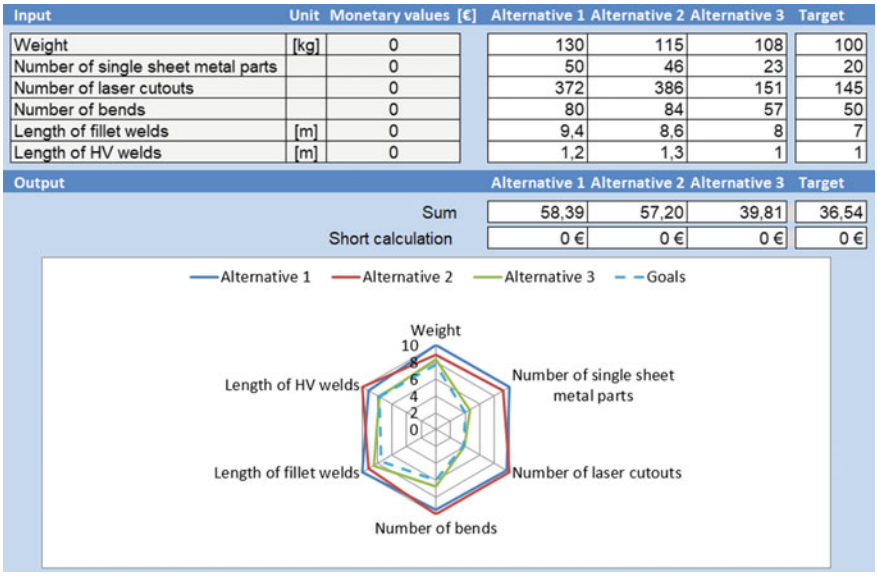


Fig. 3 GUI of the tool for engineers to assess the characteristics

4.4 Observation

After deriving and defining a measure based on the analyzed change situation, the effect of the measure was observed. The engineers gave feedback to the usability of the tool and their estimation of the potential for their future design tasks. In Addition, the overall quantity of change iterations has been monitored over time to identify the actual effect of the tool.

The users liked the simplicity of the tool and the possibility to assess their own concepts of assembly 1 in advance of review meetings. In addition, the feature to compare different concept of assembly 1 helps the engineers to choose the concept with the biggest potential regarding the target values previously introduced. Considering the overall amount of engineering changes, the change data reveals a noticeable drop in the number of engineering changes after the implementation of the tool. The plot (Fig. 4) pictures the overall effect of the applied method to reduce engineering changes on the engineering change quantity.

On average, approximately twenty percent of the engineering changes were avoided 4 months after the implementation of the tool. There is some fuzziness in the data since the number of ECR represents the change request of all assemblies within the product. Nevertheless, the project manager confirms the trend of a drop in the overall amount of changes of assembly 1. In addition, the engineers substantiate this statement and the trend of an overall reduction of engineering changes on assembly 1.

5 Discussion

The objective of the method to analyze iterations about past engineering changes is the identification of potential for future development project in past change information. Thereby, performance improvements can be derived by investigating change iterations. The application of the method in this use case was slightly refined. Since the available documentation about change iterations was not viable, the lack of information was remedied by analyzing the overall development

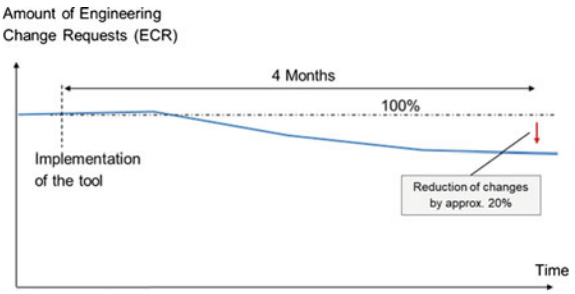


Fig. 4 Progression of ECR quantity after implementing the tool

procedure. In addition, conducting surveys include the engineers itself in the analysis. Hence, the development process and its review milestones revealed where changes occur. Furthermore, conducting a survey among the engineers extended the knowledge about change iterations on the product. Assessing the change iterations using the information gathered in the previous step has had some restrictions due to the case study environment. A major requirement for the application of the method was the applicability of the measures within the development unit responsible for the product. Thus, only variant assemblies under direct influence of the engineers were considered. Nevertheless, the benefit can increase by including standard assemblies into the assessment. This can lever the scope of effect of the measures on the quantity of change iterations.

Evaluation criteria revealed the actual change situation and change iterations of the assemblies. After choosing the assembly with the biggest potential for the reduction of engineering changes, the change causes were further investigated to define target values for future design workflows. Using target values derived from past change information to support engineers in the design phase has some benefits for the goal to reduce changes because it directly addresses triggers causing changes in the past. To embed this knowledge in the development process, a tool for monitoring and evaluating the target values was introduced. The engineers attribute the tool an improvement in future development projects since they can assess their own concept regarding the target values.

Hence, the application of the methodical approach for the investigation of technical systems regarding the change situation and its artefacts successfully resulted in an improved change situation. The method guided through the steps necessary to identify shortcomings in the design phase and supported the user in deriving measures to reduce engineering changes. The successful application of the method and its contribution to the identification of effective measures based on the investigation of past change information was also reflected by a decline of the ECR quantity after the implementation of the measures.

6 Conclusion and Outlook

Within the scope of this paper, a method for the investigation of past change information was applied in an industrial use case. The paper therefore briefly introduces the method and details its steps conducted in the specific use case context. Analyzing the development process revealed reviews and milestones where changes can occur. Additionally, the product structure was processed to define standard and variant areas. Assessing the gathered change information lead to the critical assembly where change causes and triggers have been further investigated in detail. In addition, target values where derived from the specific change triggers to address the change iterations and simplify the engineering design. After creating a tool for the engineers to monitor and evaluate the defined target values, the tool was implemented in the usual design process. Overall, the use of the method had a

positive effect on the change iterations and reduced the quantity of engineering changes. Despite the successful derivation of measures based on the assessment of past change information, the focus was specifically on the non-fulfilment of product properties. Further research could extend the assessment step to identify communication barriers within the company or investigate the structure of the design process to front-load the occurrence of changes. In Addition, an approach utilizing engineering change data derived from engineering change management tools to assess the change situation can further automate the assessment step.

Acknowledgements We thank the German Research Foundation (Deutsche Forschungsgemeinschaft—DFG) for funding this project as part of the collaborative research center „Sonderforschungsbereich 768—Managing cycles in innovation processes—Integrated development of product-service-systems based on technical products".

References

1. Pikosz P, Malmqvist J (1998) A comparative study of engineering change management in three Swedish engineering companies. In Proceedings of ASME international design engineering technical conferences, design theory and methodology
2. Ehrlenspiel K, Meerkamm H (2013) Integrierte Produktentwicklung: Denkabläufe, Methodeneinsatz, Zusammenarbeit. Hanser, München
3. Clarkson PJ, Simons C, Eckert C (2004) Predicting change propagation in complex design. J Mech Des 126(5):788–797
4. Maier A, Langer S (2011) Engineering change management report 2011: survey results on causes and effects, current practice, problems, and strategies in Denmark. Technical University of Denmark. DTU Management Engineering. Report 17
5. Hamraz B, Clarkson PJ (2015) Industrial evaluation of FBS Linkage—a method to support engineering change management. J Eng Des 26(1–3):24–47
6. Langer S, Maier A, Wilberg J, Münch T, Lindemann U (2012) Exploring differences between average and critical engineering changes: survey results from Denmark. In International design conference DESIGN. Dubrovnik, Coratia
7. Clark KB, Wheelright SC (1993) Managing new product and process development: text and cases. Free Press, New York
8. Köhler C (2009) Technische Produktänderungen – Analyse und Beurteilung von Lösungsmöglichkeiten auf Basis einer Erweiterung des CPM/PDD-Ansatzes. (Dissertation). Universität des Saarlandes
9. Elezi F, Sharafi A, Mirson A, Wolf P, Krcmar H, Lindemann U (2011) A knowledge discovery in databases (KDD) Approach for extracting causes of iterations in engineering change orders. ASME, 2011
10. Griffin A (1997) PDMA research on new product development practices: updating trends and benchmarking best practices. J Prod Innov Manage 14(6):429–458
11. Hamraz B, Caldwell NHM, Clarkson JP (2013) A holistic categorization framework for literature on engineering change management. Syst Eng 16(4):473–505
12. Cooper RG (2008) Perspective: the stage-gate ® Idea-to-launch process—Update, what's new, and NexGen systems. J Prod Innov Manage 25(3):213–232
13. Pahl G, Beitz W, Blessing L, Feldhusen J, Grote K-H, Wallace K (eds) (2007) Engineering design: a systematic approach, 3rd edn. Springer-Verlag, London Limited, London
14. Stier W (1999) Empirische Forschungsmethoden. Springer, Berlin, Heidelberg

Approximation Methods in Dantzig—Wolfe Decomposition of Variational Inequalities—A Review and Extension

William Chung

Abstract In this study, we review some approximation methods being used in Dantzig-Wolfe (DW) decomposition method for variational inequalities (VI). After applying DW decomposition method, the decomposed VI consists of one VI subproblem (*sub-VI*) and one VI master problem (*master-VI*). In each decomposition computational loop, we need to use an iterative method to solve both *sub-VI* and *master-VI* individually. To improve the computational efficiency, approximation methods in solving *sub-VI* or *master-VI* (not both) are used from the literature. Under the approximation methods, the approximate *sub-VI* is a LP or NLP. On the other hand, *master-VI* is approximately solved until a condition being met. Since both approximation methods for *sub-VI* and *master-VI* were developed separately, there is a knowledge gap that if both approximation methods can be applied at the same time in solving VI with DW decomposition method. The current study is to fill this gap. That is, we propose to apply both approximation methods of *sub-VI* and *master-VI* in one DW decomposition loop. An illustrative application is provided.

Keywords Approximation · Dantzig-Wolfe decomposition · Variational inequalities

1 Introduction

This study focuses on the application of an approximate Dantzig-Wolfe (DW) decomposition method [6] to solve a class of variational inequality (VI) problems, which may be described as follows: $VI(G, K)$: finding a vector $x^* \in K \subseteq R^n$, such that $G(x^*)^T(x - x^*) \geq 0 \; \forall x \in K$, where G is the given continuous mapping from K *to* R^n; superscript T denotes the transpose, and all vectors are

W. Chung (✉)
Department of Management Sciences, City University of Hong Kong, Hong Kong, China
e-mail: mswchung@cityu.edu.hk

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_31

considered column vectors. K is a non-empty, closed, and convex set. VI problems can be applied to energy equilibrium models [1], network equilibrium problems [14], and transportation planning models [16].

In the literature, a substantial number of techniques or algorithms are available for the numerical solution of $VI(G, K)$, e.g., by solving a sequence of optimization problems, similar to the Project Independence Evaluation Systems (PIES) algorithm [1], which can be classified as a special case of an iterative scheme for VI proposed by [5]. The interested reader may consult [9] and the references cited therein.

Meanwhile, decomposition methods allow large scale and/or complex $VI(G, K)$ to be solved in a distributed and parallel fashion. Moreover, they may also lead to a drastic simplification of the model development procedure and ease model management and maintenance [11–13]. In addition, applying existing solver, such as PATH [7], to solve a very large VI problem may be terminated because of lack of memory [10]. With these motivations, [2–4], and [10] developed the procedures to extend the application of DW decomposition principle from linear programming [6] to $VI(G, K)$, denoted as Dantzig-Wolfe decomposition of variational inequalities (*DW-VI*).

DW-VI consists of an equilibrium subproblem[1] (denoted as *sub-VI*^{k}) and an equilibrium master problem $\left(\textit{master-VI}^k\right)$, where superscript k represents the iteration number of *DW-VI*. The corresponding *DW-VI* algorithm is a computational sequence in solving *sub-VI*^{k} and *master-VI*^{k} alternatively (decomposition computational sequence). By contrast, two other computational sequences exist within the decomposition computational sequence. The reason is because VIs (*sub-VI*^{k} and *master-VI*^{k}) are to be solved within each decomposition computational sequence by solving a sequence of constrained optimization problems, such as PIES method [1]. Figure 1 summarizes the computational sequences of *DW-VI*.

Thus, for *DW-VI*, we consider that the decomposition computational sequence (decomposition loop) represents an outer layer sequence, whereas the VI computational sequence (VI loop) represents an inner layer sequence. We have one outer layer computational sequence for the decomposition method and two inner computational sequences for *sub-VI*^{k} and *master-VI*^{k} within each decomposition computational sequence.

Some approximation methods for *DW-VI* are developed to remove the inner computational sequences, thereby improving the computational performance of the *DW-VI* algorithm. Particularly, solving *sub-VI*^{k} and/or *master-VI*^{k} approximately in a way that the original VI can still be solved is ideal.

Sub-VI^k approximations are the first approximation methods of *DW-VI*. Chung and Fuller [3] developed four *sub-VI*^{k} approximation procedures in *DW-VI* for application to a multicommodity economic equilibrium model. Luna et al. [10] also

[1]Note that the subproblem can be further decomposed if the mapping and the feasible set of the subproblem are separable, like block-angular constraints in the subproblem of DW for linear programming.

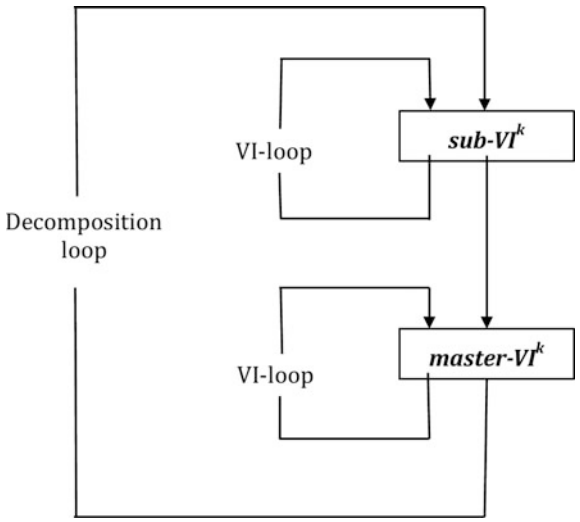


Fig. 1 Three computational sequences of *DW-VI*

studied the DW decomposition of VI problems and the approximation of *sub-VI*^{k}. They considered some algorithmic enhancements, including an inexact solution of the approximate *sub-VI*^{k} and the cheap generation of additional proposals through a projection method. Overall, no *master-VI*^{k} approximation is considered in their works.

Concerning *master-VI*^{k} approximations, Çelebi and Fuller [4] discussed the heuristic procedure of *master-VI*^{k} approximation in DW of VI. At each iteration k of *DW-VI,* the *master-VI* mapping $G(x)$ is approximated with $\widetilde{G}^k(x)$. No *sub-VI*^{k} approximation is considered. Moreover, Çelebi and Fuller [4] proposed to solve *master-VI* approximately by running a few iterations of the VI computational sequence until the desired approximation accuracy is obtained. The motivation of the master problem approximations is to alleviate their computational difficulties in solving the original master problem of *DW-VI* in [2]. Çelebi and Fuller [4] reported that they cannot reach an equilibrium solution with the standard *DW-VI* algorithm for their time-of-use (TOU) pricing models (with a 66-bus network in Nash–Cournot game setting with line limits) within a reasonable time framework (48 h).

Since Chung and Fuller [3] and Çelebi and Fuller [4] developed their approximation methods for *sub-VI*^{k} and *master-VI*^{k} respectively, there is no research work investigating if these two approximation methods can be used simultaneously for solving VI. Motivated by this knowledge gap, we develop a new approximate *DW-VI* (ADW-VI) method in which a kind of iterative solution methods for VI and DW decomposition method are merged into an iterative solution method. That is, the computational sequences of both *sub-VI*^{k} and *master-VI*^{k} can be eliminated within each decomposition step, and the resulting *ADW-VI* method shows one computational sequence. An illustrative application is provided.

We first present *DW-VI* applying to a class of VI problems with linking constraints in Sect. 2. Based on this background, the Sub-VI approximation methods of [3] and the Master-VI approximation methods of [4] are described in Sect. 3. Section 4 extends the application of both approximation methods: applying both approximation methods in a DW-VI loop. In Sect. 5, both decomposition methods are applied to compute solutions to a numerical example. We summarize our results and present conclusions in Sect. 6.

2 Dantzig-Wolfe of VI (*DW-VI*)

Fuller and Chung [8] have defined the ***DW-VI***, which is described below as a background of development of approximate version of *DW-VI* (***ADW-VI***).

All vectors are column vectors, and the symbol "T" as a superscript indicates the transpose of a vector or matrix. Let G be a mapping from R^n into itself. The feasible set is

$$K = \{x \in R^n | g(x) \geq 0, h(x) \geq 0\}$$

where g is a mapping from R^n into R^m such that g_i is concave and continuously differentiable for all $i = 1, \ldots, m$, and h is a mapping from R^n into R^l such that h_i is concave and differentiable for all $i = 1, \ldots, l$. Then, ***VI*** is defined as

$VI(G, K)$: finding $x^* \in K$, such that

$$G(x^*)^T(x - x^*) \geq 0 \forall x \in K \tag{1}$$

We assume throughout this study that Eq. (1) has at least one solution.

The constraints $h(x) \geq 0$ are the difficult (or linking) constraints, which occur in the master problem but not the subproblem of the decomposition algorithm.

2.1 *Subproblem (*Sub-VI$^{\mathbf{k}}$*)*

The subproblem at iteration k is defined with information from the master problem solved at iteration $k-1$: π^{k-1} (the dual variable vector to $h(x) \geq 0$) and $\nabla h\left(x_M^{k-1}\right)$ (the matrix of gradients of h evaluated at the master problem solution x_M^{k-1}). The feasible set of the subproblem is defined as a relaxation of K obtained by dropping $h(x) \geq 0$: $\widetilde{K} = \{x \in R^n | g(x) \geq 0\}$. *Sub-VI*^{k} is defined as follows:

sub-VI$^k\left(G - \nabla h\left(x_M^{k-1}\right)^T \pi^{k-1}, \overline{K}\right)$: finding $x_S^k \in \widetilde{K}$ such that

$$\left(G\left(x_S^k\right) - \nabla h\left(x_M^{k-1}\right)^T \pi^{k-1}\right)^T \left(x - x_S^k\right) \geq 0 \forall x \in \overline{K}.$$

The feasible set of the master problem at iteration k is defined by convex combinations of the k solution (or "proposals") that have been calculated by the first k solution of the subproblem. An additional restriction is that the constraints, $h(x) \geq 0$, need to be satisfied. We use the notation $X^k = \left[x_S^1, x_S^2, \ldots, x_S^k\right]$ to represent matrices whose columns are the k proposals collected from the subproblem at each iteration. The variables of the master problem are in the vector $\lambda \in R_+^k$ of weights on the proposals in the convex combination. The feasible set of the master problem $\Lambda^k = \left\{\lambda \in R_+^k | h\left(X^k \lambda\right) \geq 0, \left(e^k\right)^T \lambda = 1\right\}$, where $e^k \in R^k$ is a vector whose k entries are all one.

2.2 *Master Problem (*Master-VI$^{\mathbf{k}}$*)*

Master-VI^k at iteration k is defined as

master-VI$^k\left(H^k, \Lambda^k\right)$: finding $\lambda^k \in \Lambda^k$ such that $H^k\left(\lambda^k\right)^T \left(\lambda - \lambda^k\right) \geq 0 \forall \lambda \in \Lambda^k$, where the mapping H^k from R^k to R^k is defined by $H^k(\lambda)^T = G\left(X^k \lambda\right)^T X^k$.

For brevity, the notation x_M^k is sometimes used to express the solution of the master problem, λ^k, in terms of the original x variables: $x_M^k = X^k \lambda^k$, and the feasible set for *master-VI*^{k} can be alternatively denoted by K^k illustrated as follows:

$$K^k = \left\{x \in R^k | h(x) \geq 0, x \in conv\left(X^k\right)\right\},$$

where $conv\left(X^k\right)$ is the convex hull of the points represented by X^k. Thus, *master-VI*^{k} can be alternatively defined as

Master-VI$^k\left(G, K^k\right)$: finding $x_M^k \in K^k$, such that $G\left(x_M^k\right)^T \left(x - x_M^k\right) \geq 0 \forall x \in K^k$.
Note that $x_M^k \in K \subset \overline{K}$, and $K^1 \subseteq K^2 \subseteq \cdots \subseteq K^{k-1} \subseteq K^k \subseteq \cdots \subseteq K \subseteq \overline{K}$.

2.3 *Stopping Condition*

The algorithm starts by solving *sub-VI*^{1} based on a guess of the value of the mapping adjustment, such as $\nabla h\left(x_M^0\right)^T \pi^0 = 0$. Then, *master-VI*^{1} is solved (normally with artificial variables ensuring feasibility of the linking constraints). Later iterations begin with a subproblem and end with a master problem.

Convergence of the master solution x_M^k to a solution of $VI(G,K)$ is assessed by solving the next subproblem and computing a scalar quantity called the convergence gap

$$CG^k = \left(G\left(x_M^k\right) - \nabla h\left(x_M^k\right)^T \pi^k\right)^T \left(x_S^{k+1} - x_M^k\right).$$

The algorithm terminates when a predetermined convergence tolerance, $\varepsilon > 0$, is reached, i.e., $\left|CG^k\right| < \varepsilon$. The standard *DW-VI* algorithm is as follows:

[Standard *DW-VI* algorithm]

Step 0 Set $k = 0$. Choose $\nabla h\left(x_M^0\right)^T \pi^0$, $\varepsilon > 0$ and set X^0 to be the empty matrix.

Step 1 Solve *sub-VI*$^{k+1}\left(G - \nabla h\left(x_M^k\right)^T \pi^k, \overline{K}\right)$. If $k = 0$ and the subproblem is infeasible, then stop; else place the solution x_S^{k+1} in the matrix $X^{k+1} = \left[X^k, x_S^{k+1}\right]$. If $k = 0$, then proceed to Step 2; else if $CG^k > -\varepsilon$, then stop; else proceed to Step 2.

Step 2 Increment $k \leftarrow k+1$. Solve *master-VI*$^k\left(H^k, \Lambda^k\right)$. Record $\nabla h\left(x_M^k\right)^T \pi^k$. Proceed to Step 1.

We assume that $\overline{K}$ is bounded, $\{\pi^k\}_{k=1}^{\infty}$ is contained in a bounded set, and G is continuous and strictly monotone. Fuller and Chung [8] prove that if $CG^k \geq 0$, x_M^k solves $VI(G,K)$. Before convergence, $CG^k < 0$, and $\lim_{k\to\infty} CG^k = 0$.

Note that *sub-VI*^{k} and *master-VI*^{k} are VI problem, which requires computing a sequence of optimization problems to solve [5].

3 Approximation Methods in a *DW-VI* Loop

In this section, we discuss the modifications of the standard *DW-VI* by allowing both *sub-VI*^{k} and *master-VI*^{k} to be solved approximately in a DW-VI loop. We define the approximate subproblem, approximate master problem, and the stopping condition in the following subsection.

3.1 Approximating Sub-VI^k

Following the results of [3], we can consider an approximation to the mapping G in *sub-VI*^{k}. The approximation at iteration k relies on parameters in vector ξ^{k-1}, which can include the most recent solution of *master-VI*^{k}, x_M^{k-1}. We write $\widetilde{G}_S^k\left(x, \xi^{k-1}\right)$ and require that $\widetilde{G}_S$ is strictly monotone in x and $\widetilde{G}_S^k\left(x_M^{k-1}, \xi^{k-1}\right) = G\left(x_M^{k-1}\right)$. *Asub-VI*^{k} can be defined as follows:

$$Asub\text{-}VI^k\left(G_S^k - \nabla h\left(x_M^{k-1}\right)^T \pi^{k-1}, \overline{K}\right)\text{: finding } x_S^k \in \overline{K} \text{ such that}$$
$$\left(\widetilde{G}_S^k\left(x_S^k, \xi^{k-1}\right) - \nabla h\left(x_M^{k-1}\right)^T \pi^{k-1}\right)^T \left(x - x_S^k\right) \geq 0 \forall x \in \overline{K}.$$

Note that *master-VI*^{k} in [3] is a VI, which is not similar to the one we proposed, i.e., that *master-VI*^{k} is approximately solved with *Amaster-VI*^{k}.

3.2 *Approximating* Master-VI^k

At each iteration k of *DW-VI*, the master problem mapping $G(x)$ is approximated by $\widetilde{G}_M^k(x)$, and the corresponding approximated master problem mapping is solved and denoted by $Amaster\text{-}VI^k\left(\widetilde{H}^k, \Lambda^k\right)$, where $\widetilde{H}^k(\lambda) = \widetilde{G}_M^k(X^k\lambda)^T X^k$. We assumed that $lim_{k\to\infty}\left|\widetilde{G}_M^k\left(x_M^k\right) - G\left(x_M^k\right)\right| = 0$, which is the key property to have convergence theory of this approximate algorithm and used in Çelebi and Fuller [4]. In addition, the approximate mapping $\widetilde{G}_M^k(x)$ should be continuous, and we can define *Amaster-VI*^{k} with the approximate mapping as follows:

$Amaster\text{-}VI^k\left(\widetilde{H}^k, \Lambda^k\right)$: finding $\lambda^k \in \Lambda^k$ such that $\widetilde{H}^k\left(\lambda^k\right)^T\left(\lambda - \lambda^k\right) \geq 0 \forall \lambda \in \Lambda^k$, or equivalently

$Amaster\text{-}VI^k\left(\widetilde{G}_M^k, K^k\right)$: finding $x_M^k \in K^k$, such that $\widetilde{G}_M^k\left(x_M^k\right)^T\left(x - x_M^k\right) \geq 0 \forall x \in K^k$.

Note that *sub-VI*^{k} in [4] is a VI, which is not similar to the one we propose that *sub-VI*^{k} is approximately solved with *Asub-VI*^{k}.

3.3 *Stopping Conditions in Approximation Methods*

With *Asub-VI*^{k} of [3], the algorithm terminates when a predetermined convergence tolerance, $\varepsilon > 0$, is reached, i.e., $CG^k < \varepsilon$.

With *Amaster-VI*^{k} of [4], two stopping conditions exist. It is because the algorithm with *Amaster-VI*^{k} combines two convergent processes as follows: the DW algorithm of using subproblem proposals to approximate the master problem's feasible region and a steadily improving approximation of the master problem's mapping G. For the first convergent process, the convergence gap, CG^k, is used as the corresponding stopping condition. The algorithm terminates when a predetermined convergence tolerance, $\varepsilon > 0$, is reached, i.e., $\left|CG^k\right| < \varepsilon$. For the second, G approximation gap, $\left\|\widetilde{G}_M^k\left(x_M^k\right) - G\left(x_M^k\right)\right\| < \alpha\left\|\widetilde{G}_M^{k-1}\left(x_M^{k-1}\right) - G\left(x_M^{k-1}\right)\right\|$ is used where $\alpha \in (0, 1)$.

4 *DW-VI* Algorithm with Both Sub-VI and Master-VI Approximations

By combining both approximations into standard DW-VI, we define the following approximate DW-VI algorithm.

[*DW-VI* algorithm with approximations]

Step 0 Set $k = 0$. Choose $\nabla h\left(x_M^0\right)^T \pi^0$, $\varepsilon > 0$, $\alpha \in (0, 1)$, M = large number, set X^0 to the null matrix, and $CG^0 = -\infty$.

Step 1 Increment $k \leftarrow k+1$. Solve $Asub\text{-}VI^k\left(\widetilde{G}_S^k - \nabla h\left(x_M^{k-1}\right)^T \pi^{k-1}, \overline{K}\right)$ from the previous master problem (or Step 0) and place the solution x_S^k in the matrix $X^k = \left[X^{k-1}, x_S^k\right]$. If $k = 1$, then proceed to Step 2; else if $CG^k > -\varepsilon$, and $\left(\widetilde{G}_M^{k-1}\left(x_M^{k-1}\right) - G\left(x_M^{k-1}\right)\right)^T \left(x_S^k - x_M^{k-1}\right) < \varepsilon$ then stop; else proceed to Step 2.

Step 2 Choose $\widetilde{G}^k$ and calculate x_M^k such that x_M^k solve $Amaster\text{-}VI^k\left(\widetilde{G}_M^k, K^k\right)$ with $\left\|\widetilde{G}_M^k\left(x_M^k\right) - G\left(x_M^k\right)\right\| < \alpha \left\|\widetilde{G}_M^{k-1}\left(x_M^{k-1}\right) - G\left(x_M^{k-1}\right)\right\|$, then record $\nabla h\left(x_M^k\right)^T \pi^k$ and proceed to Step 1.

Note in Step 2, we may need to use the iterative method to solve $Amaster\text{-}VI^k\left(\widetilde{G}_M^k, K^k\right)$ a few times until we have $\left\|\widetilde{G}_M^k\left(x_M^k\right) - G\left(x_M^k\right)\right\| < \alpha \left\|\widetilde{G}_M^{k-1}\left(x_M^{k-1}\right) - G\left(x_M^{k-1}\right)\right\|$.

5 An Illustrative Application

To illustrate the use of the algorithm presented in the previous section, a numerical example is presented to illustrate the application of the approximation method for DW-VI. This example is based on Example 2 in [15]. Example 2 is a multiproduct multipollutant oligopolistic market model with ambient-based pollution permits and transaction costs. To this example, we add an upper limit on the summation of the demand of firms for each product. Without this constraint, no relationship between *sub-VI* and *master-VI* exists for the property of G.

The example is coded in GAMS using Windows 7 Enterprise with an Intel Core i7 processor and 8 GB of memory. To obtain reference numerical results, we first solved each example by PATH without decomposition. The example consists of monotone G same as the one in [15]. For DW-VI, we used PIES algorithm to solve *master-VI* with the convergence condition that more than 0.0001 should be present in the difference of any demand from one PIES iteration to the next. 13 decomposition steps are performed, and 609 nonlinear programs within the PIES algorithm are solved. That is, a total of 13 LP Sub-VI and 609 NLP Master-VI are solved to obtain the solution with relative $CG^k \geq -0.0001$.

Table 1 Progress of iterations of approximate DW-VI in the example

Iteration	$\lvert\tilde{\pi}^k - \pi^k\rvert$	CG^k
2	−19972.8	−3352.39
3	7.76	−1021.15
4	−2.411	−581.209
5	−1.852	−111.394
6	−0.021	−103.521
7	−0.047	−92.471
8	−0.071	−116.052
9	0.032	−11.001
10	0.003	−17.244
11	−0.003	−12.571
12	−1.32E−04	−2.913
13	2.92E−04	−0.587
14	1.48E−04	−0.106
15	3.80E−05	−0.024
16	1.15E−05	−0.005
17	2.61E−06	−0.001
18	6.31E−07	−2.83E−04
19	1.52E−07	−6.52E−05

For Approximate methods, we combine the PIES method and the DW method. It takes 19 decomposition steps, that is, 19 LP Sub-VI and 19 NLP Master-VI with $CG^k \geq -0.0001$, and $\left|\tilde{\pi}^{k-1} - \pi^{k-1}\right| < 0.0001$. In terms of the number of solved NLPs and LPs, Approximate DW-VI is better than DW-VI. As Murphy and Mudrageda [11] pointed out, although PIES method does not meet the convergent conditions of PIES provided by Ahn and Hogan [1], PIES usually does not fail to converge. On the other hand, Table 1 shows the progress of iterations of approximation DW-VI in this example.

6 Conclusions

In this paper, we first review the approximation methods to be used in solving VI with DW decomposition method. We find that there are two approximation methods, one for solving the subproblem and the other for master problem, in order to improve the computational efficiency. However, each approximation method is developed individually. There is no literature related to using both approximation methods within a DW decomposition loop. Hence, we extend these approximation methods and consider using both of them in a DW decomposition loop. The resulting algorithm was implemented to solve a small application example with positive results. Further research would include convergence properties of the resulting algorithm and computational efficiency in large-scale VI.

Acknowledgements Financial support for Chung's work came from the Research Grants Council of Hong Kong S.A.R., China (CityU 11505016).

References

1. Ahn B, Hogan WW (1982) On convergence of the PIES algorithm for computing equilibria. Oper Res 30(2):281–300
2. Chung W, Fuller JD, Wu Y (2006) A new decomposition method for multiregional economic equilibrium models. Oper Res 54(4):643–655
3. Chung W, Fuller JD (2010) Subproblem approximation in Dantzig-Wolfe decomposition of variational inequality models with an application to a multicommodity economic equilibrium model. Oper Res 58(5):1318–1327
4. Çelebi E, Fuller JD (2013) Master problem approximations in Dantzig-Wolfe decomposition of variational inequality problems with applications to two energy market models. Comput Oper Res 40:2724–2739
5. Dafermos S (1983) An iterative scheme for variational inequalities. Math Program 26:40–47
6. Dantzig GB, Wolfe P (1961) The decomposition algorithm for linear programs. Econometrica 29(4):767–778
7. Dirkse SP, Ferris MC (1996) A path search damped newton method for computing general equilibria. Ann Oper Res 68(2):211–232
8. Fuller JD, Chung W (2005) Dantzig-Wolfe decomposition of variational inequalities. Comput Econ 25(4):303–326
9. Harker PT, Pang JS (1990) Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications. Math Program 48(1):161–220
10. Luna J, Sagastizábal C, Solodov M (2014) A class of Dantzig-Wolfe type decomposition methods for variational inequality problems. Math Program Ser A 143:177–209
11. Murphy FH, Mudrageda MV (1998) A decomposition approach for a class of economic equilibrium models. Oper Res 46(3):368–377
12. Murphy FH (1993) Making large-scale models manageable: Modeling from an operations management perspective. Oper Res 41(2):241–252
13. Murphy FH, Pierru A, Smeers Y (2016) A tutorial on building policy models as mixed-complementarity problems. Interfaces 46(6):465–481. https://doi.org/10.1287/inte.2016.0842
14. Nagurney A (1993) Network economics: a variational inequality approach. Kluwer Academic Publishers, Boston, MA
15. Nagurney A, Dhanda KK (2000) Marketable pollution permits in oligopolistic markets with transaction costs. Oper Res 48(3):424–435
16. Patriksson M (1994) The traffic assignment problem: models and methods. VSP, The Netherlands

The Route Planning of Advisors for the RMUTL Cooperative Education Program

Parida Jewpanya

Abstract The Rajamangala University of Technology Lanna (RMUTL) Cooperative Education Program (CO-OP) gives students an opportunity to receive career training in industries. The students in the final academic year are required to attain this program for a semester. The RMUTL CO-OP has a team of advisors who need to visit those students in the industries at least two times during the internship period to give the students advice on projects and working life. The current plan is created based on the planner experience which sometimes ineffective in term of costs. Therefore, we formulated a mixed-integer linear programming model to help the planner to find an optimal route for the advisors in RMUTL CO-OP. The objective is to minimize the total cost of the advisor's visitation. Moreover, results show that the proposed model can obtain better solutions than the current approach.

Keywords Cooperative education program · Vehicle routing problem

1 Introduction

The Rajamangala University of Technology Lanna (RMUTL) Cooperative Education Program (CO-OP) is a paid internship program that gives students an opportunity to receive career training as they work with professionals in their major fields of study [1]. The students in this program are encouraged to build essential knowledge and skills such as teamwork and problem-solving in the real world experience. Moreover, it supports the development of graduates and making them more employable and adaptable at the workplace [2].

The organization in CO-OP program consists of competitive industry leaders and higher education institutions "cooperating" with each other to provide hands-on work experience to full-time actively enrolled students within a degree-seeking program [3].

P. Jewpanya (✉)
Department of Industrial Engineering, Rajamangala University of Technology Lanna Tak, Tak 63000, Thailand
e-mail: parida.jewpanya@gmail.com

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_32

In RMUTL CO-OP, the students in the final academic year are required to attain this program for four months (one semester). The students are eligible to apply to CO-OP after completing a minimum requirement of the university.

The RMUTL CO-OP allows the students to select the industry by themselves under the direction of the department's CO-OP advisors. Then, the RMUTL CO-OP coordinator will contact the chosen industry and process the application. During the students 'internship' period, the industries are requested to assign students the project related to their field of study. This project offers students the opportunity to gain real-world experience in both theory and practical skill.

The RMUTL CO-OP has a team of advisors who come from different departments and have different expertise. These advisors are required to visits the students in the industries at least two times during the students' internship period. This activity is very important for the RMUTL CO-OP program. The purpose of this visit is to give the students advice on their projects and working life. Moreover, the visit can bring out the stronger relationship between university and industries. More than 60% of the total program budget is spent for this activity. The CO-OP coordinators have a responsibility to arrange the visiting plan for advisors in each department. The current plan has manually created a base on the planner experience which sometimes ineffective in term of costs.

The coordinators desire a good solution to help them make significant improvements on planning this activity. Due to the dynamic of the visiting locations, there is no methodology or repeatable process that available for a coordinator which can be applied directly. Therefore, in this paper, we propose the model for assigning and routing advisors visitation to industries. The proposed model is modified from the vehicle routing problem (VRP) model which can be described as the problem of designing optimal delivery or collection of routes from a depot to some customers subject to side constraints [4, 5]. In the supply chain, the classical vehicle routing problem (VRP) plays an important role in distribution management and logistics, as well as the costs associated with operating vehicles [4]. In using the model to create a routing plan, we incorporate guidelines concerning capacity, flow balance, and available travel schedules of advisors. The model's objective is to minimize the total cost. Although routing problems have been studied and solved using a variety of methods, the proposed problem has several unique requirements.

This paper organized as follow. In Sect. 2, an overview of the current approach for generating the route of RMUTL CO-OP advisor visitation is presented. Section 3 describes the model formulation of RMUTL CO-OP advisor visitation routing problem. Experimental result and comparison with the other approach are presented in Sect. 4. Finally, conclusions and future research are identified in Sect. 5.

2 Current Approach

The RMUTL CO-OP Coordinators manually arrange the routing plan for the advisors using a greedy method of visit allocation. The steps of this method are described as follows.

Step 1 Gather all the new information for the current period, including data of factory locations. The coordinator determined the amount of time each person can work.
Step 2 Clusters the industries based on a region. For example, the industries that located in the same province are assigned into the same cluster.
Step 3 Assign a vehicle route to each cluster. The order of visiting is arbitrary.
Step 4 Determine the number of days and travel cost of each route.

Figure 1 shows the example of the current approach result. There are nine locations need to be visited. In this case, the locations are grouped into three clusters, and the route is assigned to each cluster. The order of visiting places in each route is arbitrary.

3 A New Approach: Routing Model

This study considers the vehicle routing problem. A set of routes is constructed such that all industries will be visited. Each industry can be visited at most once. The starting point and the ending point of the routes are fixed.

Set:

V Set of available vehicles $v = 1, 2, \ldots, |V|$
F Set of industries $i = 1, 2, \ldots, |F|$
D Set of a depot $i = 0$

Parameters:

t_i Visiting time at industry i; $i \in F$
d_{ij} Travel distance from industry i to j; $i, j \in F \cup D$
t'_{ij} Travel time from industry i to j; $i, j \in F \cup D$
$VCAP$ Number of industries allowed per route
$CDIST$ Cost per unit distance
$CRENT$ Cost of renting vehicle per day

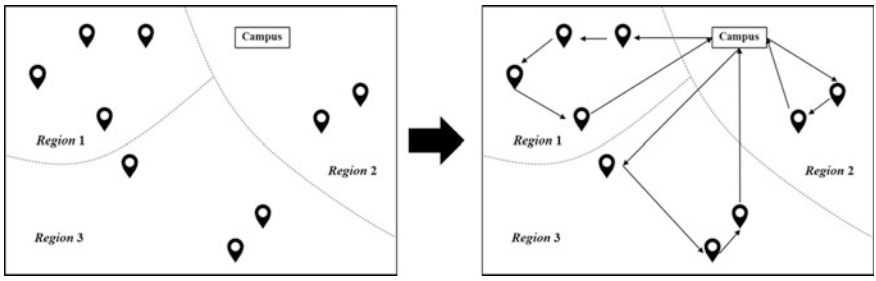


Fig. 1 Example of the recent approach

$CADV$ Allowance per day
$CHOT$ Cost of renting hotel per day
N Number of advisors assigned in each route
W Working time per day
M Big number

Variables:

D_v Number of day of each route
S_{iv} Start time of a visit to industry i in route v

$$x_{ijv} = \begin{cases} 1 & \text{If a visit to industry } i \text{ is followed by a visit to industry } j \text{ in route v} \\ 0 & \text{Otherwise} \end{cases}$$

The objective of the problem is to determine the number of routes and the best vehicle routes for advisors to visit industries. The sum of the travel costs is minimized consisting of the cost of vehicles, the allowance and hotel renting cost of advisors.

$$Min \sum_{\forall i,j \in F \cup D} \sum_{v \in V} x_{ijv} \cdot d_{ij} \cdot CDIST + \sum_{\forall v \in V} D_v \cdot (CRENT + N \cdot CADV) + \sum_{\forall v \in V} (D_v - 1) \cdot N \cdot CHOT$$

Constraints:

$$\sum_{\forall k \in F \cup D} \sum_{\forall v \in V} x_{k,k,v} = 0 \tag{1}$$

$$\sum_{\forall j \in F, i \neq j} x_{j,i,v} - \sum_{\forall j \in F, i \neq j} x_{i,j,v} = 0, \quad \forall i \in F, \forall v \in V \tag{2}$$

$$S_{jv} \geq S_{iv} + t_j + t'_{ij} - M(1 - \mathrm{x_{ijv}}), \quad \forall i,j \in F, \forall v \in V \tag{3}$$

$$S_{jv} \geq (t_j + t'_{0j}) \cdot \mathrm{x_{0jv}}, \quad \forall j \in F, \forall v \in V \tag{4}$$

$$\sum_{\forall i \in F \cup D} x_{i1v} \leq 1, \quad \forall v \in V \tag{5}$$

$$\sum_{\forall i \in F \cup D} \sum_{\forall v \in V} x_{ijv} = 1, \quad \forall j \in F \tag{6}$$

$$\sum_{\forall i \in F \cup D} \sum_{\forall j \cup F} x_{ijv} \leq VCAP, \quad \forall v \in V \tag{7}$$

$$D_v \cdot W \geq S_{jv} + t'_{j0} \cdot x_{j1v}, \quad \forall j \in F, \forall v \in V \tag{8}$$

Constraint (1) guarantees that a vehicle does not travel inside itself. Constraint (2) is the connectivity of each route. Constraint (3) determines the timeline of each route. Constraint (4) makes sure that the start service time at the first visited industry of each route should be greater than the traveling time from depot 0 to *j*. Constraint (5) ensures that vehicle *v* can be used only one time. Constraint (6) provides thatn industry is visited at most once. Constraint (7) confirms that the total visited industry does not exceed the number of industries allowed in each route. Constraint (8) determines the number of the day used in each route.

4 Computation Results

We used the routing model for all four semesters. Because of the model's size and complexity, we could not achieve optimality in any model run. To obtain our solutions, we used a LINGO version 12 solver running on an Intel PC with a 3.4 GHz processor and 4 GB of RAM, under the Windows 7 operating system.

4.1 Test Instances

We collected the data from the industrial engineering department of RMUTL, Tak campus, which has students attend the CO-OP program in every semester. The data are collected starting with the fall semester of 2016 academic year, and ending with the spring semester of 2017 academic year; this equates to 4 semesters. The data consists of the set of locations and set of routes. The number of factories and other problem parameter values is shown in Table 1.

4.2 Result Comparison

The results from the new model are compared with those results obtained from the current approach. The commercial solver is terminated after two hours if it cannot find an optimal solution. Table 2 shows that using the manual method, total costs

Table 1 The problem parameter values

No.	Name	No. of industries	Cost (Thai Baht.)						
			VCAP	*CDIST*	*CADV*	*CHOT*	*N*	*W*	*CRENT*
1	Fall2016	17	6	4	240	750	2	12	1800
2	Spring2016	9	5						
3	Fall2017	8	3						
4	Spring2017	11	3						

Table 2 The table shows total costs from the current approach and the proposed model, and the resulting savings percentages

No.	Name	Total cost (Thai Baht.)		
		Current approach	New approach	Saving (%)
1	Fall2016	55576.8	43378.4	21.9
2	Spring2016	31347.2	24589.6	21.6
3	Fall2017	37336.4	26432.0	29.2
4	Spring2017	45941.2	38965.2	15.2
Average		42550.4	33341.3	22.0

for the four periods were 42550.4 Baht; using the MILP model, they were 33341.3 Baht—a 22% reduction.

4.3 Sensitivity Analysis

Sensitivity analysis, which can be used to improve management decision making or simulate process changes, is also a benefit of using the model. For this paper, we conducted sensitivity analysis on the impact of the number of industries allowed in each route to provide more insight on the problem. Two data are tested with six values of *VCAP*. Those data are *Spring2016* and *Fall2017*. Figure 2 provides results on total costs for *Spring2016* and *Fall2017*. In summary, the higher value of *VCAP* can give the lower total costs.

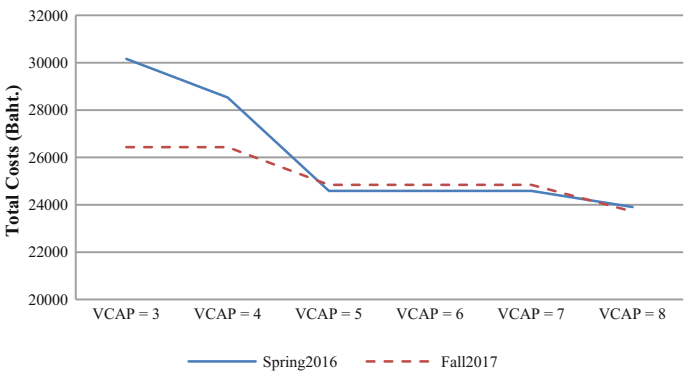


Fig. 2 Sensitivity analysis on the number of industries allowed in each route

5 Conclusion

This paper has proposed a new model for routing the advisors in the RMUTL CO-OP program. The model has applications in a wide range of settings, including planning, and strategic environments. Moreover, the model gives consulting firm management a method for calibrating the impact of a number of industries allowed in each route on the total costs.

When the RMUTL implemented the solution in its routing planning process, it realized travel cost savings. These savings have nearly achieved the projected 22% total costs reduction. Moreover, the sensitivity analysis shows that when the number of factories allowed in each route is increasing, it causes the lower total cost.

References

1. Mason RE (1989) Cooperative occupational education and work experience in the curriculum. ERIC
2. Blair BF, Millea M, Hammer J (2004) The impact of cooperative education on academic performance and compensation of engineering majors. J Eng Educ 93(4):333–338
3. Thiel GR, Hartley NT (1997) Cooperative education: a natural synergy between business and academia. SAM Adv Manag J 62(3):19
4. Barbarosoglu G, Ozgur D (1999) A tabu search algorithm for the vehicle routing problem. Comput Oper Res 26(3):255–270
5. Eksioglu B, Vural AV, Reisman A (2009) The vehicle routing problem: a taxonomic review. Comput Ind Eng 57(4):1472–1483

Optimal Design of Pipe Diameter in Water Distribution System by Multi-objective Differential Evolution Algorithm: A Case Study of Small Town in Chiang Mai

Warisa Wisittipanich and Dollaya Buakum

Abstract Water distribution system (WDS) plays a vital role in supplying water for population living in the city and urban areas. An expensive infrastructure of the WDS drives researchers to seek the least-cost design. This paper first presents the mathematical model to determine the optimal design for the WDS with two conflicting objectives; minimization of construction cost and minimization of total head loss in the network. To deal with large-scale problem in the real-world practice, metaheuristic approach is required to solve the problem. Therefore, this study proposes Differential Evolution (DE) algorithm with encoding and decoding procedures to handling the complexity of decision making in designing pipe sizes of all arcs in the water distribution network. The experiments are executed using the scenarios from the real case-study. Results obtained show that the proposed DE is able to find good a quality front with a set of non-dominated solutions in a single run without prejudice.

Keywords Water distribution · Cost minimization · Total head loss minimization Multi-objective · Differential evolution

W. Wisittipanich (✉)
Department of Industrial Engineering, Faculty of Engineering,
Chiang Mai University, 239 Huay Keaw Road,
Suthep, Muang, Chiang Mai 50200, Thailand
e-mail: warisa.o@gmail.com

D. Buakum
Excellence Center in Logistics and Supply Chain Management,
Chiang Mai University, 239 Huay Keaw Road,
Suthep, Muang, Chiang Mai 50200, Thailand
e-mail: dollaya_b@hotmail.com

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_33

1 Introduction

Water distribution system (WDS) plays a vital role in supplying water for the population living in the city and urban areas. Typically, the optimization problem of pipe network design has been studied in either single or multiple objectives. The aim of a typical single objective study is to test the proposed algorithm. When the researchers discover new algorithms or improve the current algorithms, the least-cost design of water networks would be used to test the efficiency of those optimization techniques. Vasan and Simonovic [1] developed Differential Evolution (DE) linked to hydraulic simulation solver, EPANET, called DENET, to solve two benchmark systems, New York and Hanoi water distribution system. By using DENET, lowest cost was obtained when compared to other seven algorithms. Cisty [2] proposed a combined genetic algorithm (GA) and linear programming (LP) method, named as GALP, for solving WDS design problems. The results showed that GALP provided more robust solutions in terms of closeness to the global minimum. Sedki and Ouazar [3] proposed a hybrid particle swarm optimization (PSO) and DE, linked to the hydraulic simulator, EPNET, for minimizing the cost design of WDS. The performance of the proposed PSO-DE was evaluated using three benchmark problems. It was found the PSO-DE outperformed the standard PSO and other algorithms previously presented in the literature. After that, a new heuristic called the Prescreened Heuristic Sampling Method (PHSM) was proposed by Bi et al. [4] to find optimal pipe diameter in WDS and tested on seven case studies. The result showed that PHSM clearly outperformed other algorithms, both in terms of computational efficiency and the ability to find near optimal solution. Since minimizing the network cost reduces network reliability, the multi-objective (MO) design has been studied for real world problem. Pareto optimization concept has been used to solve MO problem in order to find trade-off solutions between cost and reliability, with the latter being often expressed through compact indexes, such as the network resilience index [5, 6] or head loss [7–9]. Creaco and Franchini [5] and Alvisi and Franchini [6] improved efficiency of optimization using a decomposition approach to find optimal pipe sizing of the new WDS for two objectives; minimizing cost and minimizing head loss. Results obtained showed that the proposed method significantly outperformed other existing methods. Bureerat and Sriworamas [8] designed WDS layout using MO algorithms for simultaneous topology and sizing design of piping networks. The design problem included two objective functions which were network cost and total head loss in pipes. Rahmani and Behzadian [10] presented a methodology based on a three-stage MO optimization model for solving the problem of Battle of Background Leakage Assessment for Water Networks (BBLAWN). The optimal design of pipeline rehabilitation, pump scheduling and tank sizing was formulated and solved on the skeletonized network by optimizing the costs of pipes, pumps and tank upgrading and the cost of water losses and energy.

This paper considers MO problem with Pareto-based approach to design optimal pipe diameter in the water distribution system in a case study of new construction in a small town in Chiang Mai, Thailand.

2 Problem Formulation

In this study, the WDS is formulated as a MO optimization problem to minimize total cost and total head loss of the system. In particular, the proposed model aims to find the optimal pipe diameter of all arcs in the network subjected to the constraints of mass conservation, energy conservation, minimum pressure, flow requirement and pipe size available while the pipe layout and its connectivity, nodal demand, and minimum head requirements are imposed. In addition, this study proposes the new term of flow conservation constraint in which the total quantity of water that flow through any pipe is equal to the demand of destination node plus with all remaining flow along the pipe path. The WDS is mathematically formulated as follows.
Indices

i, j, k node ($i, j, k = 1,2,\ldots, N$)
a_{ij} arc or pipe that connects from start node i to destination node j ($a_{ij} = 1,2,\ldots, npipe$)

Parameters

$L_{a_{ij}}$	Length of arc or pipe a_{ij}
$c\left(D_{a_{ij}}\right)$	Cost per unit length of pipe diameter $D_{a_{ij}}$
$npipe$	Number of pipes in the network
N	Number of nodes in the network
$hl_{a_{ij}}$	Head loss along pipe a_{ij}
H_i	Head of start node i
H_j	Head of destination node j
CW	Hazen-Williams loss coefficient = 150 (Pipe made from PVC)
$FLOW\ a_{ij}$	Total quantity of water flow through pipe a_{ij}
$DEMAND_j$	Demand of destination node j
$H_j^{\min}$	Minimum required pressure at node j
$\{D\}$	Set of commercially available pipe diameter

Decision variable

Da_{ij} Pipe diameter of pipe a_{ij} ($Da_{ij} \in \{D\}$)

The assumptions considered in this research are listed as follows.

- Water is incompressible fluid and flows as laminar flow.
- The pipe system contains a relatively long length of pipe, and thereby the major loss is very large compared to the minor loss. Thus, head loss in this paper considers the major loss only.

- Head of the source node is obtained from gravity only which is equal to the height of tank 15 m.
- Demand of water is equal to 120 L/person/day according to the technical standard of user's demand in municipality by Royal Irrigation Department of Thailand.

As mentioned earlier, this study considers two objective functions which are minimize construction cost and minimize total head loss in the network as shown in Eqs. (1) and (2), respectively.

$$Minimize\, C = \sum_{a_{ij}=1}^{npipe} c\left(Da_{ij}\right) \times L_{a_{ij}} \tag{1}$$

$$Minimize\, HL = \sum_{a_{ij}=1}^{npipe} hl_{a_{ij}}\left(Da_{ij}\right) \times L_{a_{ij}} \tag{2}$$

Constraint (3) explains the head loss along pipe $\left(hl_{a_{ij}}\right)$.

$$\Delta H_{a_{ij}} = H_i - H_j = hl_{a_{ij}} = 10.458 \times \frac{L_{a_{ij}} \times FLOW_{a_{ij}}^{1.852}}{CW^{1.852} \times D_{a_{ij}}^{4.871}}, \quad \forall a_{ij} \tag{3}$$

Constraint (4) is conservation of flow in which the total quantity of water that flow through any pipe is equal to the demand of destination node j plus all remaining flow along the pipe path.

$$FLOW\, a_{ij} = DEMAND_j + \sum_{a_{jk}}^{npipe} FLOWa_{jk}, \quad \forall a_{ij} \tag{4}$$

Constraint (5) is energy conservation that represents the head of destination node j.

$$H_j = H_i - hl_{a_{ij}}, \quad \forall j \tag{5}$$

Constraint (6) determines minimum pressure for each node in the network.

$$H_j \geq H_j^{min}, \quad j = 1, 2, 3, \ldots, N, \quad \forall j \tag{6}$$

Constraint (7) determines pipe size availability which represents the diameter of each pipe that belongs to a commercial size set.

$$Da_{ij} \in \{D\}, \quad \forall a_{ij} \in npipe \tag{7}$$

3 Case Study

This paper studies the water distribution network in a small town in Chiang Mai. As shown in Fig. 1, the network consists of one reservoir, 109 pipes and 101 demand nodes. The design of pipe sizes in the network is based on the available commercial pipe size varying from 2 to 6 inches. The cost per length associated with each pipe size is shown in Table 1.

4 Solution Representation of DE for WDS

The encoding and decoding procedures are proposed to transform a vector into the design of pipe diameter for each arc in WDS. The pipe diameters can only be selected from the predetermined discrete commercial size set D. Therefore, the continuous values in vector dimensions must be converted to the discrete pipe diameters. The minimum pressure head ($Hmin$) is taken as the constraints for each demand node. A penalty approach is adopted to handle the minimum head requirement constraint as stated in Eq. (7). The solution representation used in this study is described as follows.

4.1 Encoding Procedures

Each vector is generated with its dimension equal to the number of arcs in the water distribution network. Initially, each dimension value is randomly filled with uniform random number between 0 and 1. The example of vector initialization in the WDS with 10 arcs is shown in Fig. 2.

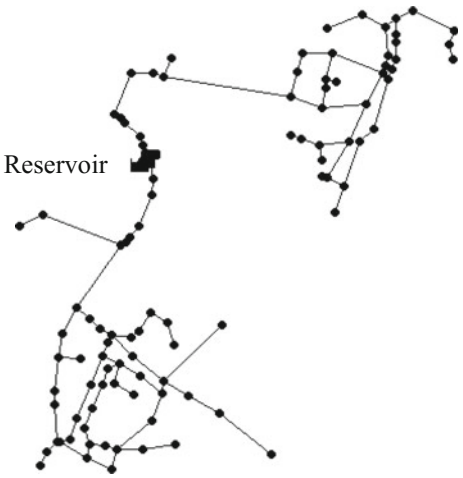


Fig. 1 Connectivity of water distribution network in a small town in Chiang Mai

Table 1 Cost of each pipe diameter

Diameter (inch)	Cost (Bath/ft)
2	14
3	30
4	49
5	57
6	103

Dimension	1	2	3	4	5	6	7	8	9	10
Dimension Value	0.23	0.34	0.19	0.71	0.58	0.46	0.29	0.81	0.65	0.38

Fig. 2 Number of dimensions representing total arcs in a network

4.2 Decoding Procedures

To transform the random number in vector's dimension value into the practical solution, the following five steps are executed.

Step 1 Define the boundary: The number of boundary corresponding to the number of available pipe size in a network is computed using Eq. (11).

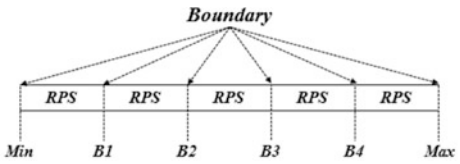
$$Number\ of\ boundary = Number\ of\ available\ pipe\ size + 1 \tag{11}$$

Next, a Range of each Pipe size (RPS) is computed using the maximum and the minimum of dimension values in a current vector and number of discrete pipe size in set D. RPS is calculated using Eq. (12).

$$RPS = \frac{Maximum - Minimum\ dimension\ value\ in\ the\ current\ vector}{Number\ of\ pipe\ size} \tag{12}$$

In this study, the number of available pipe size is 5 (2, 3, 4, 5 and 6 inches), thus the number of boundary is 6 (Min, B1, B2, B3, B4 and Max). The first boundary (Min) is equally to minimum value of dimension values in the current vector. Then, the next boundary is computed using the previous boundary plus RPS, and the last boundary (Max) is equally to maximum value of dimension values of the current vector. The example of 6 boundaries in the WDS with 5 available pipe sizes is shown in Fig. 3.

Step 2 Define the range for each pipe size: The boundaries defined in the first step are used to determine the pipe size for each arc in the network. Figure 4 shows the range for five pipe diameters between boundary values.



Dimension	1	2	3	4	5	6	7	8	9	10
Dimension Value	0.23	0.34	0.19	0.71	0.58	0.46	0.29	0.81	0.65	0.38

Max = 0.81 , *Min* = 0.19
RPS = (*Max* - *Min*)/ *No. Pipe size*
= (0.81- 0.19)/5 = 0.12

Sequence of Boundary	The first boundary	B1	B2	B3	B4	The last boundary
Boundary Computation	Min	B1 = Min+RPS	B2 = B1+RPS	B3 = B2+RPS	B4 = B3+RPS	Max
Boundary Vulue	0.19	0.19+0.12 = 0.31	0.31+0.12 = 0.43	0.43+0.12 = 0.55	0.55+0.12 = 0.67	0.81

Fig. 3 Defining the boundary from dimension values of a current vector

Size Computation	Min < Size1 < B1	B1 < Size2 < B2	B2 < Size3 < B3	B3 < Size4 < B4	B4 < Size5 < Max
Size Vulue	0.19 ≤ Size1 < 0.31	0.31 ≤ Size2 < 0.43	0.43 ≤ Size3 <0.56	0.56 ≤ Size4 < 0.68	0.68 ≤ Size5 ≤ 0.81
Pipe Diameter	2	3	4	5	6
Cost	14	30	49	74	103

Fig. 4 Defining the rank of pipe size from boundary

Step 3 Assign the pipe size and its associated cost for each arc: After the pipe size is defined with range, the pipe diameter and the cost for each arc in a network are assigned correspondingly as shown in Fig. 5.

Step 4 Compute head loss each arc and head each node: Head loss of each arc and head of each destination node are computed after pipe diameter and its cost are assigned using Eqs. (4) and (7) respectively.

Step 5 Compute total cost and total head loss: Total cost and total head loss of the network are computed using Eqs. (1) and (3) respectively. In this step, the minimum pressure head of destination node is taken as the constraints using a penalty approach. Details of computation are given below.

Arc	1	2	3	4	5	6	7	8	9	10
Dimension Value	0.23	0.34	0.19	0.71	0.58	0.46	0.29	0.81	0.65	0.38
Pipe Diameter	2	3	2	6	5	4	2	6	5	3
Cost	14	30	14	103	57	49	14	103	57	30

Fig. 5 Assigning the pipe size and its cost for each arc

1. Begin

   If the minimum pressure head of destination node $\geq$ Hmin, return total cost
   If else the minimum pressure head of destination node < Hmin, return very large value

2. For Total cost computation
3. For Total head loss computation

   If the minimum pressure head of destination node $\geq$ Hmin, return total head loss
   If else the minimum pressure head of destination node < Hmin, return very large value

4. End

5 Experimental Results of the Proposed DE

The proposed DE is implemented in C# programming language under the Microsoft Visual Studio Solution version 2013. The numerical experiments run on the platform of Intel® Pentium CPU NS3540 @2.16 GHz with 4 GB RAM.

The experiments are executed using the scenarios from the real case of water distribution network in a small town in Chiang Mai, Thailand. The DE population size is set as 20 vectors, scale factor is as random value between 0.4 and 0.9, crossover rate is linearly increased from 0.6 to 1.0, and the maximum DE iteration is set as 100 iterations. As shown in Fig. 6, the optimal front for the case study is obtained by DE in a single run with a set of non-dominated solutions.

According to the experimental results, non-dominated solutions obtained from the proposed DE provides the decision maker with good insights into possible alternative for the final decision.

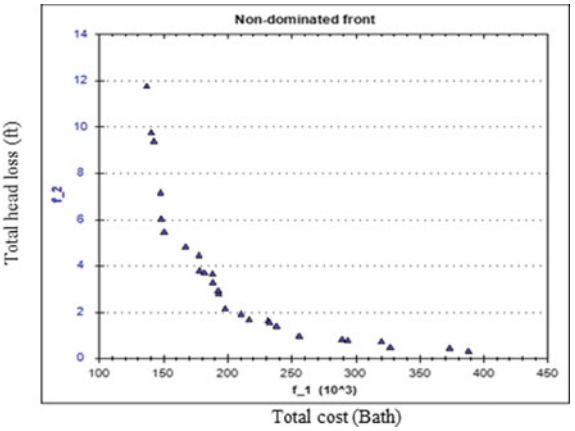


Fig. 6 The Pareto front of a case study obtained by the proposed DE

6 Conclusion

The aim of this work is to find optimal pipe diameter design in the WDS with respect to minimization of construction cost and minimization of total head loss in the system. While the exact method cannot provide solution in a reasonable time, Differential evolution (DE) is implemented to handling the complexity of decision making in real-world practices. To solve the problem, the designing of encoding and decoding process are developed for transform vector solution into the design of pipe size of all arcs in the network. Experimental results obtained shows that the proposed DE is able to find solution for complex problem. Furthermore, good quality optimal front is observed and a set of non-dominated solutions are obtained in a single run without prejudice for consideration of decision makers.

Acknowledgements This work was supported by Excellence Center in Logistics and Supply Chain Management, Chiang Mai University, Thailand.

References

1. Vasan A, Simonovic SP (2010) Optimization of water distribution network design using differential evolution. J Water Resour Plan Manage 136(2):279–287
2. Cisty M (2010) Hybrid genetic algorithm and linear programming method for least-cost design of water distribution systems. Water Resour Manage 24(1):1–24
3. Sedki A, Ouazar D (2012) Hybrid particle swarm optimization and differential evolution for optimal design of water distribution systems. Adv Eng Inform 26(3):582–591
4. Bi W, Dandy GC, Maier HR (2015) Improved genetic algorithm optimization of water distribution system design by incorporating domain knowledge. Environ Model Softw 69:370–381
5. Creaco E, Franchini M (2014) Low level hybrid procedure for the multi-objective design of water distribution networks. Procedia Engineering 70:369–378

6. Alvisi S, Franchini M (2015) A linearization approach for improving the computational efficiency of water distribution system ranking-based optimization algorithms. Procedia Engineering 119:516–525
7. Li M et al (2012) Non-dominated sorting genetic algorithms-iibased on multi-objective optimization model in the water distribution system. Procedia Engineering 37:309–313
8. Bureerat S, Sriworamas K (2013) Simultaneous topology and sizing optimization of a water distribution network using a hybrid multiobjective evolutionary algorithm. Appl Soft Comput 13(8):3693–3702
9. Zheng F, Zecchin A (2014) An efficient decomposition and dual-stage multi-objective optimization method for water distribution systems with multiple supply sources. Environ Model Softw 55:143–155
10. Rahmani F, Behzadian K (2014) Sequential multi-objective evolutionary algorithm for a real-world water distribution system design. Procedia Engineering 89:95–102

Investigation of Characteristics of Noise Storm Solar Burst Type I on 11th March 2013

Z. S. Hamidi and N. N. M. Shariff

Abstract Most type I bursts appear in chains of five or more individual bursts. We studied the event of solar type I solar bursts recorded by the Compact Astronomical Low-cost Low-frequency Instrument for Spectroscopy and Transportable Observatory (CALLISTO-ROSEWELL) spectrometer on 11th March 2013 in the frequency range 320–450 MHz. It was observed 5 s after the flare which happens at 16:24:05UT. The drift rate of this burst is 8 MHz/s with a range of energy between 1.4081 × 106 eV to 1.739 × 106 eV. There are six sunspots were producing flares which are AR1691, AR1690, AR19692, AR1693, AR1694, and AR1689. The sunspot number is 89 and the level of energy is proportional to 680 SFU. Solar Burst Type I is normally in a storm structure and become one of the pre-burst stage. In this case, the burst is associated with a large solar flares. Type I solar burst can be as an indicator of pre-solar flare and CMEs if the intensity of the burst is high. The physical conditions inside and outside the source and the emission mechanisms are discussed.

Keywords Solar burst type I · Noise storm · Solar flare · Coronal mass ejections (CMEs)

Z. S. Hamidi (✉)
Institute of Science, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia
e-mail: zetysh@salam.uitm.edu.my

N. N. M. Shariff
Academy of Islamic and Contemporary Studies, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia

N. N. M. Shariff
Faculty of Applied Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_34

1 Introduction

1.1 Solar Burst Type I

Metric solar radio emissions indicate the existence of non-thermal electrons in the solar corona at altitudes of 0.1–1.0 solar radii (R). Type I storms last from a few hours to several days [1]. Type-I noise storms are one of the solar radio phenomena observed at a meter wavelength with a flux in the range of about 1–100 sfu, and a brightness temperature higher than 10^9 K [2]. The emission mechanism is fundamental plasma emission, due to the coalescence between Langmuir waves and low-frequency waves, and the driver of Langmuir waves is a population of energetic particles trapped in closed magnetic-field structures over an active region [3]. The storms have two components; one is a discrete type-I burst of short duration with a narrow band, and the other is a noise storm continuum of long duration with a broad band. The drift rate can be as a positive or negative frequency drift-rate in the interval of a few tens of MHzs^{-1} to about 200 MHzs^{-1} Elgarøy (1961) and limited range of frequency [4]. This components show highly circular polarization, corresponding to the ordinary mode. It is generally considered that non-thermal electrons that are trapped by closed magnetic field lines excite plasma waves. These excited plasma waves are converted into O-mode waves, which are finally observed as type-I noise storms. It is believed that A storm consists of many individual bursts with short duration, and it is assumed that each burst corresponds to an acceleration of suprathermal electrons [5]. It is also believed that the radiation is attributed to plasma waves excited by suprathermal electrons accelerated between two packets of Alfvén waves traveling in opposite directions [6]. Although the process to produce non-thermal electrons is not well understood, some models have been proposed. What is so unique is that the coronal magnetic fields change very rapidly because the appearance of the new magnetic structures and consequently radio emissions are generated by the interaction between accelerated particles and the turbulent plasma over active regions [7, 8]. The problem that we can studied is the lack of direct measurement of coronal magnetic field in the region make the evolution of coronal phenomena is difficult to be discussed briefly. Furthermore, the formation of this is very complex, long duration of this loop is the magnetic reconnection and disruption of the loops which is observed during the evolution of the active region is not fully understood [9].

Meanwhile, the classification of X-ray solar flare we can conclude with the solar flare explosion on the sun [10]. It will happen when the energy stored in twisted magnetic field is suddenly released. The flares will produce a burst of radiation across the magnetic spectrum, which is from the radio waves to X-ray and gamma rays. With that, the scientist was classified solar flares according to their X-ray brightness in the wavelength range 1–8 Armstrong. There was divide into three categories, which is X-class flares are big. They are major events that can trigger planet-wide radio blackouts and long-lasting radiation storms [11]. Lastly, C-class flares are small with few noticeable consequences here on Earth.

The objectives of this study is to calculate the formula of the drift rate of noise storm and to evaluate the range of energy of the burst. By study about noise storm, we know the activity and the dependence of some of the storm characteristics in meter wavelength region such as mean intensity and number of storms per year [12]. The type I solar burst is the starting point to predict the Sun's phenomena such as solar flare and Coronal Mass Ejections (CMEs).

2 Methodology

The CALLISTO is the short form from the Compact Astronomical Low-cost Low-frequency Instrument for Spectroscopy and Transportable Observatory was introduced by Arnold O.Benz. The E-CALLISTO network was set up and operates identical spectrographs in the different location in the globe for 24 h coverage [13]. CALLISTO spectrometer eC48 have a RS-232 shielded cable and have a 12 V power supply donated by the Institute of Astronomy of ETH Zurich, Switzerland [14]. An individual channel has a bandwidth of 300 kHz and can be tuned by the controlling software in steps of 62.5 kHz [15]. The CALLISTO spectrometer was covering a frequency range from 45 MHz up to the higher is about 870 MHz [16]. We use a tool available from the e-CALLISTO network for background subtraction to study the 255 chains of solar type I solar bursts recorded by the CALLISTO Roswell New Mexico, USA spectrograph on 11th March 2013 in the frequency range 320–450 MHz. We used a software in Java (RAPP Viewer) provides a set of procedures to analyze and visualize spectrograms, including a tool for background subtraction [17]. The calibration process to deduct the Radio Frequency Interference (RFI) is need to obtain a better structure of the burst [18]. Figure 1 shows the CALLISTO system.

In this analysis, the important parameter that has been taken into account are such burst duration, drift rate, energy of photon, and unique structure of the burst. We applied the background-subtract or tool to all recorded spectra. The following parameters were determined for each chain of type I bursts: start frequency (f_i), end frequency (f_f), start time (t_i), end time (t_f), total duration (Δt), total frequency range (Δf), and frequency-drift rate ($\Delta fc/fc f\Delta t$). The information about activities and events such eruptions and solar flares on the Sun were retrieved from spaceweather.com. Spaceweather.com is an official website that provides news and information about the Sun-Earth environment. Meanwhile, Coronal Mass Ejection data and images were retrieved from Large Angle and Spectrometric Coronagraph (LASCO) online database. LASCO is an instrument of 11 instruments included on the joint NASA/ESASOHO (Solar and Heliospheric Observatory) spacecraft. The data of X-ray flux which observed by GOES satellite was retrieved from Space Weather Prediction Centre official website. This website provides alerts and warnings for any disruptions that might affect people and equipment working in space and earth to the nation and the world.

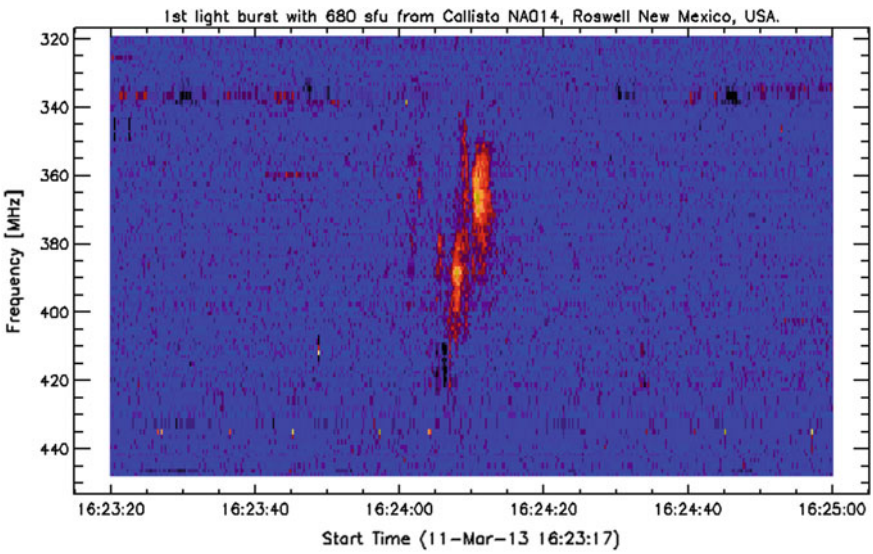


Fig. 1 SRBT I on 11th March 2013. Credit to e-CALLISTO

3 Results and Discussion

It was found that on 11th March 2013, there are also a single SRBT I observed. Whereby, it happens for ten seconds within 16:24:05UT till 16:24:15UT. It also drifted from 420 to 340 MHz, from a higher intensity of burst to low intensity. The event of solar type I solar bursts recorded by the Compact Astronomical Low-cost Low-frequency Instrument for Spectroscopy and Transportable Observatory (CALLISTO-ROSEWELL) spectrometer on 11th March 2013 in the frequency range 320–450 MHz. The drift rate of this burst is 8 MHz/s with a range of energy between 1.4081×10^6 eV and 1.739×10^6 eV. The level of energy is proportional to 680 SFU.

During the day, there are 89 sunspot number were detected by SolarMonitor.org. However, there are six of them were producing flares which are AR1691, AR1690, AR19692, AR1693, AR1694, and AR1689. The solar radio burst type I was produced type C from the active region AR1691. Active region is usually indicated by sunspot. However, not all active regions have sunspot. The plentiful of magnetic field can reach 1000 times stronger in the active region as compared to the average magnetic field of the Sun. Active regions are frequently observed when the Sun's magnetic field is tremendously disturbing during peak of the sunspot cycle. There are also other types of solar dramatic activities that also frequently occurred around active regions such solar prominence and coronal loops. This is because the SRBT I was observed 5 s after the flare which happens at 16:24:05UT. The emission came from a plasma cloud ejected in association with an extensive solar prominence.

Based on the X-Ray flux data above, there are at the type of M5 (5×10^{-5} Wm^{-2}), X2 (2×10^{-4} Wm^{-2}) and X6 (6×10^{-4} Wm^{-2}) classes of solar flare suddenly ejected from the active region AR1081. While this eruption, a magnetic field filament connected to the sunspot AR1689 also erupted. The two wild filaments combined to produce a bright Coronal Mass Ejections (Spaceweather.com). The filament is the largest release of plasma and magnetic field from the solar corona. The long-term evolution of ARs confirmed that the evolution of an active region in each and every detail reflects the evolution of its magnetic field. It was found that flares mainly occur when the magnetic field of the AR has the highest complexity and magnetic flux density, and this is when the variability of AR cores are the highest, too (Fig. 2).

Based on the calculation have made, it was found that the electron density of the burst is 5.406×10^{14} e/m^3 and the drift rate of the solar radio burst are 8 MHz/s. Besides, the photon energy are ranging from 1.408×10^6 eV to 1.739×10^6 eV. The non-thermal electrons are generated through some physical process, are trapped in closed magnetic-field lines, and generate Langmuir waves, which are converted into O-mode radio waves and are later observed as type I bursts Table 1.

Solar Burst Type I is normally in a storm structure and become one of the pre-burst stage. In many cases, if we observed the solar flare phenomena, this burst is associated with C classes of flares. However, in this case, the burst is associated with a large solar flare. The aggregate lifetime of all short-lasting burst is approximately equal to aggregate lifetime of burst of any other duration. The energy of short-lasting burst with duration of 0.2–0.4 s is five times smaller than the energy longer bursts, it is constituted only 2–5% of the energy of the noise storm burst component. Solar burst chains can be some superposition of short-lasting burst on one long burst.

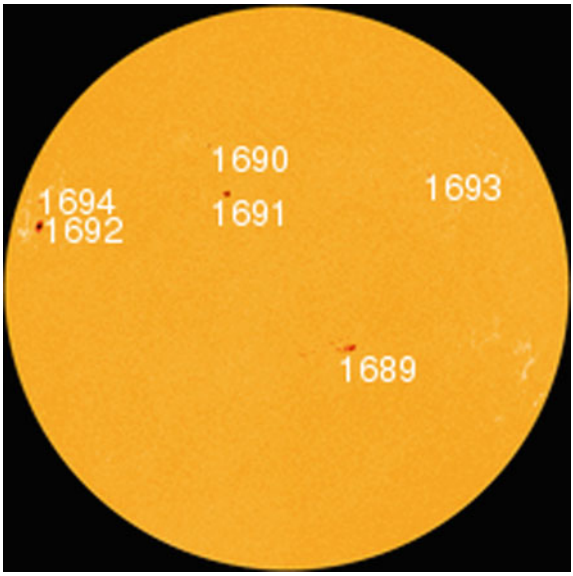


Fig. 2 Active region AR1691. Credited to spaceweather.com

Table 1 The current condition of the Sun on11th March 2013. Credited to Spaceweather

Parameter	Value
Solar wind speed	304.5 km/s
Photon density	5.7 photons/cm^3

4 Conclusion

Type I solar burst can be as an indicator of pre-solar flare and CMEs if the intensity of the burst is high. Even we could not directly confirmed that is only possible because it just based on selected event, we need consider other process to explain the detailed the injection, energy loss and the mechanism of the acceleration of the particles for the further study.

Acknowledgements We are grateful to the CALLISTO network, STEREO, LASCO, SDO/AIA, NOAA and SWPC make their data available online. This work was partially supported by the FRGS grant, 600-IRMI/PERDANA 5/3/MITRA (005/2018)-2 from the Kementerian Pengajian Tinggi Malaysia. Special thanks to the National Space Agency and the National Space Centre for giving us a site to set up this project and support this project. Solar burst monitoring is a project of cooperation between the Institute of Astronomy, ETH Zurich, and FHNW Windisch, Switzerland, Universiti Teknologi MARA and University of Malaya. This paper also used the NOAA Space Weather Prediction Centre (SWPC) for the sunspot, radio flux and solar flare data for comparison purpose. The research has made use of the National Space Centre Facility and a part of an initiative of the International Space Weather Initiative (ISWI) program.

References

1. Krucker S, Lin RP (2000) Two classes of solar proton events derived from onset time analysis. Astrophys J 542:61–64
2. Kattenberg A (1981) Solar radio bursts and their relation to coronal magnetic structures. Rijksuniversiteit te Utrecht
3. Melrose DB (1970) On the theory of Type II and Type III solar radio bursts. Aust J Phys 23:885–903
4. Malville JM (1962) Astrophys J 136:266
5. Suzuki S, Dulk GA (1985) Bursts of type III and type V. In: McLean DJ, Labrum NR (eds) Solar radiophysics: studies of emission from the sun at metre wavelengths. Cambridge Univ. Press, New York
6. Takakura T (1963) Origin of solar radio type I bursts. Publ Astron Soc Jpn 15:462
7. Hamidi Z, Shariff N, Ibrahim Z, Monstein C, Zulkifli WW, Ibrahim M, Amran N (2014) Magnetic reconnection of solar flare detected by solar radio burst Type III. J Phys: Conf Ser 539(1):012006
8. Lantos P, Kerdraon A, Rapley G, Bentley R (1981) Relationship between a soft X-ray long duration event and an intense metric noise storm. Astron Astrophys 101:33–38
9. Shariff N, Hamidi Z, Zainol N (2017) Analysis of slow partial Halo CME events with velocity of 100-200 km/s: an observation through CALLISTO system. Paper presented at the 2017 International Conference on industrial engineering, management science and application (ICIMSA)

10. Hamidi ZS, Anim N, Shariff NNM, Abidin ZZ, Ibrahim ZA, Monstein C (2013) Dynamical structure of solar radio burst type III as evidence of energy of solar flares. Paper presented at the PERFIK 2012, Malaysia
11. Hamidi Z, Zainol N, Ali M, Sabri S, Shariff N, Faid M, … Monstein C (2016) Signal detection of the solar radio burst Type III based on the CALLISTO system project management. Paper presented at the 2016 international conference on industrial engineering, management science and application (ICIMSA)
12. Husien N, Zainol N, Hamidi Z, Sabri S, Shariff N, Faid M, … Monstein C (2016) Solar radio bursts detected by CALLISTO system and their related events. Paper presented at the Industrial Engineering, Management Science and Application (ICIMSA), 2016 International Conference on
13. Zainol N, Sabri S, Hamidi Z, Ali M, Shariff N, Husien N, … Faid M (2016) Effective data collection and analysis of solar radio burst type II event using automated CALLISTO network system. Paper presented at the 2016 International Conference on industrial engineering, management science and application (ICIMSA)
14. Sabri S, Zainol N, Ali M, Shariff N, Hussien N, Faid M, … Monstein C (2016) The dependence of log periodic dipole antenna (LPDA) and e-CALLISTO software to determine the type of solar radio burst (I-V). Paper presented at the 2016 international conference on industrial engineering, management science and application (ICIMSA)
15. Ali M, Sabri S, Hamidi Z, Husien N, Shariff N, Zainol N, …, Monstein C (2016) e-CALLISTO network system and the observation of structure of solar radio burst type III. Paper presented at the 2016 international conference on industrial engineering, management science and application (ICIMSA)
16. Sabri SNU, Hamidi ZS, Shariff NNM, Monstein C (2016) The significance of e-CALLISTO system and construction of log periodic dipole antenna (LPDA) and CALLISTO system for solar radio burst study. Information science and applications (ICISA), 181–188
17. Hamidi Z, Shariff N (2017) Automatic solar radio burst Type II and III image processing tracking by using CALLISTO system. Paper presented at the 2017 international conference on industrial engineering, management science and application (ICIMSA)
18. Abidin ZZ, Anim NM, Hamidi ZS, Monstein C, Ibrahim ZA, Umar R, Sukma I (2015) Radio frequency interference in solar monitoring using CALLISTO. New Astron Rev 67:18–33

A Framework for Measuring Value Chain Performance in the Sugar Industry

Tanapun Srichanthamit and Korrakot Yaibuathet Tippayawong

Abstract This paper aims to design and develop value chain performance measurement for Sugar Industry. The proposed performance measurement framework was designed considering empirical studies and gathering preliminary information from case studies under the Value Chain Management (VCM) concept. Critical success factors were collected from the literature reviews on sugar and other agricultural products and surveys from the experts' perspective. Then, the Quality Function Deployment (QFD) was applied to match the value chain activities and obtained success factors to identify appropriate weights and ranks of those factors. The developed measurement framework was trial in one sugar factory to assess the practical. It was identified that this framework can actually be applied in the sugar case study for internal benchmarking. This paper addresses how to combine VCM and QFD concepts to develop measurement framework for specific industry.

Keywords Value chain management · Performance measurement
Quality function deployment · Critical success factors · Sugar industry

1 Introduction

Food and agribusiness sector plays significant role in driving Thailand's economy. It contributes approximately 10% of the country's GDP [1]. Despite the small proportion, this sector is still of great important since it has been supported by many

T. Srichanthamit
Graduate Program in Industrial Engineering, Faculty of Engineering,
Chiang Mai University, Chiang Mai 50200, Thailand

K. Y. Tippayawong (✉)
Excellence Center in Logistics and Supply Chain Management,
Faculty of Engineering, Chiang Mai University, Chiang Mai 50200, Thailand
e-mail: korrakot@eng.cmu.ac.th

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_35

small stakeholders and about one-third of 70 million population is presently employed in agribusiness sector. This industry generates great value added to domestic agricultural products. However, the food and agribusiness has been facing production problems such as lacking of raw materials, shortage of labor, and the behavior of consumers that always rapidly changed.

Sugar industry is one of the significant parts of the agribusiness sector in Thailand since it is the second most exported product after canned and processed seafood and Thailand is the second largest export sugar to the world (US 2.3$ billion) after Brazil (US 10.4$ billion) [2].

The sugar industry in Thailand has its own characteristics and different from other agricultural products. The sugar quota is distributed to 53 sugar factories in Thailand. The production is strictly controlled by the national office of sugarcane and sugar board to control the volume and price of the sugar produced in Thailand. The stakeholders involved in sugar supply chain include sugarcane farmers, factories and the Office of Sugarcane and Sugar Board.

Value chain analysis is a useful method for identifying how the value can be created along the agri-product supply chain which considers from the upstream farmers toward the downstream consumers domestics and worldwide. Therefore, the investor of agri-business sector should consider adopting this method to analyze and improve their value chain. Value chain analysis could be employed in investigating farmers' activities leading to increase in crop production, reduce non-necessary plantation cost. The production efficiency can be increased through the mid-stream analysis which could constitute further competitive advantage to the industry.

This study aims to develop value chain performance indicators and to evaluate value chain performance measurement of sugar industry. Consequently, this paper addresses how to combine VCM concept and QFD to designing and developing a framework for measuring performance.

This study begins with the value chain analysis of sugar industry in Thailand. The literature review on key success factors of sugar and related agricultural products are then reviewed. The QFD is subsequently introduced to combine obtained success factors and value chain activities. The performance measurement KPIs are finally presented as a result of this research finding.

2 Literature Review

2.1 Value Chain Analysis

Value chain analysis is an analytical method for accessing the sequence of connections in value creation activities that can be occurred from raw materials process to end customer process [3]. Moreover, the added value of each sequence in value

chain activities can increase value to the customer's perspective. In addition, the implementation of the value chain analysis can indicate the problems encountered within the manufacturing process that can be reduced cost or differentiation can be enhanced [4]. Value chain was applied to planning methods and to improve operational efficiency [5]. The VCA method has been developed and adapted to several sectors (iron, cocoa bean, red meat) [6, 7].

Value chain analysis is applied in many industries. In this article, we have applied the value chain analysis to the sugar supply chain and to show the relationship between the activities that are linked. Value chain can be managed to add value to inputs, services and make the final product or service available to end customer.

2.2 Sugar Industry in Thailand

In Thailand, sugarcane is intensively grown in the west and northeast of the country. Recently, there are 53 sugar factories in Thailand. Sugar industry has been classified as a controlled product under office of the cane and sugar board. This control includes harvesting, production quota and delivery [8]. The sugar supply chain in Thailand consists of cane growers, factories and end customers. The upstream starts from preparing soil and sugarcane planting until the sugarcane is harvested before delivering to the plant. The transportation time and approaches of sugarcane to the factory are considered crucial because it relates to the sugar content available when it arrives the factory. Various kind of trucks are used depends on the amount of sugarcane transported and distance. The sugar industry mainly operates four processes includes (1) hauling and milling, (2) raw sugar production process, (3) white sugar/refined sugar production process and (4) packing. The downstream customers are divided into 3 main categories: industrial group, wholesaler and dealer.

2.3 Quality Function Deployment (QFD)

Quality Function Deployment is a technique commonly used to develop products to meet voice of customers. House of Quality is an initial tool and a key component of QFD. It is a tool for converting voice of customer into technical requirements for planning process [9].

However, QFD can be applied in different ways depending on the purpose of implementation and necessity of each case. In this paper, the QFD was applied to find the relationship between the sugar industry's success factors under the activity analysis of the sugar value chain.

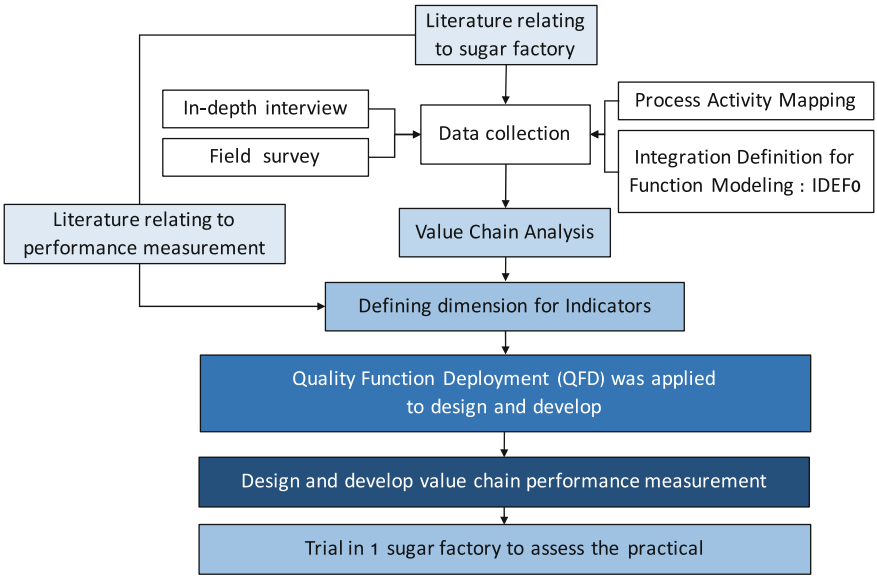


Fig. 1 The methodology to develop performance measurement framework for sugar industry

3 Methodology

The scope for this study is limited to the sugar industry in Thailand. The research process consists of 6 steps. The process of the framework development is illustrated in Fig. 1. The value chain analysis and QFD were applied in this particular issue.

3.1 Value Chain Analysis

The value chain activities of sugar industry were intensively analyzed using process activity mapping and Integration Definition for Function Modeling (IDEF0). The value chain activities encompass 5 primary and 4 supporting activities. Five primary activities are (1) inbound logistics (2) operations (3) outbound logistics (4) marketing and sales and (5) customer services. The literature reviews on critical success factors (CSFs) in each value chain activity were conducted. Those obtained factors were categorized to each activity together with five performance dimensions of sugar industry. The result from this stage indicated primary critical success factors related to each value chain activity and performance dimension. This primary result was the input of later stages, expert opinion and QFD forming.

3.2 Defining Dimensions for Performance Measurement

Defining dimensions for performance measurement depends on the suitability of each activity under the value chain concept by considering the research, literature review and the opinions of experts. Not only dimensions for performance measurement but also indicators in each dimension will be considering the research, literature review and the expert's opinions.

3.3 Application of QFD in Value Chain Concept

In this approach, 7 experts from sugar industry, academic and government agent are requested to put weight and ranks including the relationship level between each critical success factor and value chain activity together with performance dimension of sugar industry through the QFD methods.

QFD can be used to identify a set of indicators that are related to value chain activities. The steps involved in this methodology are listed below.

1. Identify critical success factors.
2. Identify indicators.
3. Build QFD chart.

Identify critical success factors. The critical success factors can be obtained in several ways including literature reviews, factory surveys, the expertise's perspective. This data will then be used in the QFD chart.

Identify indicators. The indicators used in the QFD chart are identified. These indicators reflected what are viewed as being importance(effect) in the value chain activities. In the QFD chart, these indicators will be assessed to observe if they meet any of the activities which have been identified.

Build QFD chart. This step involves building a QFD chart using the value chain activities and the indicators.

Table 1 The signs of relationship used in the QFD analysis

Scale		Legend
9	⦿	Strong relationship
3	◎	Moderate relationship
1	▲	Weak relationship
	–	Negative correction
	+	Positive correction
	▲	Objective is to maximize
	▼	Objective is to minimize

The importance of this step is that a QFD chart shows the relationship between value chain activities and the indicators under five performance dimensions. The relationship between the indicators will find out with the signs given in Table 1. This step represents the indicator of each value chain activities and demonstrates which indicators the sugar factories should focus on.

4 Results and Discussion

The results are divided into 2 parts, which are: (1) A Value Chain Analysis and (2) Application of QFD in Value Chain concept.

4.1 Value Chain Analysis

The literature reviews based on the success factors of each value chain activities were conducted. After achieving the primary success factors, experts from Thai sugar industry were requested to provide opinions and importance regarding those obtained factors. The initial result on sugar value chain success factors is revealed in terms of nine value chain activities (Table 2) and five value chain dimensions (Fig. 2).

Table 2 Critical success factors for sugar industry in terms of five main value chain activities

Value chain activity	Critical success factor	References
1. Inbound logistics	Raw material handling system Raw material stability	[10–14] [15–17]
2. Operations	Increasing efficiency in production process and packing Standard production system Maintenance system Laboratory	[18–21] [15, 18–20] [17, 21] [17, 18, 21]
3. Outbound logistics	Delivery system Location	[13, 18] [14, 22]
4. Marketing and sales	Marketing capabilities Brand image	[14, 17, 21, 23–25] [26–28]
5. Customer services	Consumer confidence Flexibility in response customer	[21, 29, 30] [18, 30, 31]

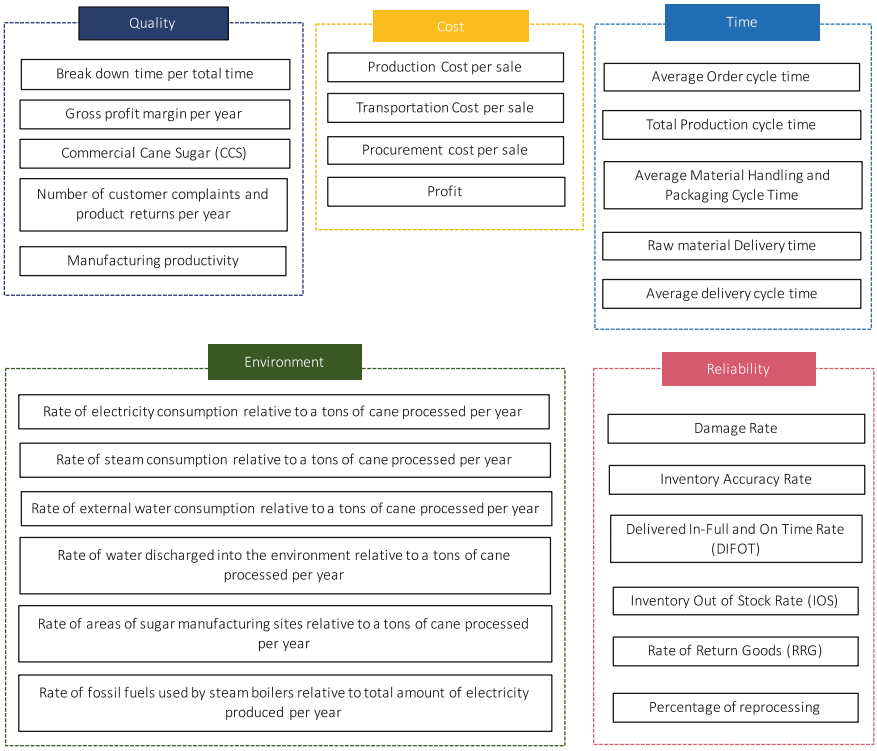
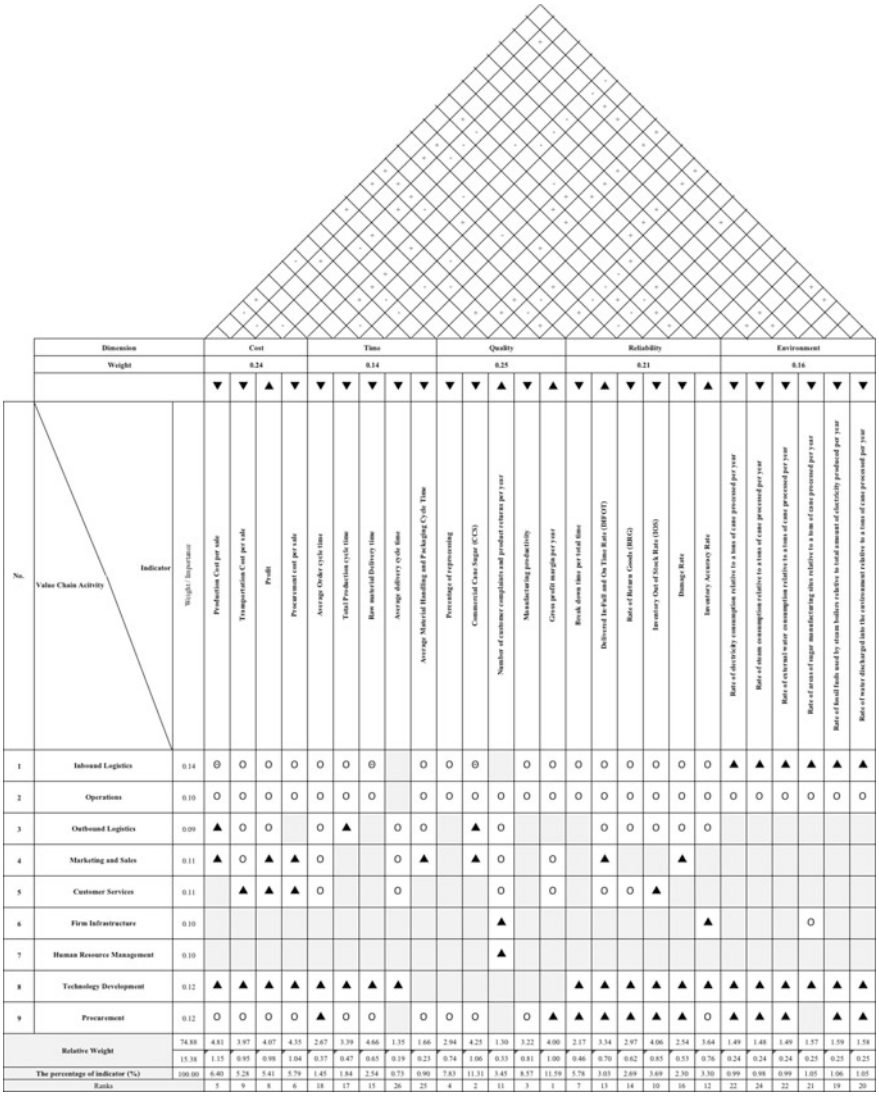


Fig. 2 Critical success factors for sugar industry in terms of five value dimensions

4.2 Application of QFD in Value Chain Concept

From the previous step, the critical success factors and the initial dimensions for the value chain activities and indicators were achieved. This step represents the indicator of each value chain activity and the result indicates that what factors the sugar industry should focus on.

The QFD chart for the sugar industry is shown in Fig. 3. This is a guideline for designing and setting indicators for performance measurement of value chains for sugar factory. Conclusions on Performance Indicators for Value Chain see Table 3.



No.	Value Chain Activity / Indicator	Weight / Importance	Production Cost per sale	Transportation Cost per sale	Profit	Procurement cost per sale	Average Order cycle time	Total Production cycle time	Raw material Delivery time	Average delivery cycle time	Average Material Handling and Packaging Cycle Time	Percentage of reprocessing	Commercial Cane Sugar (CCS)	Number of customer complaints and product returns per year	Manufacturing productivity	Gross profit margin per year	Break down time per total time	Delivered In-Full and On Time Rate (DIFOT)	Rate of Return Goods (ERG)	Inventory Out of Stock Rate (IOS)	Damage Rate	Inventory Accuracy Rate	Rate of electricity consumption relative to a tons of cane processed per year	Rate of steam consumption relative to a tons of cane processed per year	Rate of external water consumption relative to a tons of cane processed per year	Rate of areas of sugar manufacturing sites relative to a tons of cane processed per year	Rate of fossil fuels used by steam boilers relative to total amount of electricity produced per year	Rate of water discharged into the environment relative to a tons of cane processed per year
	Dimension		Cost				Time					Quality					Reliability						Environment					
	Weight		0.24				0.14					0.25					0.21						0.16					
			▼	▼	▲	▼	▼	▼	▼	▼	▼	▼	▼	▲	▼	▲	▼	▲	▼	▼	▼	▲	▼	▼	▼	▼	▼
1	Inbound Logistics	0.14	⊖	O	O	O	O	O	⊖		O	O	⊖		O	O	O	O	O	O	O	O	▲	▲	▲	▲	▲	▲
2	Operations	0.10	O	O	O	O	O	O	O		O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
3	Outbound Logistics	0.09	▲	O	O		O	▲		O	O		▲	O				O	O	O	O	O						
4	Marketing and Sales	0.11	▲	O	▲	▲	O			O	▲		▲	O		O		▲			▲							
5	Customer Services	0.11		▲	▲	▲	O			O				O		O		O	O	▲								
6	Firm Infrastructure	0.10												▲								▲				O		
7	Human Resource Management	0.10												▲														
8	Technology Development	0.12	▲	▲	▲	▲	▲	▲	▲	▲							▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
9	Procurement	0.12	O	O	O	O	▲	O	O		O	O	O		O	▲	▲	▲	▲	▲	▲	O	▲	▲	▲		▲	▲
	Relative Weight	74.88	4.81	3.97	4.07	4.35	2.67	3.39	4.66	1.35	1.66	2.94	4.25	1.30	3.22	4.00	2.17	3.34	2.97	4.06	2.54	3.64	1.49	1.48	1.49	1.57	1.59	1.58
		15.38	1.15	0.95	0.98	1.04	0.37	0.47	0.65	0.19	0.23	0.74	1.06	0.33	0.81	1.00	0.46	0.70	0.62	0.85	0.53	0.76	0.24	0.24	0.24	0.25	0.25	0.25
	The percentage of indicator (%)	100.00	6.40	5.28	5.41	5.79	1.45	1.84	2.54	0.73	0.90	7.83	11.31	3.45	8.57	11.59	5.78	3.03	2.69	3.69	2.30	3.30	0.99	0.98	0.99	1.05	1.06	1.05
	Ranks		5	9	8	6	18	17	15	26	25	4	2	11	3	1	7	13	14	10	16	12	22	24	22	21	19	20

Fig. 3 The QFD chart for the sugar industry indicate 9 value chain activities (left) and indicators of five performance attribute (top)

Table 3 Proposed value chain performance KPIs as a result from QFD analysis

Value chain activities	Quality	Cost	Reliability	Environment	Time
Inbound logistics	Commercial cane sugar	Raw material cost per sale	–	–	Raw material delivery time
Operations	Manufacturing productivity	Production cost per sale	–	Greenhouse gas emissions	Total production cycle time
Outbound logistics	–	Transportation cost per sale	Delivered in-full and on time rate	–	Average delivery cycle time
Marketing and sales	–	–	–	–	–
Customer services	Number of customer complaints and product returns per year	–	Delivered in-full and on time rate	–	–

5 Conclusion

In this paper, a framework for value chain performance measurement in the sugar factory was proposed. A framework composed of 12 KPIs in four value chain primary activities (inbound logistics, operations, outbound logistics, and customer service) together with five performance attributes (time, cost, quality reliability, environment) of sugar industry. The verification of 12 KPIs was undertaken in one sugar factory in Thailand and the practical is proved.

The proposed performance measurement model was designed by considering empirical studies and gathering of preliminary success factors from the case studies from either sugar or other agricultural products in terms of their value chain management. The QFD technique was then employed to integrate those obtained success factors into value chain activities and performance attributes of sugar industry. However, the initial performance measurement framework was trial in only one company, more case studies should be included for the better verification result.

Although the performance measurement model from this research was modified specifically for sugar industry, other industries could duplicate similar approach to develop specific performance measurement for specific industry. This can be one of the contributions from this research.

Acknowledgements The authors gratefully acknowledge the Excellent Center in Logistics and Supply Chain Management (E-LSCM), Chiang Mai University for the technical and financial supports.

References

1. Singhapreecha C (2014) Economy and agriculture in Thailand. FFTC Agriculture Policy Articles. http://ap.fftc.agnet.org/ap_db.php?id=246. Last accessed 2018/2/15
2. Workman D (2017) Sugar exports by Country. http://www.worldstopexports.com/sugar-exports-country/. Last accessed 2018/2/20
3. Porter ME (1985) Competitive advantage: creating and sustaining superior performance. The Free Press, Macmillan Publishing, New York
4. Shank JK, Govindarajan V (1993) What "drives" cost? A strategic cost management perspective. Adv Manage Acc 2:27–46
5. Kess P, Law K, Kanchana R, Phusavat K (2010) Critical factors for an effective business value chain. Ind Manage Data Syst 110(1):63–77
6. Dahlström K, Ekins P (2006) Combining economic and environmental dimensions: value chain analysis of UK iron and steel flows. Ecol Econ 58(3):507–519
7. Panlibuton H, Lusby F (2006) Indonesia cocoa bean value chain case study. USAID Micro report 65
8. Chiadamrong N, Kawtummachai R (2008) A methodology to support decision-making on sugar distribution for export channel: a case study of Thai sugar industry. Comput Electron Agric 64(2):248–261
9. Akao Y (1990) History of quality function deployment in Japan. Best Qual Targets Improv Syst 3:183–196
10. Özer Ö (2003) Replenishment strategies for distribution systems under advance demand information. Manage Sci 49(3):255–272
11. Childerhouse P, Towill DR (2004) Reducing uncertainty in European supply chains. J Manuf Technol Manage 15(7):585–598
12. Disney SM, Naim MM, Potter A (2004) Assessing the impact of e-business on supply chain dynamics. Int J Prod Econ 89(2):109–118
13. Ramdas K, Spekman RE (2000) Chain or shackles: understanding what drives supply-chain performance. Interfaces 30(4):3–21
14. Singh RK, Kumar R, Shankar R (2012) Supply chain management in SMEs: a case study. Int J Manuf Res 7(2):165–180
15. Kumar R, Singh RK, Shankar R (2013) Study on coordination issues for flexibility in supply chain of SMEs: a case study. Glob J Flex Syst Manage 14(2):81–92
16. Xu L, Mathiyazhagan K, Govindan K, Haq AN, Ramachandran NV, Ashokkumar A (2013) Multiple comparative studies of green supply chain management: pressures analysis. Resour Conserv Recycl 78:26–35
17. Kumar R, Singh RK, Shankar R (2016) Study on collaboration and information sharing practices for SCM in Indian SMEs. Int J Bus Inf Syst 22(4):455–475
18. Higgins A, Thorburn P, Archer A, Jakku E (2007) Opportunities for value chain research in sugar industries. Agric Syst 94(3):611–621
19. Bailey K, Francis M (2008) Managing information flows for improved value chain performance. Int J Prod Econ 111(1):2–12
20. Green K, Morton B, New S (1996) Purchasing and environmental management: interactions, policies and opportunities. Bus Strategy Environ 5(3):188–197
21. Rieple A, Singh R (2010) A value chain analysis of the organic cotton industry: the case of UK retailers and Indian suppliers. Ecol Econ 69(11):2292–2302
22. Fang D, Weng W (2010) KPI evaluation system of location decision for plant relocation from the view of the entire supply chain optimization. In: IEEE international conference on Automation and Logistics (ICAL), pp 659–663
23. Chen YS, Lai SB, Wen CT (2006) The influence of green innovation performance on corporate advantage in Taiwan. J Bus Ethics 67(4):331–339
24. Bailey K, Francis M (2008) Managing information flows for improved value chain performance. Int J Prod Econ 111(1):2–12

25. Luthra S, Kumar V, Kumar S, Haleem A (2011) Barriers to implement green supply chain management in automobile industry using interpretive structural modeling technique: an Indian perspective. J Ind Eng Manag 4(2):231–257
26. Khiewnavawongsa S, Schmidt EK (2008) Green power to the supply chain. Adv Mark 244
27. Nishat Faisal M (2010) Sustainable supply chains: a study of interaction among the enablers. Bus Process Manage J 16(3):508–529
28. Lin RJ (2011) Moderating effects of total quality environmental management on environmental performance. Afr J Bus Manage 5(20):8088
29. Hubbard G (2009) Measuring organizational performance: beyond the triple bottom line. Bus Strategy Environ 18(3):177–191
30. Luthra S, Garg D, Haleem A (2015) Critical success factors of green supply chain management for achieving sustainability in Indian automobile industry. Prod Plann Control 26(5):339–362
31. Perdan S, Azapagic A, Clift R (2000) Teaching sustainable development to engineering students. Int J Sustain High Educ 1(3):267–279

Design and Optimization of a Multi-echelon Supply Chain Network for Product Distribution with Cross-Route Costs and Traffic Factor Values

Asnaf Aziz, Razaullah and Iftikhar Hussain

Abstract Network design and optimization problems for product flow appear widely in shipping and production applications. We present a new variation to such class of problems in which the shipping cost linked with a route depends not only on the product flow moving across that route but on the product flow on other routes in the supply chain network as well. Selecting a route with fewer hurdles increases the product flow effectiveness. We consider the entire supply chain network i.e., from raw material supply to production, finished products to warehouses and then to the demand points. We formulate an integer mathematical model and present computational results for a set of test problems arising from shipping and production applications to analyze how the model performs with the varying network characteristics.

Keywords Network design · Product distribution · Cross-route costs
Traffic factor

1 Introduction

Network design and product distribution problems determine the network structure i.e., what supply routes and nodes to be included in the network and what amount of products to supply on them. Basically, shipping cost minimization problems involve reducing the shipping charges of the finished goods from one site to another, such that the requirements of each site are met and each supply node

A. Aziz (✉) · Razaullah
Department of Mechanical Engineering Technology,
University of Technology, Nowshera, Pakistan
e-mail: Asnaf.aziz@uotnowshera.edu.pk

I. Hussain
Department of Industrial Engineering, University
of Engineering and Technology, Peshawar, Pakistan

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_36

operates within its capacity. The network design of supply chain has a great impact on its performance because it determines its configuration and sets constraints to minimize the supply chain cost and maximize responsiveness. A good decision for the location of network nodes can help a supply chain be responsive while minimizing its costs. Building many facilities reduces shipping costs and provides faster responsiveness but increases the inventory and facility costs incurred by the firm. In the entire supply chain, products are shipped from node to node by different shipping firms, according to some discount for a range of products. A supply route with more hurdles decreases the responsiveness and results in tedious delivery of the products. Thus, the main objective is to minimize the supply chain cost and to satisfy the customer demand within the available resources.

There is a vast body of literature related to supply chain network design and product distribution problems. General supply chain network for product flow surveys can be found in Longinidis et al. [1] Wang et al. [2] and Nagurney et al. [3]. The warehouse-customer interface model presented by Gill [4] minimizes the number of warehouse location while ensuring that each customer is assigned to at least one selected warehouse and such an assignment satisfies the maximum allowable distance. The objective of the model presented by Cohn et al. [5] is to assign freights to arcs so as to minimize cost, trading off the use of the lowest-cost arc with the use of highest-cost arc to increase the number of products assigned to a given air carrier and thus decreases total cost for product flow assigned to that air carrier. Eliiyi et al. [6] obtained an optimal route bus-garage allocations that minimize the total distance covered in all pull-out and pull-in trips, and reached significant improvement levels with respect to the current situation and minimized the fuel consumption in city bus transportation. The main purposes of the research presented by Zho et al. [7] are to establish a constrained maximum-capacity linear-integer programming model and to develop a routing method that allows a particular vehicle using the maximum-capacity route within a given consumption of the fuel. Leblanc and Boyce [8] consider problems in which one decision maker designs the network and a second set of decision makers assign the flow. The work presented by Machado et al. [9] proposes a network flow linear program model to solve the problem of minimizing costs of product, distribution of multi-commodities and the class of multi-commodity network flow problems. Bozorgirad et al. [10] developed an algorithm to solve the multi-source multi-product flexible multistage logistic network problem using Genetic Algorithm technique to solve this kind of problem to obtain an acceptable solution within a reasonable time. The work presented by Kalaitzidou et al. [11] introduced a general mathematical programming framework that employs an innovative composition coupled with forward and reverse logistics activities. Barnhart et al. [12] presented a column-generation model for the class of integer multi-commodity flow problems and a branch-and-bound solution approach involving column and row generation. Nielsen et al. [13] introduced optimization models to quickly and effectively design the channel route network and to provide a method by which channel route schedulers can generate and evaluate several, rather than a single, candidate channel route structures. Pant et al. [14] presented a framework for

transparency, traceability and information flow for management of dairy supply chain networks. Yolmeh et al. [15] proposed a model for the integrated problem of supply chain network designing and assembly line balancing under demand uncertainty. The problem presented involves determining the location of manufacturers and assemblers in the network, balancing the assembly lines, and shipping of materials and products throughout the supply chain network.

We contribute to this literature by presenting a variation of the supply chain network design and product flow problems that, to the best of our knowledge, has not previously appeared in the literature, that in which the cost associated with product flow across a supply route depends not only on that route's flow, but on the product flow across other routes in the network as well. We consider the cost per unit per mile flow of products and the traffic factor issue is also incorporated in this research as a supply route with less or no hurdles increases the responsiveness and results in better services. The flow of products through different routes in the supply chain network by a shipping firm with relatively greater discount decreases the total cost. The proposed procedure is based on an algorithm which determines the product flow across the routes with no or less hurdles of the multi-echelon supply chain network and the demand of each node is satisfied. A numerical example clarifies the procedure and the computational results further provide insights.

The remainder of the article is organized as follows. In Sect. 2, the problem is formally stated and an integer programming formulation is presented to solve it. In Sect. 3, a case study is tested on a network designed for product flow across the entire supply chain. Finally, computational extensions are provided to assess how the proposed model performs with the variation in the network characteristics, and areas for further study are suggested.

2 The Design and Optimization of Multi-echelon SCN Problem and Its Formulation

Current problem, as defined above, focuses on the entire supply chain network design and its optimization. The nodes represent the locations of raw materials, production plants, warehouses and customers. We assumed that one unit of raw material produces one unit of finished product. The supply routes make the arcs of the network. Multiple shipping firms perform the delivery function between a pair of nodes. The shipping cost associated with a route is simply a linear function of the product flow over that route. The shipping firms are often willing to negotiate discounts based on the volume of products. Thus, the cost of moving over one route depends on the quantity of product flow not only on that route but on the product flow on all other routes in the network associated with that shipping firm as well. The selection of a supply route with less or no hurdles results in a convenient product delivery. The cost structure for such multi-echelon supply chain network is therefore based on four components—the cost per unit per mile, the distance

between a pair of nodes, the corresponding traffic factor value and the discount factor that depends on the amount of products moving overall supply routes associated with a specific shipping firm.

In this problem, we considered a set of products, a set of network nodes and a set of supply routes. The number of products is partitioned into different flow ranges and an associated discount factor with each of these ranges. For example, if the total product flow across all supply routes is in the range of 250 to 499, then there may be 5% discount. The problem can be viewed as how much product flow to be assigned to a supply route in each stage of the multi-echelon supply chain network so as to best utilize the capacity of each node and satisfy the customer demand with the objective of minimizing the total cost.

Before presenting the problem formulation, we introduce the following notation.

Sets

a number of shipping firms
b set of discount ranges
l number of supply nodes
m number of demand locations
n number of potential production plant locations
t number of potential warehouse locations

Data

D_j annual demand from customer at location j
K_i potential capacity of plant at location i
S_h supply capacity of supplier at location h
W_e potential capacity of warehouse at location e
F_i fixed cost of locating a plant at location i
f_e fixed cost of locating a warehouse at location e
C_{hi} cost per mile per unit from supply source h to plant i
C_{ie} cost per mile per unit from plant i to warehouse e
C_{ej} cost per mile per unit from warehouse e to customer at location j
T_{hi} traffic factor value of the route from supply source h to plant at location i
T_{ie} traffic factor value of the route from plant i to warehouse at location e
T_{ej} traffic factor value of the route from warehouse e to customer at location j
l_r^c lower bound on flow for discount range r by shipping firm c
u_r^c upper bound on flow for discount range r by shipping firm c
p_{cr} percent discount offered by shipping firm c for products in range r

Variables

z_r^c 1 if total flow of products shipped by firm c is in the range r, else 0
Y_i 1 if plant is located at location i, else 0
y_e 1 if warehouse is located at site e, else 0
x_{hi}^{cr} number of units shipped from supplier h to plant i priced according to discount range r by shipping firm c

x_{ie}^{cr} number of units shipped from plant i to warehouse e priced according to discount range r by shipping firm c

x_{ej}^{cr} number of units shipped from warehouse e to demand point j priced according to discount range r by shipping firm c

The problem is formulated as the following integer program:
Minimize:

$$\sum_{i=1}^{n} F_i Y_i + \sum_{e=1}^{t} f_e y_e + \sum_{c=1}^{a}\sum_{r=1}^{b}\sum_{h=1}^{l}\sum_{i=1}^{n} (1-p_{cr})\frac{d_{hi}}{T_{hi}} C_{hi} x_{hi}^{cr} + \sum_{c=1}^{a}\sum_{r=1}^{b}\sum_{i=1}^{n}\sum_{e=1}^{t} (1-p_{cr})\frac{d_{ie}}{T_{ie}} C_{ie} x_{ie}^{cr} + \sum_{c=1}^{a}\sum_{r=1}^{b}\sum_{e=1}^{t}\sum_{j=1}^{m} (1-p_{cr})\frac{d_{ej}}{T_{ej}} C_{ej} x_{ej}^{cr}$$

The objective function being the total cost of the entire supply chain network is required to be minimized subject to the following constraints:

$$\sum_{c=1}^{a}\sum_{r=1}^{b}\sum_{i=1}^{n} x_{hi}^{cr} \le S_h \quad \forall h \in l \tag{1}$$

$$\sum_{c=1}^{a}\sum_{r=1}^{b}\sum_{h=1}^{l} x_{hi}^{cr} - \sum_{c=1}^{a}\sum_{r=1}^{b}\sum_{e=1}^{t} x_{ie}^{cr} \ge 0 \quad \forall i \in n \tag{2}$$

$$\sum_{c=1}^{a}\sum_{r=1}^{b}\sum_{e=1}^{t} x_{ie}^{cr} \le K_i y_i \quad \forall i \in n \tag{3}$$

$$\sum_{r=1}^{b}\sum_{i=1}^{n} x_{ie}^{cr} - \sum_{r=1}^{b}\sum_{j=1}^{m} x_{ej}^{cr} \ge 0 \quad \forall c \in a,\ e \in t \tag{4}$$

$$\sum_{r=1}^{b}\sum_{j=1}^{m} x_{ej}^{cr} \le W_e y_e \quad \forall c \in a,\ e \in t \tag{5}$$

$$\sum_{r=1}^{b}\sum_{e=1}^{t} x_{ej}^{cr} \ge D_j \quad \forall c \in a,\ j \in m \tag{6}$$

$$\sum_{r=1}^{b}\sum_{i=1}^{n} x_{hi}^{cr} \ge l_r^c z_r^c \quad \forall c \in a,\ h \in l \tag{7}$$

$$\sum_{r=1}^{b}\sum_{e=1}^{t} x_{ie}^{cr} \ge l_r^c z_r^c \quad \forall c \in a,\ i \in n \tag{8}$$

$$\sum_{r=1}^{b}\sum_{j=1}^{m} x_{ej}^{cr} \geq l_r^c z_r^c \quad \forall c \in a,\ e \in t \tag{9}$$

$$\sum_{r=1}^{b}\sum_{i=1}^{n} x_{hi}^{cr} \leq u_r^c z_r^c \quad \forall c \in a,\ h \in l \tag{10}$$

$$\sum_{r=1}^{b}\sum_{e=1}^{t} x_{ie}^{cr} \leq u_r^c z_r^c \quad \forall c \in a,\ i \in n \tag{11}$$

$$\sum_{r=1}^{b}\sum_{j=1}^{m} x_{ej}^{cr} \leq u_r^c z_r^c \quad \forall c \in a,\ e \in t \tag{12}$$

$$z_r^c \in \{0,1\} \quad \forall c \in a,\ r \in b \tag{13}$$

$$Y_i, y_e \in \{0,1\} \quad \forall i \in n,\ e \in t \tag{14}$$

$$x_{hi}^{cr}, x_{ie}^{cr}, x_{ej}^{cr} \geq 0, \quad \textit{integers} \quad \forall c \in a, r \in b, h \in l, i \in n, j \in m, e \in t \tag{15}$$

The shipping firm $c \in a$ delivers the raw materials and finished products from one location to another location according to the discount range $r \in b$.

The objective of the problem sums the total product flow over each arc of the network, priced according to the cost per unit per mile times the corresponding distance divided by the traffic factor value times one minus the discount offered for a given range of products by a shipping firm.

The 1st constraint specifies that the total amount shipped from a supplier cannot exceed the supplier's capacity while the 2nd constraint states that the amount shipped out of a plant cannot exceed the quantity of raw material received. The 3rd constraint enforces that the amount produced in a production plant cannot exceed its capacity and the 4th constraint specifies that the amount shipped out of a warehouse cannot exceed the quantity received from the production plants. The 5th constraint states that the amount shipped through a warehouse cannot exceed its capacity whereas the 6th constraint states that the amount shipped to a customer must cover the demand. The constraints from 7–12 enforce that the total flow across all routes by a shipping firm in a given range is either zero or satisfies the upper and lower bounds of the range while the 13th constraint ensures that z_r^c is equal to 1 if total flow of products shipped by firm a is in the range b, else 0. The 14th constraint enforces that each production plant or warehouse is either open or closed and the 15th constraint ensures that the number of units shipped in the supply chain network is both greater than or equal zero and integers.

3 Case Study

In this case study, two shipping firms perform the product delivery function on different routes of the entire supply chain network. Table 1 summarizes the discounts offered by the two shipping firms for different ranges of products. It shows that there is no discount from shipping firm A, if this firm is assigned less than 250 units of product flow across all routes in total. If the total flow assigned to shipping firm A is 250 or more but less than 500, there is 5% discount on product flow across all routes in total. Similarly, if the total flow assigned to shipping firm A is between 500 and 1000, then all of this product flow will receive a 10% discount. The discounts offered for the same ranges of product flow by shipping firm B, are 2, 7 and 10%, respectively. In this case study, the three different product flow ranges are the same for both the firms, this is only for simplicity and these ranges may differ from one another for the shipping firms.

The real distances of the entire supply chain network are shown in Tables 2, 3 and 4. The selection of the shortest route for product flow results in minimizing shipping costs and response time. Customer preferences may be in terms of desired response time and the choice of shipping firm. If the response time is longer than that acceptable to the customers, we simply drop the product flow route designed for a shipping firm from our model.

Table 1 Discount factors

Range	Units	Shipping firm A (%)	Shipping firm B (%)
1	0–249	0	2
2	250–499	5	7
3	500–1000	10	10

Table 2 Distance from supply node to production plant

d_{hi} (miles)		
$d_{14} = 45$	$d_{24} = 85$	$d_{34} = 90$
$d_{15} = 90$	$d_{25} = 50$	$d_{35} = 95$
$d_{16} = 90$	$d_{26} = 80$	$d_{36} = 90$
$d_{17} = 85$	$d_{27} = 60$	$d_{37} = 70$

Table 3 Distance from production plant to warehouse

d_{ie} (miles)			
$d_{48} = 70$	$d_{58} = 85$	$d_{68} = 40$	$d_{78} = 75$
$d_{49} = 80$	$d_{59} = 60$	$d_{69} = 95$	$d_{79} = 60$
$d_{4,10} = 40$	$d_{5,10} = 95$	$d_{6,10} = 85$	$d_{7,10} = 90$

The traffic factor value is 1 (maximum) when there are no hurdles on a supply route. Tables 5, 6 and 7 indicate the corresponding traffic factor values on each supply route of the network between a pair of nodes. There may be construction, check posts, curves and uneven road surfaces on a supply route which increases the response time and hence shipping costs. The selection of supply route with less or no hurdles (maximum traffic factor value) increases the product delivery effectiveness and results in convenient shipping.

The cost of shipping one unit of commodities for one mile between a pair of two nodes is considered in this model. The costs per unit per mile for each of the supply

Table 4 Distance from warehouse to demand point

d_{ej} (miles)		
$d_{8,11} = 50$	$d_{9,11} = 90$	$d_{10,11} = 75$
$d_{8,12} = 95$	$d_{9,12} = 95$	$d_{10,12} = 85$
$d_{8,13} = 80$	$d_{9,13} = 85$	$d_{10,13} = 85$
$d_{8,14} = 70$	$d_{9,14} = 90$	$d_{10,14} = 90$

Table 5 Traffic factor from supply node to production plant

T_{hi} Traffic factor		
$T_{14} = 1.00$	$T_{24} = 0.85$	$T_{34} = 0.90$
$T_{15} = 0.90$	$T_{25} = 1.00$	$T_{35} = 0.95$
$T_{16} = 0.90$	$T_{26} = 1.00$	$T_{36} = 0.90$
$T_{17} = 0.85$	$T_{27} = 0.95$	$T_{37} = 1.00$

Table 6 Traffic factor from production plant to warehouse

T_{ie} Traffic Factor			
$T_{48} = 1.00$	$T_{58} = 1.00$	$T_{68} = 1.00$	$T_{78} = 1.00$
$T_{49} = 0.90$	$T_{59} = 0.90$	$T_{69} = 0.95$	$T_{79} = 0.95$
$T_{4,10} = 0.85$	$T_{5,10} = 0.95$	$T_{6,10} = 0.95$	$T_{7,10} = 0.90$

Table 7 Traffic factor from warehouse to demand point

T_{ej} Traffic Factor		
$T_{8,11} = 1.00$	$T_{9,11} = 1.00$	$T_{10,11} = 1.00$
$T_{8,12} = 0.95$	$T_{9,12} = 0.95$	$T_{10,12} = 0.85$
$T_{8,13} = 0.95$	$T_{9,13} = 0.95$	$T_{10,13} = 0.85$
$T_{8,14} = 1.00$	$T_{9,14} = 0.90$	$T_{10,14} = 0.85$

routes are shown in Tables 8, 9 and 10. A supply route with more hurdles and less number of units per unit supply increases the cost. The selection of optimal locations for the network nodes minimizes the shipping costs.

The capacities of suppliers 1, 2 and 3 are 1700, 800 and 1400 respectively. The fixed costs of the plants 4, 5, 6 and 7 are \$2000, \$3000, \$2500 and \$4000 respectively each with production capacities of 600 units. The fixed costs of the warehouses 8, 9 and 10 are \$2500, \$3000 and \$3000 with storage capacities of 1000, 1500 and 1500 units, respectively. The demands from nodes 11, 12, 13 and 14 are 700, 400, 250 and 450, respectively. Also, in the mathematical model of the selected problem, $h \in l$, $i = l + 1$, $i \in n$, $e = n + 1$, $e \in t$, $j = t + 1$ and $t \in m$.

The result produced shows that both the shipping firms are involved in product flow across each stage of the multi-echelon supply chain network and satisfies the upper and lower bounds. The product flow follows the cheapest route between a pair of nodes in the entire network and the constraints of each node are satisfied. The total cost \$1,032,300, includes the fixed costs of the facilities (plants and warehouses), the production costs of commodities and the shipping cost across all routes of the entire supply chain network.

In Fig. 1, the product flow across the supply routes of the multi-echelon supply chain network is shown. Both the shipping firms (A and B) are involved in shipping commodities at 10% discount. There is no need of plant 7 and warehouse 10. We have to keep these two facilities closed.

Table 8 Shipping cost from supply node to production plant

C_{hi} Cost per unit per mile		
$C_{14} = 2$	$C_{24} = 4$	$C_{34} = 6$
$C_{15} = 3$	$C_{25} = 5$	$C_{35} = 7$
$C_{16} = 6$	$C_{26} = 3$	$C_{36} = 2$
$C_{17} = 7$	$C_{27} = 4$	$C_{37} = 1$

Table 9 Shipping cost from production plant to warehouse

C_{ie} Cost per unit per mile			
$C_{48} = 2$	$C_{58} = 1$	$C_{68} = 3$	$C_{78} = 1$
$C_{49} = 4$	$C_{59} = 3$	$C_{69} = 5$	$C_{79} = 5$
$C_{4,10} = 6$	$C_{5,10} = 5$	$C_{6,10} = 7$	$C_{7,10} = 9$

Table 10 Shipping cost from production plant to warehouse

C_{ej} Cost per unit per mile		
$C_{8,11} = 2$	$C_{9,11} = 1$	$C_{10,11} = 3$
$C_{8,12} = 4$	$C_{9,12} = 3$	$C_{10,12} = 5$
$C_{8,13} = 6$	$C_{9,13} = 5$	$C_{10,13} = 7$
$C_{8,14} = 6$	$C_{9,14} = 5$	$C_{10,14} = 7$

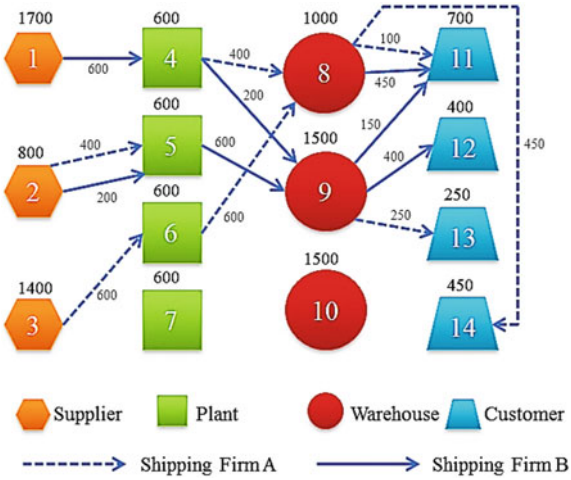


Fig. 1 The product flow across the multi-echelon supply chain network

4 Conclusion and Future Work

We have addressed an extension to the traditional multi-echelon supply chain network for product distribution in which the shipping cost linked with a route depends not only on the product flow moving across that route but also on the product flow on other routes in the network. The incorporation of traffic factor issue and discount factor broadened the study of network design and product flow. The presented computational experiments and tested a case study to assess the tractability of the proposed model.

The potential areas for future research include the generic model of this approach and inclusion of non-linear objective functions. The consideration of demand uncertainty and shipping modes may further enhance its applicability to real-world scenarios.

Acknowledgements I wish to thank Prof. Dr. Imtiaz Hakeem, Department of Mechanical Engineering, SUIT, Peshawar, for his support, helpful suggestions and critical comments.

References

1. Longinidis P, Georgiadis MC (2014) Integration of sale and leaseback in the optimal design of supply chain networks. Omega 47:73–89
2. Wang H, Mastragostino R, Swartz CLE (2016) Flexibility analysis of process supply chain networks. Comput Chem Eng 84:409–421
3. Nagurney A, Liu Z, Cojocaru MG, Daniele P (2007) Dynamic electric power supply chains and transportation networks: an evolutionary variational inequality formulation. Transp Res Part E 43:624–646

4. Gill A (2011) A supply chain design approach to petroleum distribution. Int J Bus Res Manage 2(1):33–44
5. Cohn A, Davey M, Schkade L, Siegel A, Wong C (2008) Network design and flow problems with cross-arc costs. Eur J Oper Res 189:890–901
6. Eliiyi U, Nasibov E, Özkılçık M, Kuvvetli U (2012) Minimization of fuel consumption in city bus transportation: a case study for Izmir. Procedia-Soc Behav Sci 54:231–239
7. Zhu X, Garcia-Diaz A, Jin M, Zhang Y (2014) Vehicle fuel consumption minimization in routing over-dimensioned and overweight trucks in capacitated transportation networks. J Clean Prod 85:331–336
8. Leblanc LJ, Boyce DE (1986) A bi-level programming algorithm for exact solution of the network design problem with user-optimal flows. Transp Res 3:259–265
9. Machado CMS, Mayerle SF, Trevisan V (2010) A linear model for compound multicommodity network flow problems. Comput Oper Res 37:1075–1086
10. Bozorgirad S, Desa MI, Wibowo A (2012) Genetic algorithm enhancement to solve multi-source multi-product flexible multi-stage logistics network. IJCSI Int J Comput Sci 9 (3):2, 157–164
11. Kalaitzidou MA, Longinidis P, Georgiadis MC (2015) Optimal design of closed-loop supply chain networks with multifunctional nodes. Comput Chem Eng 80:73–91
12. Barnhart C, Hane CA, Vance PH (2000) Using branch-and-price-and-cut to solve origin-destination integer multicommodity flow problems. Oper Res 48(2):318–326
13. Nielsen CA, Armacost AP, Barnhart C, Kolitz SE (2004) Network design formulations for scheduling U.S. Air Force channel route missions. Math Comput Model 39:925–943
14. Pant RR, Prakash G, Farooquie JA (2015) A framework for traceability and transparency in the dairy supply chain networks. Procedia-Soc Behav Sci 189:385–394
15. Yolmeh A, Salehi N (2015) An outer approximation method for an integration of supply chain network designing and assembly line balancing under uncertainty. Comput Ind Eng 83:297–306

Study on the Free Trial of IT Services from Users' Decision Perspective: A Conceptual Framework

Weiling Jiao, Hao Chen and Yufei Yuan

Abstract Nowadays, information technology services (ITS) are offered to individual end users from free to fee, free trial has become a necessary and crucial promotion strategy to many information technology companies, which calls for advancing our understanding on free trial phenomena. In this paper, we define free trial as an inseparable whole process including the initial trial, experience, and fared use stage from users' decision prospective, and build a theoretical framework conceptualizing the dynamics surrounding users' decision making in free trial. Using this framework, we further develop the research model; investigate corresponding decision variables and present propositions. The integrated conceptual framework provides a comprehensive understanding of users' adoption, purchase intentions, and behaviors for ITS promoted by free trial.

Keywords Free trial · IT service · Behavior dynamics · Three-stage framework

W. Jiao (✉)
Economics and Management School, Yancheng Institute of Technology, Yancheng 224051, Jiangsu, China
e-mail: jiaoweiling@foxmail.com

H. Chen
Faculty of Management and Economics, Dalian University of Technology, Dalian 116023, Liaoning, China
e-mail: ch9569@mail.dlut.edu.cn

Y. Yuan
DeGroote School of Business, McMaster University, Hamilton L8S 4M4, ON, Canada
e-mail: yuanyuf@mcmaster.ca

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_37

1 Introduction

Free trial of information technology service (ITS) refers to the practice in which IT companies offer users free access to experience their ITS, either with restricted functionality or with a full-fledged version in a limited trial period, and allow the users to opt out or stay after trial without penalty [1, 2]. Examples are abundant especially in software and high-tech services, such as Kindle Unlimited eBook, and online movies Netflix. Free trial is a behavioral influence strategy in marketing, which is guided by the assumption that direct experience with a product or service constitutes an important basis for subsequent purchase action [3]. Unlike physical goods, technology-based ITS has peculiar properties of uncertainty [4], which reduces users' willingness to adopt these ITSs [5]. Free trial can significantly ease users' uncertainty to the quality of ITS because it is easier for users to produce positive attitude and purchase intention via gaining personal experience from trial [4].

Our study of the free trial of ITS is different from previous studies of IT adoption and continued usage. Firstly, when ITS become consumer goods, individual users need to consider more about purchase decision [2] rather than adoption. Secondly, IT innovation adopters should be treated as consumers rather than simply as adopters. The major difference is that consumers will estimate the value of each alternative based on a trade-off of the benefits and sacrifices [6]. However, traditional adoption model (e.g. TAM) fails to capture it [7]. Thirdly, factors that influence the formation of users' beliefs have changed in free trial context. The price of paid ITS need to be considered [4]. The existing IT adoption and post-adoption studies do not specifically examine how their beliefs can be influenced in the first place [8]. To the best of our knowledge, an integrated theoretical framework based on the whole process level has not been reported in information system literature.

To fill these gaps, we built an integrated conceptual framework to take the free trial of ITSs as a whole process rather than as several independent stages. We integrate TAM, the value-based adoption model (VAM) and expectation confirmation theory (ECT) to reflect the nature of benefits and sacrifices evaluation and to explain users' rational behavior during the whole process of free trial. We also identify and clarify relating factors which may have impact on users' beliefs of willingness to trial (WTT) before trial and the willingness to pay (WTP) after trial.

2 Conceptual Framework for Free Trial of ITS

We suggest the free trial of ITS is a dynamics process with three stages linked closely, including the initial trial stage, the experience stage, and the fared use stage. The free trial conceptual framework is shown in Fig. 1.

Fig. 1 Conceptual framework of the free trial of ITSs

In the initial trial stage, the WOMs of these ITSs are helpful to decide whether spending time and energy to trial. Users also need to *evaluate of the usability* to judge whether the ITSs are usefulness and easy to use, and estimate the *trial trade-off* of benefits and costs then to decide whether the free trial is worth adopting. In the experience stage, *experience evaluation* of the utility and immersion will helpful to judge whether the quality and value will meet their expectation based on their own perception and WOM (*confirmation*). Then, users decide whether *satisfied* with the trial. In the fared use stage, users need further consider more to make decision on WTP: (1) *added value assessment.* Whether there is more values (e.g. not constrained by time and function) they could get from paid ITS; (2) *price fairness evaluation.* Whether the price of the paid ITS is charged reasonable and acceptable.

The three stages of free trial are interdependent and interactional. The initial trial stage is the beginning and premise of the following two stages. Only when users intend to adopt free trial, i.e. the users have a high level of WTT, will they enter into the following trial stage. The experience stage is the link between users' different decisions before and after trial. It is firstly the adoptability behavior forming stage after adoption decision. If users find that they themselves are not able to adapt, they will likely to choose to give up. Secondly, users' direct experience in this stage will be the foundation and basis of their following purchase decision [3]. Thirdly, experiences in this stage are more likely to be reported back in the form of WOM to those potential users who are considering whether to accept the free trial or not. The positive WOM can help IT companies to acquire new users [9], while negative WOM will accelerate potential users to refuse free trial [10]. At last, the fared use stage is the result of the previous two stages as well as the beginning of a new round of free trial. If they refuse to purchase, they may abandon the ITS or choose to retrial.

3 Research Model Development

Based on the above work, we developed a research model shown in Fig. 2. There are several variables that determine users' decisions and efforts in each corresponding stage, while the decisions and efforts determine the stage transition.

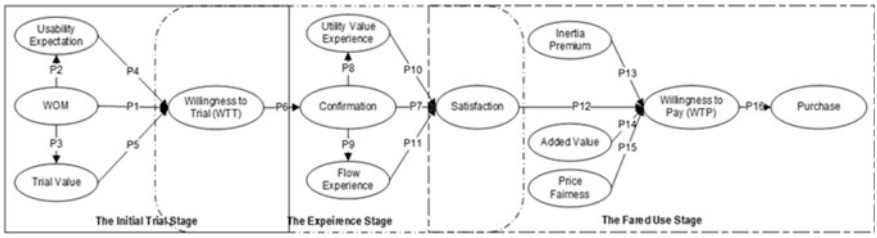


Fig. 2 Research model of users' decision for free trial of ITS

3.1 Additional Variables and Propositions for Users' WTT

Users who have experience of the ITS use and purchase will generate WOMs through social medias for others to reference. WOM has become a primary reference for users and user-generated information plays a significant role not only on users' adoption and purchase decision directly [8]. In this case, though WOM impacts users' intentions during the whole process of free trial, we take it as a crucial variable to determine users' free trial adoption decision before trial. Thus, we suggested that:

P1. WOM is positively associated with WTT.
P2. WOM is positively associated with usability expectation.
P3. WOM is positively associated with trial value.

Usability expectation means the degree to which new technology is estimated by users as being technically operable, including: *perceived usefulness* and *perceived ease of use*. The TAM proves that perceived usefulness and perceived ease of use are two fundamental drivers to users' behavioral attitude and intention [11]. In a free trial context, users may be concerned with the usefulness and ease of use of technology based ITS before they decide to avail the free trial. Therefore, we suggested that:

P4. Usability expectation is positively associated with WTT.

Base on VAM [7], the trial adoption intention is determined by the net value of a comparison between trial benefits and sacrifices. Trial benefits contains *free and no commitment* [12], *uncertainty reduction* [9], and *gratification* [13]. Trial sacrifices include *researching and learning costs* [14, 15] and *privacy and security risks* [16, 17]. Therefore, in line with past studies, we propose the following proposition:

P5. Trial value is positively associated with WTT.

3.2 *Variables and Propositions for Users' Expectation Confirmation*

The main purpose of users to trial is to make confirmation whether the actual personal experience could match their expectation. Based on the ECT [18], users have a trendy to make an evaluation to compare perceived performance with the initial expectation.

P6. WTT is positively associated with confirmation.

The ECT suggested that higher satisfaction can be observed when expectation is either high [18]. That is to say, users tend to be more satisfied when their expectations are confirmed. We propose that:

P7. Confirmation is positively associated with satisfaction.

When users become consumers, they will expect to receive both utilitarian and hedonic benefits from usage. Therefore, perceived performance in this model should be substituted by value experience and flow experience so as to reflect the utilization and entertainment benefits. Based on prior studies [19], confirmed users tend to evaluate experienced value and flow highly, and a feeling of disconfirmation will reduce such perception. Hence, we propose that:

P8. Confirmation is positively associated with utility value experience.
P9. Confirmation is positively associated with flow experience.

Utility value experience refers to the perceived ITS utilitarian outcomes that users gain from trial for utilization purposes. ECT suggested the positive relationship between perception and satisfaction [19]. Therefore,

P10. Utility value experience is positively associated with satisfaction.

Flow experience refers to the perceived ITS hedonic outcomes that users gain from trial experience for entertainment purpose. Flow experience is a sensation that occurs as a result of significant cognitive involvement, which determines users' satisfaction and future behavior [20]. Thus, we propose that:

P11. Flow experience is positively associated with satisfaction.

Satisfaction is the sense of pleasure or appointment after the comparison of the perceived effects and expectations. Prior studies found the satisfied users possess a higher purchase intentions compared with those who are dissatisfied [19]. Thus,

P12. Satisfaction is positively associated with WTP.

3.3 Variables and Propositions for Users' WTP

Inertia premium implies a reduction of efforts and costs if users continue to use the ITS after completing the free trial [21], which may urge users to become paying adopters for the following reasons: (1) *Efforts reduction.* Continued usage does not require any installment efforts because the ITSs are already in place [2]; (2) *Switching costs.* Users have to invest additional time and energy to learn and familiar with the new ITS. Therefore, we propose that:

P13. Inertia premium is positively associated with WTP.

Perceived added value refers to the availability of additional usage time and more or additional functions or features from paid ITS compared with those of free trial. Users usually will evaluate this factor before making a purchase decision [22]. Besides, Price fairness is one of the important variables that determines adoption intention in the development of IT innovation [22]. We propose that:

P14. Added value is positively associated with WTP.
P15. Price fairness is positively associated with WTP.

According to [23], users will be more willing to subscribe and select the paid ITS because of their high WTP. Hence, we propose that:

P16. WTT is positively impact on the actual behavior to purchase the paid ITS.

4 Conclusions

Our study built a conceptual framework to provide a comprehensive understanding of the dynamics around users' adoption and purchase decision in free trials of ITSs. Free trial is an integrated process with interdependent and interactional three stages. WTT, WTP, and satisfaction are three important requirements for stage transition. We also clarify relevant decision problems and corresponding efforts in different stages on the whole process from users' decision perspectives, and then investigate and present the relating decision variables and propositions.

Our study makes both theoretical contribution and practical implications. Our framework conceptualizes the dynamics surrounding the free trial towards ITS in the areas of IS adoption and purchase to continue to use and demonstrated the relationship and requirement of stage transition clearly, which can help researchers in IS to identify the achievement and limitation of existing research and to make better understanding of the integrated process and decision making in free trial. In practice, this study can help IT companies to understand users' decision making process completely and to understand the main important factors of the free trial strategy so as to fully utilize the advantage of free trial and select the best strategy for its success. This study also provides some guidelines in which users can take advantage of free trial and make right choices for paid ITS.

Acknowledgements Science and Technology Department of Jiangsu Province: Study on the integration platform development and key technology of big data on industry in the context of intelligent cloud manufacturing

References

1. Cheng HK, Tang QC (2010) Free trial or no free trial: optimal software product design with network effects. Eur J Oper Res 20(52):437–447
2. Foubert B, Gijsbrechts E (2016) Try it, you'll like it—or will you? The perils of early free-trial promotions for high-tech service adoption. Mark Sci 35(5):810–826
3. Scott CA (1976) The effects of trial and incentives on repeat purchase behavior. J Mark Res 13(3):263–269
4. Zhu DH, Chang YP (2014) Investigating consumer attitude and intention toward free trials of technology based services. Comput Hum Behav 30:328–334
5. Cheng HK, Liu Y (2012) Optimal software free trial strategy: the impact of network externalities and consumer uncertainty. Inf Syst Res 23(2):488–504
6. Kim HW, Chan HC, Gupta S (2007) Value-based adoption of mobile internet: an empirical investigation. Decis Support Syst 43(1):111–126
7. Kahneman, Tversky A (1979) Prospect theory: an analysis of decision under risk. Econometrica (47):263–291
8. Oestreicher SG, Zalmanson L (2013) Content or community? A digital business strategy for content providers in the social age. MIS Q 37(2):591–616
9. Oh H (2016) Free versus for-a -fee of a paywall on the pattern and effectiveness of word-of-mouth via social media. MIS Q 40(1):31–56
10. Bone PF (1995) Word-of-mouth effects on short-term and long-term product judgments. J Bus Res 32(3):213–223
11. Davis FD (1989) Perceived usefulness, perceived ease of use, and users acceptance of information technology. MIS Q 13(3):319–340
12. Sriram S, Chintagunta PK, Manchanda P (2015) Service quality variability and termination behavior. Manage Sci 61(11):2739–2759
13. Bicen P, Madhavaram S (2013) Research on smart shopper feelings. Mark Theor Pract 21 (2):221–234
14. Liu CZ, Yoris AA, Choi HS (2014) Effects of freemium strategy in the mobile app market: an empirical study of google play. J Manage Inf Syst 31(3):326–354
15. McKinney V, Yoon K, Zahedi F (2002) The measurement of web-customer satisfaction: an expectation and disconfirmation approach. Inf Syst Res 13(3):296–315
16. Cheng HK, Li S, Liu Y (2015) Optimal software free trial strategy: limited version, time-locked, or hybrid? Prod Oper Manage 24(3):504–517
17. Vinod P, Jaipur R, Laxmi V (2009) Survey on malware detection methods. In: Proceedings of the 3rd international conferences on software, pp 74–79
18. Bhattacherjee A (2001) Understanding information systems continuance: an expectation confirmation model. MIS Q(25):351–370
19. Oliver RL, DeSarbo WS (1988) Response determinants in satisfaction judgments. J Consum Res 14:495–507
20. Csikszentmihalyi M (2000) The costs and benefits of consuming. J Consum Res 27(2): 267–272
21. Su X (2009) A model of consumer inertia with applications to dynamic pricing. Prod Oper Manage 18(4):365–380

22. Bolton LE, Warlp L, Alba JW (2003) Customer perceptions of price (un)fairness. J Consum Res 29:474–491
23. Bhattacherjee A, Premkumar G (2004) Understanding changes in belief and attitude toward information technology usage: a theoretical model and longitudinal test. MIS Q 28(2): 229–254
24. Lin K, Lu H (2011) Why people use social networking sites: an empirical study integrating network externalities and motivation theory. Comput Hum Behav 27:1152–1161
25. Lynch R, Dembo M (2004) The relationship between self-regulation and online learning in a blended learning context. Int Rev Res Open Distrib Learn 5:1–6

The Impact of Transformational Training Programs on Employee Loyalty: A Structural Equation Modeling

Nidal Fawwaz Al Qudah, Yang Yang, Saad Alsaidan, Syed Asad Ali Shah and Syed Jamal Shah

Abstract This study aimed to investigate the impact of transformational training programs on the loyalty of employees. The research sample consisted employees working in private Jordanian universities (deans, heads of departments and faculty members). The sample was obtained by simple random sampling to avoid bias. The total sample size comprised 625 university employees. The study applied both inferential and descriptive statistics in data analysis. SPSS and AMOS were used to do the analysis. The study revealed that there is significant relationship between Reaction and Loyalty to supervisor, Loyalty to working group and Loyalty to organization; there is a significant relationship between Behavior and the Loyalty to supervisor, Loyalty to working group and Loyalty to organization; there is a significant relationship between Results and Loyalty to supervisor and Loyalty to working group and Loyalty to organization. Finally, the results found that there is an insignificant relationship between Behavior and Loyalty to supervisor, Loyalty to working group and Loyalty to organization. The study recommends the extension of this research to examine the moderating role of employees' satisfaction on the relationship between transformational training programs and employees' loyalty.

Keywords Employee loyalty · Structural equation modeling · Transformational training programs

N. F. Al Qudah (✉) · Y. Yang · S. Alsaidan · S. A. A. Shah · S. J. Shah
School of Management, Harbin Institute of Technology, Harbin, China
e-mail: naq682002@yahoo.com

Y. Yang
e-mail: yfield@163.com

S. Alsaidan
e-mail: s_alsaidan@hotmail.com

S. A. A. Shah
e-mail: asadshah2014@gmail.com

S. J. Shah
e-mail: jimmikakakhail@stu.hit.edu.cn

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_38

1 Introduction

Loyalty is a two-way path. If an organization desires its employees to be loyal, they must earn it by creating a stable and challenging workplace [1]. Many scholars also agree that loyalty is an emotional assurance of an employee's ambition and determination to maintain a constant and responsible association with a specific director due to their satisfaction with the support they receive [2].

Loyalty develops through a firm emotional relationship and any form of temporary dissatisfaction with the job is endured [3]. This means that the fundamental principle underlying the concept of employee loyalty is emotional attachment [4]. Loyalty is determined from two approaches: the attitudinal approach and the behavioral approach. In the attitudinal approach, loyalty refers to a psychological inclination to feelings, identification, attachment or commitment to the organization [5]. Loyalty has been suggested as one of four possible reactions to dissatisfaction, the other three being exit, voice, and neglect; Naus (2007) adds cynicism to this list [6]. Training programs that are consistent with the goals, needs of employees and organizations and which fit with the business strategy, will meet with greater success than those that are not [7]. However, studies have indicated positive effects of training programs on employee satisfaction and suggest that there is a strong relationship between employee loyalty and employee satisfaction [8].

We suppose that there is a positive relationship between transformational training program and Employee Loyalty. As well, there is still a lack of research analyzing the relationship between transformational training programs and Employee Loyalty. Therefore, the basic assumption of our research is that transformational training programs and employee loyalty can help universities achieve high quality orientation of their employees. More specifically, transformational training programs directly affect employee loyalty. The theoretical basis for this research is founded on two cognitive dialectics. It suggests that transformational training programs are the best path for employee loyalty. Employee loyalty enables universities to mediate between their resources and the requirements of a competitive environment.

1.1 Objectives of the Study

The objective of the study is to investigate the effect of transformational training programs (Reaction, Learning, Behavior and Results) on employee loyalty (Loyalty to Supervisor, Loyalty to Working Group and Loyalty to Organization).

1.2 Research Question

Based on the research objectives, the following research question need to be answered:

Do different dimensions of transformational training programs (Reaction, Learning, Behavior and Results) have any effect on employee loyalty (Loyalty to Supervisor, Loyalty to Working Group and Loyalty to Organization)?

2 Literature Review

2.1 Transformational Training Programs

Training program provides the trainer with a determine level of knowledge or skills designed to shape a desired behavior or action in the process [9]. The importance of training transfer through explaining the theories of motivation, the ability to transfer, and the trainer's attitudes, He found that regardless of whether the transfer effect was positive or negative, the participant is not the only factor that affects the transfer, but it is also influenced by other necessary factors [10].

Kirkpatrick identifies the four levels of evaluation in Kirkpatrick's model as reaction criteria, learning criteria, behavioral criteria, and results criteria [11]. The reaction criteria are trainees' perceptions of the training program in organizations [11]. Learning criteria in higher education are assessed by the learning outcome and evaluated by applying different forms of knowledge tests by which performance, presentation, and demonstration of skill in the training context is measured [12]. Behavioral criteria levels gauge the performance of actual jobs to assess the influence of training on job performance [13]. Results criteria in organizational settings are measured by efficiency and productivity achievements, increased satisfaction of customers, and increased profitability [14].

2.2 Employee Loyalty

Loyalty is a two-way path. If an organization desires its employees to be loyal, they must earn it by creating a stable and challenging workplace [15]. Many scholars also agree in giving a common statement that loyalty is an emotional assure of an employee's ambition to involve and remain a determinedly constant and responsible association with a specific director due to the satisfaction about the support [2]. Martensen and Grønholdt (2006) assert that the loyal develops through a firm emotional relation and any form of temporary dissatisfaction with the job is endured, which means that the fundamental principle underlying the concept of employee loyalty is that of emotional attachment [3]. Cook divided employee loyalty into active and passive, the active loyalty refers to the subjective staff loyal to the organization with the desire.

This desire is often due to a high degree of organization and employee goals and now there are consistent with organizational help for employees 'self-development and self-realization factors. Passive loyalty is when the employees themselves do not wish to remain in the organization, but due to some constraints, such as high

wages, etc., have to stay in the organization [16]. Meschke present a tripartite employee loyalty term that included three reference objects: supervisor, working group, and organization [17]. The three reference objects could be of crucial importance for investigations into employee loyalty outcomes, as loyalties towards different reference objects may conflict with each other [18].

2.3 Relationship Between Transformational Training Programs and Employee Loyalty

Sivanathan et al. [19] Explore whether and how transformational training interventions can improve occupational safety. A pre-test/post-test quasi-experiment with swimming pool supervisors (n = 18) and their swim instructors (n = 39) was conducted to test whether a safety-specific transformational training program intervention for supervisors affected instructors' safety behaviors. It revealed that, when compared to the control group (8 supervisors, 13 instructors), the intervention significantly improved instructors' perceptions of their supervisors' transformational training behaviors, maintained instructors' safety compliance, and improved instructors' safety participation at a 6-week follow up. A change in instructors' perceptions of their supervisors' transformational training behaviors served as the mechanism by which changes in the safety behaviors occurred.

Another study evaluated the effects of a transformational training program on Unit Charge Nurses' leadership practices. Based on the Practices Inventory-Self and Observer ratings, the results indicate that leadership practices increased statistically significantly after the implementation of the program. The Unit Charge Nurses Leadership Practices Inventory self-ratings were significantly higher than those of the observers [20].

Another study explores the relationship between training and employees' commitment to their organization. The study is based on a survey of 250 employees and management staff of a financial firm based in southwestern Nigeria [21]. The analysis revealed some evidence suggesting a significant positive statistical relationship between different levels of training and employees' commitment to the organization. The study revealed a significant positive statistical relationship between the different levels of training and employees' commitment to the organization.

3 Methodology

3.1 The Research Model and Hypotheses

Our research model is shown in Fig. 1, which also shows the definitions of various constructs in the model. In this study, the author constructs the research model based on reviewing the selected literature on Transformational Training Programs and Employee Loyalty.

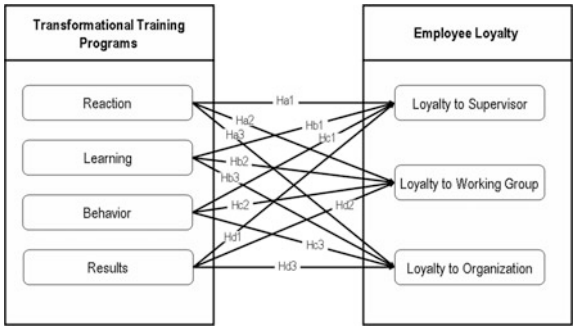


Fig. 1 Research model

Based on the above discussions, the researcher proposes the following hypotheses:

Ha1. The Reaction of Transformational Training Programs positively affects Loyalty to Supervisor.
Ha2. The Reaction of Transformational Training Programs positively affects Loyalty to Working Group.
Ha3. The Reaction of Transformational Training Programs positively affects Loyalty to Organization.
Hb1. The Learning of Transformational Training Programs positively affects Loyalty to Supervisor.
Hb2. The Learning of Transformational Training Programs positively affects Loyalty to Working Group.
Hb3. The Learning of Transformational Training Programs positively affects Loyalty to Organization.
Hc1. The Behavior of Transformational Training Programs positively affects Loyalty to Supervisor.
Hc2. The Behavior of Transformational Training Programs positively affects Loyalty to Working Group.
Hc3. The Behavior of Transformational Training Programs positively affects Loyalty to Organization.
Hd1. The Results of Transformational Training Programs positively affects Loyalty to Supervisor.
Hd2. The Results of Transformational Training Programs positively affects Loyalty to Working Group.
Hd3. The Results of Transformational Training Programs positively affects Loyalty to Organization.

3.2 Study Population and Sample

The study population in the current study consisted of all individuals working in private Jordanian universities in the capital, Amman. A convenience sample comprised 625 deans, heads of departments and faculty members from private Jordanian universities in Amman. The sample was obtained by simple random sampling to avoid bias (Table 1).

3.3 Data Collection and Instruments

Empirical data were collected and analyzed through a quantitative approach. This approach was chosen because the current study was interested in testing the validity of the questionnaire. As well, the research considered the most suitable technique to measure the quantitative data [17]. A structured survey questionnaire was prepared in order to obtain unambiguous answers to specific questions. The questions were mainly closed, which allowed subsequent statistical analysis. Likert Scale-type ratings were mainly based on a scale ranging from 1 (strongly disagree) to 5 (strongly agree). The data was then transferred to AMOS version 22 for further analysis. The researchers measured employee loyalty using measurements developed by Meschke [17]. As well, to measure the Transformational Training Programs, the researchers adapted Al-Qudah and others' [22] scale.

3.4 Data Analysis

The study applied Structural Equation Modeling (SEM) in data analysis, using the AMOS programming to do the analysis.

Table 1 Research sample

No	University	No. of college	Founded
1	Applied Science Private University	8	1989
2	Al-Isra University	10	1989
3	Al-Ahliyya Amman University	8	1989
4	Princess Sumaya University for Technology	4	1991
5	Petra University	8	1991
6	Al-Zaytoonah University of Jordan	7	1993
7	Amman Arab University	7	1999
8	Middle East University	11	2005

3.5 *Measuring Reliability*

A reliability test was performed one more time to verify whether the Transformational Training Programs and employee loyalty constructs were reliable. The reliability results (shown in Table 2) clarify that all constructs were higher than the acceptable level stated by Murtagh et al. [23].

4 Analysis and Results

Structural equation modeling (SEM) was conducted to test the hypotheses in order to assess the effect and the significance level of each path in the model. The model shown in Fig. 1 was examined and tested using AMOS 22.0. The model fit determines the degree to which the structural equation model fits the sample data. However, the model was tested and provided good indicators of fit as: $\chi 2/df = 1.925$. The goodness of fit index (GFI) and adjusted goodness of fit index (AGFI) was 0.994 and 0.927; the normed fit index (NFI) was 0.931. The Tucker-Lewis coefficient (TLI) was 0.924, the comparative fit index (CFI) was 0.973 and the root mean square error of approximation (RMSEA) was 0.037, indicating a good fit between the theoretical model and the data. Table 3 presents each parameter's C.R., Estimate and S.E.

Hence, Reaction has a significant effect on Loyalty to Supervisor, Loyalty to Working Group and Loyalty to Organization at private Jordanian universities ($\beta = 0.118$, 0.309 and 0.346; C.R = 3.933, 8.351 and 6.528; *P*-value = ***, *** and ***). Ha1, Ha2 and Ha3 are therefore supported.

Learning does not a significant direct impact on Loyalty to Supervisor, Loyalty to Working Group and Loyalty to Organization at private Jordanian universities ($\beta = 0.0.071$, −0.033 and 0.065; C.R = 1.651, −0.635 and 0.867; *P*-value = 0.095, 0.511 and 0.187) or Hb1, Hb2 and Hb3 are therefore not supported.

Table 2 Reliability analysis results for measurement items

Construct (or factor)	Cronbach's alpha	
	No of items	Value
Reaction	3	0.709
Learning	4	0.721
Behavior	3	0.705
Results	3	0.711
Transformational training programs	13	0.824
Loyalty to supervisor	6	0.797
Loyalty to working group	6	0.805
Loyalty to organization	6	0.770
Employee loyalty	18	0.843

Table 3 Hypotheses testing result

Hypothesis	Regression weights		Estimate	SE	C.R.	P Value	Results
	From	To					
Ha1	Reaction	LS	0.118	0.030	3.933	***	Accepted
Ha2	Reaction	LW	0.309	0.037	8.351	***	Accepted
Ha3	Reaction	LO	0.346	0.053	6.528	***	Accepted
Hb1	Learning	LS	0.071	0.043	1.651	0.095	Not accepted
Hb2	Learning	LW	-0.033	0.052	-0.635	0.511	Not accepted
Hb3	Learning	LO	0.065	0.075	0.867	0.187	Not accepted
Hc1	Behavior	LS	0.284	0.034	7.294	***	Accepted
Hc2	Behavior	LW	0.339	0.041	8.268	***	Accepted
Hc3	Behavior	LO	0.313	0.060	5.217	***	Accepted
Hd1	Results	LS	0.423	0.029	14.586	***	Accepted
Hd2	Results	LW	0.162	0.036	4.500	***	Accepted
Hd3	Results	LO	0.077	0.052	1.481	0.073	Not accepted

Behavior has a significant effect on Loyalty to Supervisor, Loyalty to Working Group and Loyalty to Organization at private Jordanian universities (β = 0.284, 0.339 and 0.313; C.R = 7.294, 8268 and 5.217; *P*-value = ***, *** and ***) or Hc1, Hc2 and Hc3 are therefore supported.

Results has a significant effect on Loyalty to Supervisor, Loyalty to Working Group at private Jordanian universities (β = 0.423 and 0.162; C.R = 14.586 and 4.500; *P*-value = *** and ***) or Hd1 and Hd2 are therefore supported. Finally, Results does not have a significant direct impact on Loyalty to Organization at private Jordanian universities (β = 0.077; C.R = 1.481; *P*-value = 0.073) or Hd3 is therefore not supported.

5 Discussion

From the analysis, it was observed that there are significant relationship between the Reaction and the Loyalty to supervisor, Loyalty to working group and Loyalty to organization. This was in line with the research that found that employee loyalty is a deliberate commitment to further the best interests of one's employer, even when meeting so many demands involves sacrificing some aspects of one's self-interest beyond what would be required by one's legal or other moral duties [24].

As well, there are significant relationships between the Behavior and the Loyalty to supervisor, Loyalty to working group and Loyalty to organization. This was consistent with the behavioral approach to loyalty, which holds that loyalty is an observable phenomenon, obvious, and materialized in the relationship between the employee and the organization [25].

We observed that there is a significant relationship between the Results and the Loyalty to supervisor and Loyalty to working group. Achievement of employee loyalty is the uppermost level of positive interpersonal relationships that a director in a corporation can attain, which means that achieving loyalty requires a positive status [26]. Kumar and Shekhar [1] demonstrated that loyalty is a two-way path. If an organization desires its employees to be loyal, they must earn it by creating a stable and challenging workplace.

Finally, the results found that there is an insignificant relationship between the Learning and the Loyalty to supervisor, Loyalty to working group and Loyalty to organization.

6 Conclusion, Limitation and Future Research

From the analysis, it can be concluded that there exists a statistically significant relationship between the dependent variables (transformational training programs) and loyalty to organization, loyalty to working group, and loyalty to supervisors (independent variables). The study found that transformational training programs made a big contribution to employees' loyalty to the universities.

The study was limited to universities in Jordan. The study was also limited to employees of the universities and no other members of the university such as students and public users. The authors recommend the extension of this research will be to examine the effect of the moderating role of employees' satisfaction on the relationship between transformational training programs and employees' loyalty.

References

1. Kumar D, Shekhar N (2012) Perspectives envisaging employee loyalty: a case analysis. J Manage Res 12:100
2. Bloemer J, Odekerken-Schröder G (2006) The role of employee relationship proneness in creating employee loyalty. Int J Bank Mark 24:252–264
3. Martensen A, Grønholdt L (2006) Internal marketing: a study of employee loyalty, its determinants and consequences. Innovative Mark 2:92–116
4. Mehta S, Singh T, Bhakar S, Sinha B (2010) Employee loyalty towards organization—a study of academician. Int J Bus Manage Econ Res 1:98–108
5. Guillon O, Cezanne C (2014) Employee loyalty and organizational performance: a critical survey. J Organ Change Manage 27:839–850
6. Naus F, Van Iterson A, Roe R (2007) Organizational cynicism: extending the exit, voice, loyalty, and neglect model of employees' responses to adverse conditions in the workplace. Hum Relat 60:683–718
7. Wexley KN, Latham GP (1991) Developing and training human resources in organizations. Harper Collins Publishers

8. Hang NTL (2016) An investigation of factors affecting employee satisfaction and loyalty in the Fire Fighting and Prevention Police of Ho Chi Minh City Vietnam
9. Dullien T (2008) Transactional or transformational training? Indian Gaming, pp 88–119
10. Alawneh MK (2008) Factors affecting training transfer: participants' motivation to transfer training, literature review, Online Submission
11. Kirkpatrick DL (1996) Invited reaction: reaction to Holton article. Human Resour Dev Q 7:23–25
12. Lievens F, Sackett PR (2012) The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. J Appl Psychol 97:460
13. Praslova L (2010) Adaptation of Kirkpatrick's four level model of training criteria to assessment of learning outcomes and program evaluation in higher education. Educ Assess Eval Accountability 22:215–225
14. Landy FJ, Conte JM (2016) Work in the 21st century, Binder ready version: an introduction to industrial and organizational psychology. Wiley, Hoboken
15. Al Qudah NF, Yang Y, Anjum MA (2018) Transformational training programs and quality orientation of employees: does employees' loyalty matter? Sustainability 10:465
16. Cook S (2008) The essential guide to employee engagement: better business performance through staff satisfaction. Kogan Page Publishers
17. Meschke S (2016) Tripartite Employee Loyalty (TEL): validation, measurement and selected outcomes of a new concept
18. Provis C (2005) Dirty hands and loyalty in organisational politics. Bus Ethics Q 15:283–298
19. Sivanathan N, Turner N, Barling J (2005) Effects of transformational leadership training on employee safety performance: a quasi-experiment study. In: Academy of management proceedings, pp N1–N6
20. Duygulu S, Kublay G (2011) Transformational leadership training programme for charge nurses. J Adv Nurs 67:633–642
21. Owoyemi OA, Oyelere M, Elegbede T, Gbajumo-Sheriff M (2011) Enhancing employees' commitment to organisation through training. Int J Bus Manage 6:280
22. Al Qudah NF, Yang Y, Li Z (2017) Perceived effectiveness of transformational training programs: dimensions, measurement and validation. In: 2017 International Conference on Industrial Engineering, Management Science and Application (ICIMSA), pp 1–6
23. Murtagh F, Heck A (2012) Multivariate data analysis vol. 131. Springer Science & Business Media
24. Elegido JM (2013) Does it make sense to be a loyal employee? J Bus Ethics 116:495–511
25. Rusbult CE, Farrell D, Rogers G, Mainous AG (1988) Impact of exchange variables on exit, voice, loyalty, and neglect: an integrative model of responses to declining job satisfaction. Acad Manag J 31:599–627
26. Klopotan I, Buntak K, Drožđek I (2016) Employee loyalty: differences between genders and the public and the private sector. Interdisc Description Complex Syst 14:303–313

What Factors Contributed to the Success of WeChat?—SIMPLE Model

Weiling Jiao, Hao Chen and Yufei Yuan

Abstract WeChat is one of the most successful social media in the world. This paper tries to explore the factors that make WeChat so attractive so as to gain real insight into the potential for current social media to be successful in other market. We built a model of SIMPLE (Socialization, Integration, Mobilization, Popularization, Localization and Ecologization) to understand the secrets that make WeChat. Our research is useful for both academics and practitioners to understand the reasons why and how to ensure that social media become attractive in mobile era.

Keywords Social media · WeChat · Success factors · SIMPLE model

1 Introduction

WeChat, known as Weixin in Chinese, was launched on January 21, 2011 and has been growing stunningly fast since then. It only spent nearly 15 months to have its first 100 million registered users [1], which is far less than many other social media, for example, Facebook took 54 months and Twitter took 49 months. Only about six years later, it has had 943 million register users in China, in addition to more than

W. Jiao (✉)
Economics and Management School, Yancheng Institute of Technology,
Yancheng 224051, Jiangsu, China
e-mail: jiaoweiling@foxmail.com

H. Chen
Faculty of Management and Economics, Dalian University of Technology,
Dalian 116023, Liaoning, China
e-mail: ch9569@mail.dlut.edu.cn

Y. Yuan
DeGroote School of Business, McMaster University,
Hamilton, ON L8S 4M4, Canada
e-mail: yuanyuf@mcmaster.ca

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_39

800 million registered users in the world according to its latest official statistical data [2]. Its average revenue per user has been estimated to be at least 7 times the one of WhatsApp, which makes WeChat becomes the largest messaging platform [3].

Due to its great attractiveness, WeChat has now inspired many oversea companies to monitor it closely. However, the world still knows less about it so that the huge potential and value owned by it just like a precious but hazy scroll. For those western social media companies who want to imitate and copy WeChat, it is possible to reap huge gains only when they can understand it clearly. Then what really is WeChat, why has it been so attractive? The objective of this paper is to provide evidence showing "SIMPLE", the core philosophy of WeChat, have contributed to attractiveness and to summarize the experiences can be learned from it.

2 What Is WeChat?

WeChat has been keeping SIMPLE as its philosophy and advocated using the simple services to meet users' basic needs in a simplicity way. We focus on this philosophy and interpret it as "Socialization, Integration, Mobilization, Popularization, Localization and Ecologization" to understand what exactly WeChat is in reality. The six degrees of SIMPLE is shown in Fig. 1.

Socialization. WeChat bridges the geographic space barriers and tries to invent novel ways for users to connect with each other. It sets up one of the largest social network and the hottest social platform in China today. WeChat has strong user

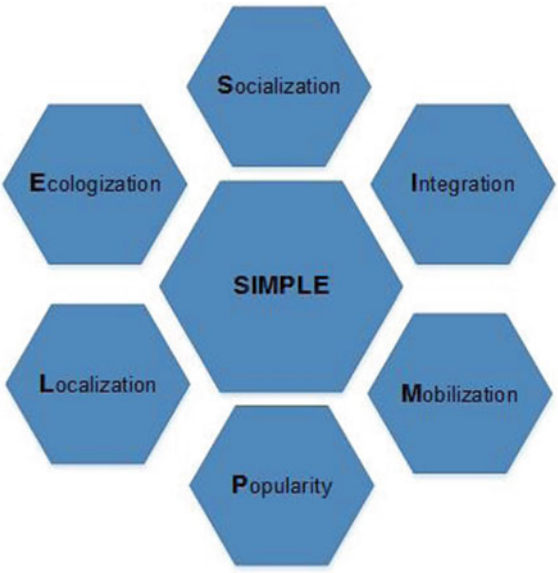


Fig. 1 Six degrees of SIMPLE

stickiness [4]. As to usage frequency, among its 756 million daily active users, more than 90% users use WeChat at least once per day. As to usage time, more than 50% users use WeChat at least 1 h per day. As to the number of friends of users, more than 50% users have more than 100 friends in their contacts list.

Integration. WeChat is easily used to do everything to facilitate users' life, such as to pay bills, book flights, replace passport, keep in touch with friends and your doctors, financial investment, and even provide heat maps to show the crowdedness of user's favorite shopping mall or popular tourist site. All of these functions are tightly integrated, thus users can use any of them at the same time without leaving WeChat. Furthermore, over 200 thousand offline stores have been integrated into WeChat platform. Users can get coupons, discounts, promotion messages, and enjoy the convenience of buying.

Mobilization. The full functions of WeChat can only be used on mobile phones. There are more than 70% Chinese mobile phone users are WeChat users and the amount has been up to 889 million by the end of 2016 [2].

Popularization. Popularization describes the high market penetration level of WeChat. WeChat is always focus on meeting the basic demand (i.e. connection and communication) of common users, not specific demands and not of minor users, which makes it gain massive adoption on a huge scale in China.

Localization. WeChat matches local information with personal preference and push them the most relative business advertisements and services information, such as group buying; as well as helps stores to attract consumers nearby. The sum of life consumption of Chinese directed by WeChat is about ¥ 11 billion [5].

Ecologization. Beyond a social instant communication tool, WeChat has been established an ecosystem, including innovation businesses, services to the people's livelihood, e-commerce and finance, and industry upgrading, integrating the area of innovative and high technology, basic city service, communication, commerce, payment, entertainment, and travel.

3 Analysis of Factors that Makes WeChat Attractive

3.1 Socialization

Flexible social strategy. WeChat carries out flexible social connectivity strategy to make users easy to connect with others. An acquaintances social circle is formed based on user's acquaintances offline and the match with users contact list in the phone. While strangers and other social circles, often connected based on LBS or through shake function (others who shake their mobile phones at the same time with you will be listed on your screen), people nearby function (users who are using WeChat within a certain area around you can be found) or drift bottle function (making friends with strangers who are randomly pushed by WeChat), make social connectivity become more interesting and less distant.

Various communication ways. WeChat provides users with various information interaction ways. WeChat is firstly an instant communication app so that it supports one-to-one and group chat in the form of text, pictures, voice, video, stickers, and locations, which makes the communication conveniently and efficiently. It also supports users to share their opinions, favorite things and recent life through the function of moments, where their friends in the contact lists often give response by sending comments or like. More than 80% WeChat users are sticky to moments [5]. WeChat enriches functions by launching Red Packets, WeRun, and WeReading to make communication more interesting.

3.2 Integration

Functions Integration. The Chinese apps represented by WeChat are tending to combine as many features as possible into one application, which is in stark contrast to Western apps. WeChat realizes functions integration by its pioneering model, named "apps within an app", also known as "app constellations", i.e. all apps are in a single and an integrated app, which makes WeChat more like a browser for mobile websites, or, arguably, a mobile operating system. For example, there are millions of lightweight apps integrated inside WeChat platform. As these apps are "lightweight", is relative to native and full-featured, users are spared the trouble of downloading separate native full-featured apps.

Information Integration. WeChat was first used as an information communication tool, and has now developed into a huge information input and output platform and created new information input and output approaches. For example, shake and scan functions can turn text, picture, music, video, and web pages into data and then add them in a form of two-dimension code into WeChat. Thus, users can create information in WeChat with lower cost and higher speed, while WeChat can offer user with useful information they wanted by means of accurate indentation and calculation.

Online & Offline Integration. WeChat is not limited itself to the virtual world but moves into the physical world. Mon-Mon, a Bluetooth-enabled animal toy, integrates WeChat with the offline world. Parents can use the Mon-Mon official account in WeChat to send personal voice messages and pre-recorded bedtime stories to the toy while they are at work or traveling. Kids get those stories or messages immediately, and can even reply to their parents' WeChat account via a voice mail. Another example is WeChat pay function can be used not only online, but also offline via QR codes (which is similar to ID of users') at any physical stores at any time. It is simple and convenient, thus it is no need for users to carry cash. Beside, compared with cash payment, there are always favorable discount offered by business. WeChat pay helps china become a cashless society.

3.3 Mobilization

Make full use of mobile technology. WeChat nearly harnesses advantages of smartphone completely and thoughtfully, from GPS location to sensors to voice to camera. For example, WeChat engages the camera to scan English text and translate it into Chinese, or to pay directly for a transaction. WeChat also better utilizes all the other smartphone sensors as sources of data input. It uses GPS when users search for businesses nearby. It calls upon the microphone to identify a TV show or a song on the radio. It uses the accelerometer when a user shakes a device to find strangers nearby to chat with. And it uses bluetooth when users add friends in their vicinity.

WeChat commerce. Based on WeChat mall, a social e-commerce systems, users can inquire, choose, experience, interact with, order and pay products online and offline by their smart-phone with their fragmented time. While enterprises make use of it to release information, carry out marketing, interact with and manage customers, offer after-sale service, conduct big data analysis, and recruit talents and so on. This commerce model provides advantages for users and enterprises, especially individual sellers and family workshops, because WeChat not only offers a platform to demonstrate merchandises and payment interface, but also provide marking channels through social relations.

3.4 Popularization

Users are King. WeChat always puts user needs on the first place and pay more attention to users' experience and demands, and take them as priority. For example, WeChat always holds back sellers' desire when design its products, which means ads aren't everywhere on WeChat. Another example is that all technology innovation of WeChat is needs-oriented. Philosophically, while Facebook and WhatsApp measure growth by the number of daily and monthly active users on their networks, WeChat cares more about how relevant and central WeChat is in addressing the daily, even hourly needs of its users. Instead of focusing on building the largest social network in the world, WeChat has focused on building a brand new mobile lifestyle for their users.

Free. Free is one of the most important factors that make WeChat gain massive adoption on a huge scale. All functions of WeChat are totally free for any users, official account developers and sellers.

Easy to Use. As WeChat target the most common people in China, its goal is to meet users demand in the simplest way and the lowest barrier. The way of information input in WeChat is the best example. WeChat creates voice memos to instead of text messaging. It solved the challenge of the Chinese typing and gave people an alternative. Besides, WeChat developed sticker and make it a new language to represent anything virtually. Stickers are animated and standalone

expressions, which are not tied to any specific alphabet and can transcend language. All the people around the world can understand the expression for sticker is a universal language.

Hedonism. The attractiveness of WeChat lies in its unique creativity, which make sure its product, is enjoyable for users to use. One of the most notable examples is red packet. It was clever of WeChat to turn dutiful gift-giving into an exciting game. It encouraged users to bind together into groups to send money, often in randomized amounts (if you send ¥ 3000 to 30 friends, they may not get ¥ 100 each; WeChat decides how much. In this case, everyone is curious about how much money I can get). Last year, over 400 m users (both as individuals and in groups) sent 32 billion red packets of digital cash during the celebration.

3.5 Localization

WeChat is engaging to integrate life-services for users as many as possible so as to ensure users can experience different kinds of service seamlessly and make users' life and entertainment is more convenient and well-organized. Traditional services offered here include mobile phone recharge, ticket purchase, entertainment, lottery purchase and so on. Some other life service, such as cards & offers, sports, community charity donation, reading, investment and so on. Many of these services seem to have changed users' life. Take WeChat sport as an example, which is increasing value of the fragmentation time of Chinese people. Nearly 25% users increase their workout, 40% users increase their interaction with friends, and 25% users pay more attention to their health [5]. It seems that WeChat have transformed into a multi-dimension service platform from social platform.

3.6 Ecologization

As the development of WeChat, it has been an ecology system. To keep old users and attract more new users, it is suggested to offer all users, including individual users, enterprises user, organization users, new value and a healthy ecological environment. WeChat advocates efficiency, fairness and justness and legality. It always insists on that to make information flow and reach more conveniently and rapidly, but to avoid information overload; to make more object carry our information input and output more efficiently, but to avoid unfair system; to encourage more users to create and communicate content to gain value and return.

4 Discussion

This research developed a theoretical SIMPLE model including all the factors contributed to the attractiveness of WeChat. Academics can use the theoretical model generated through this research with the knowledge that they have been proven via the actual practice of WeChat. Often, theoretical models are developed and then applied to real-life business situations. We have taken the opposite approach in this paper and developed a theoretical SIMPLE model based on actual business successes. This paper provides academics with a framework that can be used to examine the attractiveness of other social media. It is expected to expand on and improve the theoretical model we have developed here.

Practitioners will benefit from this research as it provides a 'roadmap' of how to achieve attractiveness in the social media industry. By applying the experiences learned and detailed in this paper, social media businesses can have a much higher likelihood of WeChat.

Acknowledgements Science and Technology Department of Jiangsu Province: Study on the integration platform development and key technology of big data on industry in the context of intelligent cloud manufacturing.

References

1. Chinese Internet Data Center (2013) Market analysis report of WeChat in 2012. Available at: http://www.199it.com/2013/01/15
2. CuriosityChina (2016) WeChat user data report. Available at: http://weibo.com/curiositychina?is_all=1
3. Chan C. When one app rules them all: the case of WeChat and Mobile in China. Available from: http://www.economist.com/news/business/21703428-chinas-wechat-shows-way-social-medias-future-wechats-world
4. BizNext (2016) WeChat user data report. Available at: http://www.gzhphb.com/article/13/134226.html
5. Theotokis A, Doukidis G (2009) When adoption brings addiction: a use diffusion model for social information systems. In: Thirtieth international conference on information systems, Phoenix, pp 1–13

Reducing the Risks of Occupational Dermatitis on Welding Industry: An Epidemiological Study

Marianne B. Calayag

Abstract Epidemiology is a study that concentrates on the analysis of the cause and effect of a health-related states or events (disease) to see its prevalence in a specified population (Last 1988). A critical premise of epidemiology is that disease and other health events do not occur randomly in a population, but are more likely to occur in some members of the population than others because of the risk factors that may not be distributed randomly in the population. This mostly happens in workplaces like in welding industry. Multiple and distinctive factors in welding process can cause occupational diseases on the workers including occupational dermatitis (OD). Based on the results, personal and environmental factors affect the severity and frequency of the OD among the workers. In conclusion, the research findings support the hypothesis that welding work is a risk factor for occupational dermatitis and that professional welding workers may not be sufficiently protected from dermal hazards at work. Occupational medicine and dermatology specialists and others who provide healthcare to professional welders should be aware of the health risks of performing welding work and handling specific materials and equipment and the possibility that symptoms of this cutaneous disease are under-reported and underdiagnosed in this population.

Keywords Occupational dermatitis · Epidemiology · Risk factors
Welders · Reference group

1 Introduction

Epidemiology is a study that concentrates on the analysis of the cause and effect of a health-related states or events (disease) to see its prevalence in a specified population. In other words, epidemiology deals with collective health of humans. Epidemiology offers insight into why disease and injury afflict some people more

M. B. Calayag (✉)
Industrial Engineering Department, Bulacan State University, Malolos, Philippines
e-mail: marianne_calayag@yahoo.com

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_40

than others, and why they occur more frequently in some locations and times than in others—knowledge necessary for finding the most effective ways to prevent and treat health problems. The term "epidemiology" appears to have first been used to describe the study of epidemics by the Spanish physician Villalba in Epidemiología Española.

Workers are the large contributors in the operation of the company. To maintain or improve the operation, workers should experience safe working process and workplace. One job that is often considered risky is welding. Welding provides a powerful manufacturing tool for the high-quality joining of metallic components. Multiple and distinctive factors in welding process can cause occupational diseases on the workers including Occupational Dermatitis (OD).

OD is defined as a reactive eczematous inflammation of the skin which occurs after the direct contact with a chemical but occasionally by biologic or physical agents (Holness 2014; Chew and Maibach 2003). Exposure in the workplace is responsible for a wide range of this cutaneous problem.

Long term exposure to risk factors associated in development of manifestations of OD can lead to work absenteeism, disability, increased use (and therefore cost) of medical care and pharmaceuticals, reduced quality of life and increased stress in individuals with these conditions. The aim of this epidemiological study is to identify the cause and risk factors of OD to provide information to the workers on the hazards they are facing and to develop preventive measures for the workers at risk.

1.1 Problem Statement

According to Philippine Statistics Authority, OD is one of the occupational diseases experienced by workers in the Philippines with a total recorded case of 1044 (2011) to 2792 cases (2013). Large build-up of OD cases between 2011 and 2013 (62.61%) suggest that workers who have ODs are continually increasing. It was observed that the workers from construction and welding experienced this most (Philippine Statistics Authority 2015).

With the analysis of related articles and researches, the researcher came up with the problem: "**How to minimize the exposure of the welders to the risks that affects the development of occupational dermatitis?**"

1.2 Objective of the Study

The main objective of the study is to minimize the exposure of the welders to the risk factors that affects the prevalence of the manifestations of OD among the welders. Align to this goal, the researchers laid out the following specific objectives that if attained, the main objective of this study would be concretized:

- to know the causes (or etiology) of the disease
- to identify and evaluate the risk factors associating to the disease
- to determine the severity and frequency level of the manifestations of OD among the welders; and
- to provide a basis for developing disease control and prevention measures for the workers at risk.

1.3 Rationale of the Study

Health and safety are important for the workers to sustain good and continuous production. Since welders are exposed to different chemicals and materials in the welding process, this study was conducted to assess the factors that may lead to occupational contact dermatitis (OCD). To have a different assessment to the said occupational disease, the study focused on the risk factors affecting the manifestations of OD on the workers' skin.

This study was performed to provide additional information about the possible adverse effect or the factors that may trigger the occupational contact dermatitis to generate. The understanding on the factors and possible outcome of welding fume is essential to risk assessment and development of safety precautions and prevention strategies that will be helpful to large number of population.

1.4 Scope and Limitation

Exposure on chemicals and hazardous elements present in the work process and workplace threatens the health and safety of the workers. In light to this problem, this study was conducted to assess the prevalence of manifestations of occupational dermatitis among the welders. Researcher limits population of respondents working around Bulacan, Pampanga and Quezon City with a sample size of 257.

OD was selected because of the obliviousness of the selected population on the hazardous effects of welding on the skin and the continuous increase of affected workers here in the Philippines.

2 Related Studies

Occupational dermatitis is one of the most common work-related illnesses, and has been estimated to account for up to 40–70% of the total burden of occupational disease. Overall in the general workforce, irritant occupational contact dermatitis occurs more commonly than allergic occupational contact dermatitis (Dickel 2002a,

b; Lodi 2000; Meyer 2000; Sun 1995; Turner 2007). The most common site of occupational dermatitis is hand, and it should be noted which hand and which site of the hand (dorsum, fingers, palm) was first involved. Incidence rates estimated from disease registry data have been reported to be in the order of 0.5–1.9 cases per 1000 workers per year, with wide variations in estimates between countries and specific occupations.

The true incidence is likely to be higher since occupational dermatitis is often severely under reported. Factors responsible for underreporting include: (1) under-recognition of work-relatedness of the symptoms; (2) reporting discrepancies arising from distinctions between observable and compensable skin disease; and (3) the perception that occupational dermatitis is trivial because it is not life threatening.

To diagnose an occupational dermatitis carefully, the researcher must observe history of dermatitis and occupation, as well. All of these information must be achieved by exact questions: description of present illness, current and previous job and description of duties, water exposure, protective device usage, affection of other co-workers, out of workplace exposures, history of atopy previous skin disease, and use of medications, factor for development of OD in the workplace is atopy, and the most common predisposing occurring in 15–20% of population. Dry skin and advancing age are important in this field, as well. Diagnosis of a work-related skin disease requires more time than a general dermatologic workup, and a premature diagnosis before studying all the evidences should be avoided because an incorrect diagnosis can have detrimental effects. Patch testing is a well-established and widely used method of diagnosing occupational dermatitis. Low concentration of allergens is applied to the upper back in individual square plastics or round aluminum chambers. They are kept in place with adhesive tape and then left in place for 48 h. When the patches are removed, the extent of the reaction (a red, raised itchy spot underneath the disc) is recorded and another reading is performed 2–3 days later (Fig. 1).

Reading of a patch test is based on morphological criteria and should be performed by an adequately trained patch test reader. Most studies indicate that readings on day 3–4 (D 3–4) and on day 7 (D7) give the most reliable results. Late-appearing positive patch test reactions can appear for most allergens and are common for some. These reactions will be missed if only one reading is performed.

Many factors can induce irritant reactions such as intrinsic nature of a substance (i.e., pH, solubility, physical state and concentration), environmental factors (i.e., temperature, humidity, and pressure), and predisposing individual factors (i.e., age, gender, ethnicity, concurrent and pre-existing skin disease, and skin region exposed). However, patch testing is routinely performed by applying a baseline series of the most frequently occurring allergens. The European baseline series contain 28 items but some of them are mixes, which mean that the number of allergens is greater than 28. National and international occupational dermatitis groups evaluate the series regularly. The baseline series detect approximately 75–80% of all contact allergens (Fig. 2).

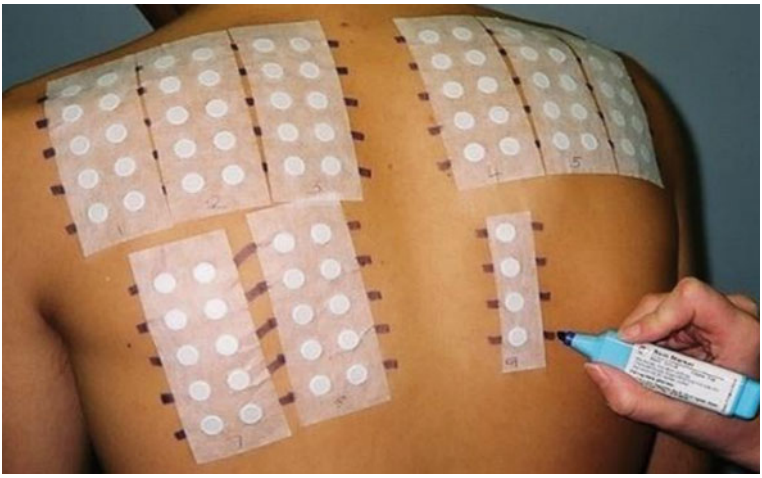


Fig. 1 Example of patch testing performed (Skin & Cancer Foundation 2014)

+?	Doubtful reaction; faint erythema only
+	Weak positive reaction; erythema, infiltration, possibly papules
++	Strong positive reaction; erythema, infiltration, papules, vesicles
+++	Extreme positive reaction; intense erythema and infiltration, coalescing vesicles
-	Negative reaction
IR	Irritant reaction

Fig. 2 Recording of patch test reactions (IODRG 2014)

2.1 The Skin as a Barrier

When thinking about the role of skin exposure in the development of sensitization and potentially the development of subsequent respiratory disease, the role of the skin barrier must be considered. It is possible that damage to the skin either through physical, chemical, or biological processes may hamper the skin's ability to block penetration of allergens.

For patients with OD, the skin barrier is degraded from the onset; alterations persisted throughout the disease course and cause chronic skin lesions. The most important agents that modify the skin barrier are: water, detergents, solvents, chemicals, dry skin, and dehydration. The most important protective mechanism in

OD is an intact skin barrier. Destruction of the skin barrier entails increased permeability of the skin, opening the way for the penetration of microbial agents, allergens, and irritants. It starts as an inflammatory reaction in the skin, with the possibility of cutaneous immune activation.

Any injury to the skin (erosion, ulceration, fissure, xerosis, burning, etc.) may precede the occurrence of OD lesions by altering the skin barrier and removing its protective role.

2.2 Exposure-Response Relationships

The relationship between airborne exposure and respiratory disease has been studied in numerous workplaces for a wide variety of exposures and outcomes ranging from reported symptoms to confirmed clinical diagnoses of respiratory disease (Jaakkola et al. 2009; Heldal et al. 2010; Lillienberg et al. 2010; Smit et al. 2008; Pronk et al. 2007; Jacobs et al. 2008; Cox-Ganser et al. 2009).

An earlier study (2007) by Van Wendel do Joode et al. investigated the association between skin symptoms and exposure to semi-synthetic metalworking fluids (de Joode et al. 2007). Skin exposure was measured in two ways: first by a semi-quantitative dermal exposure assessment tool, DREAM (Van-Wendel-de-Joode et al. 2003), and second using a quantitative tracer method, VITAE (Fenske et al. 1986a, b). Skin outcomes were also measured in two ways: first a standard symptoms questionnaire and second, a picture based screening list. In separate models, exposure (yes/no) and workers with high exposure (none/low/high) had increased prevalence ratios (PR) for reporting skin symptoms on their hands, forearms or face (PR range 2.3–2.4) (de Joode et al. 2007).

The pathway from environmental contaminant onto, into, and through the skin is seen to be more complicated. The exposure agent can be deposited directly onto the skin during work tasks, picked up by the skin when touching surfaces (settled airborne exposure), deposited from the airborne compartment directly on to the skin or onto surfaces, or deposited onto a clothing layer covering the skin (Schneider et al. 1999). From here, the contamination can be transferred between the compartments, ultimately reaching the skin surface where it may or may not penetrate depending on the chemical composition of the exposure and the integrity of the skin barrier. Much of the recent work focusing on dermal exposure and dermal exposure assessment has stemmed from the conceptual model published in 1999 by Schneider et al. (1999).

In this model, Schneider et al. acknowledge the interaction between the source of contamination, the air component and surface contamination. Additionally, the model acknowledges the complexity that exists due to the exchange of contaminant between the outer clothing layer, the inner clothing layer and the skin itself. All of these layers can receive contaminant from either the airborne component or the surface contaminant layer, and can also exchange contaminant between each other.

Previous research supports that self-reported skin symptoms are predictive of skin disease. However, some results suggest that self-reported skin symptoms may over estimate (Smit et al. 1992; Lynde et al. 2009) or underestimate (Holness et al. 1995) the prevalence of disease when compared with a physician examination. The use of picture based questionnaires and self-reported doctor-diagnosed dermatitis may provide a prevalence estimate closer to that of physician diagnoses, but may also underestimate prevalence (Smit et al. 1992).

It is also known that occupational airborne and skin exposures may be correlated (Fent et al. 2008; Liljelind et al. 2010). The skin exposure model proposed by Schneider et al. suggests that there is a compartmental connection between airborne and skin exposure based on contribution of one to the other (and vice versa) (Schneider et al. 1999). This connection means it is plausible that when the contribution of airborne exposure to skin exposure via deposition is high, the two exposures (airborne and skin) will be correlated. In addition to correlations between skin and airborne exposure levels, there is evidence that several common sensitizers are also associated with occupational asthma (Arrandale et al. 2012).

Individually, exposure-response relationships have been reported between occupational exposure and skin symptoms (de Joode et al. 2007; Sripaiboonkij et al. 2009a), though studies of exposure-response for respiratory symptoms are far more common. There has been little research on causal exposures or risk factors for reporting concurrent skin and respiratory symptoms, despite a number of case studies of workers with allergic occupational dermatitis and occupational asthma in response to the same occupational exposure (Moulin et al. 2009; De Raeve et al. 1998).

3 Conceptual Framework and Development of Hypothesis

Conceptual framework (Fig. 3) explains key constructs and terms, introduces and clarifies any theoretical models, situates the study within the prior theory and research, and identifies the phenomena the researcher propose to analyze and justification for studying it (Halstead 2010).

3.1 Defining Variables

Every experiment has at least two types of variables: independent and dependent. The independent variable (IV) is often thought of as our input variable. It is a variable that stands alone and isn't changed by the other variables you are trying to measure. The dependent variable (DV), or outcome variable is something that depends on other factors (independent variable). For the symptoms to manifest on the skin of the workers, the workers must be subjected to the risk factors through direct contact or inhalation. This means that the manifestations of OD on the skin of

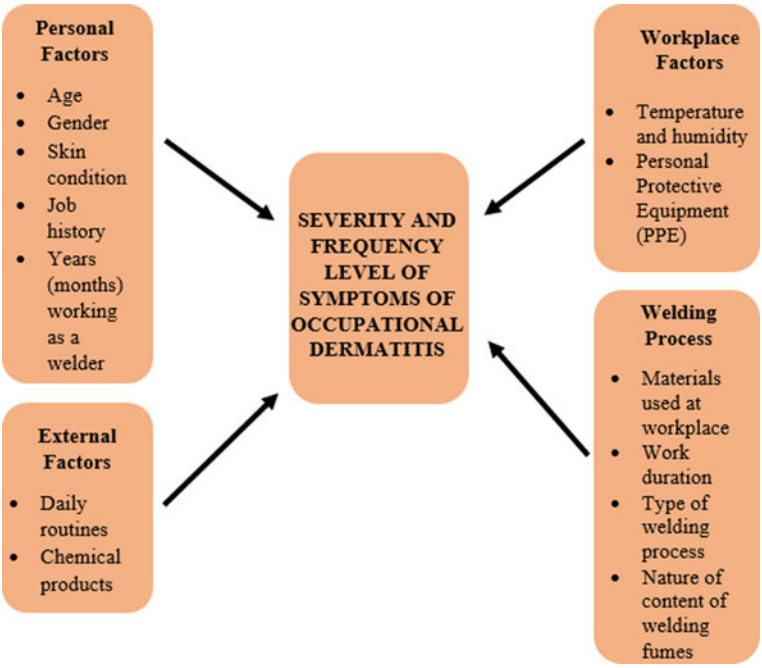


Fig. 3 Conceptual framework of the study

the welders depends on the risk factors the welders are exposed to. Therefore, the dependent variables for this study are the causative agents present in the workplace and the independent variables are the manifestations of Occupational Dermatitis.

4 Methodology

4.1 Research Design

The main aim of explanatory research is to identify any causal links between the factors or variables that pertain to the research problem. Hypothesis testing provides an understanding of the relationships that exist between variables which is applicable to the epidemiological studies. This method was used in order to obtain a clearer picture from the quantitative data, and then to use the qualitative data to provide better understanding and explanation of the study in question. This design first begins with the collection and analysis of quantitative data and followed by the collection and analysis of qualitative data. In the explanatory design, the researcher recognizes particular quantitative findings that need further explanation.

4.2 Measurement and Instrumentation of Variable

For this epidemiological study, the researchers employed to use Nordic Occupational Skin Questionnaire (NOSQ-2002) (Nordic Council of Ministers, 2002) EASI (Eczema Area and Severity Index) Score (Hon A/Prof Amanda Oakley, 2015) and POEM (Patient-Oriented Eczema Measure) (University of Nottingham, 2013).

The Nordic Occupational Skin Questionnaire (NOSQ) was developed by a group of Nordic occupational dermatology researchers to survey work-related skin dermatoses and exposures to environmental factors. The questionnaire covers demographics, occupational history, atopic symptoms, self-reported hand or forearm eczema, exacerbating factors, consequences and life impact of dermatoses, self-reported contact urticaria on hands or forearms, skin symptoms, skin tests, exposures, protective glove use, general health and household size.

EASI (Eczema Area and Severity Index) Score is a validated scoring system that grades the physical signs of dermatitis/eczema.

Severity score is recorded for each of the four regions of the body. The severity score is the sum of the intensity scores for four signs. The four signs are: redness, thickness (soreness, inflammation and weeping), splitting and cracking; and flaking. The average intensity of each sign in each body region is assessed as: none (0), mild (1), moderate (2) and severe (3).

4.3 Calculations

For each region, record the intensity for each of four signs and calculate the severity score by adding up the redness intensity, thickness intensity, splitting and cracking intensity, and flaking intensity.

For each region, multiply the severity score by the area score and by a multiplier. The multiplier is different for each body site.

- Head and neck: severity score x area score x 0.1 (in children 0–7 years, x 0.2)
- Trunk: severity score x area score x 0.3
- Upper limbs: severity score x area score x 0.2
- Lower limbs: severity score x area score x 0.4 (in children 0–7 years, x 0.3).

Add up the total scores for each region to determine the final score. The minimum score is 0 and the maximum score is 72.

POEM (Patient-Oriented Eczema Measure) is a tool used for monitoring eczema severity and frequency. It focuses on the illness as experienced by the patient. It is suitable for use in outpatient clinic, audit, epidemiological studies and clinical trials (University of Nottingham 2013). It is composed of seven questions and each of the seven questions carries equal weight and is scored from 0 to 4.

4.4 Epidemiological Measures

The frequency of occurrence of disease, injury and death often varies over time and between populations. Epidemiological principles and methods are used to describe the frequency and the determinants of these events. There are two main types of disease frequency—prevalence and incidence. Both measures relate to the number of cases in a population, and also include some concept of the time over which measurement is carried out.

Measures of frequency characterize the occurrence of health outcomes, disease, or death in a population. These measures are descriptive in nature and indicate how likely one is to develop a health outcome in a specified population.

Prevalence, or more correctly point prevalence, is the frequency of existing disease in a defined population at a particular point in time.

$$Prevalence = \frac{Number\ of\ cases\ with\ disease\ at\ a\ point\ in\ time}{Total\ population\ at\ same\ point\ in\ time}$$

Whereas prevalence is the frequency of existing cases of disease in a population, incidence is the frequency of new cases of disease in a defined population during a specified time period. There are different ways of measuring incidence.

Risk or cumulative incidence

The first measure of incidence is called cumulative incidence or risk, since it refers to the occurrence of risk events (disease, injury or death) in a group of people studied over time. It is calculated in much the same way as prevalence, but rather than existing cases, only new cases are counted over the specified time interval.

$$Cumulative\ Incidence = \frac{Number\ of\ new\ cases\ with\ disease\ at\ a\ point\ in\ time}{Total\ population\ at\ same\ point\ in\ time}$$

It is important to note that the denominator is the total number of people who were free of disease at the start of the time period. This is defined as the 'population at risk'. Also, like prevalence, this measure is a proportion and is dimensionless. However, because cumulative incidence will increase over time, the time period over which it is measured must be clearly stated.

Measures of exposure effect and impact

According to Bailey et al. (2006) to investigate a possible association between a risk factor and a particular disease, the incidence of disease in the people exposed to the risk factor should be compared with the incidence in a group of people who were not exposed. This comparison can be calculated by various methods, which are outlined below.

Relative measures estimate the size of an association between exposure and disease, and indicate how much more likely people in an exposed group are to develop the disease than those in an unexposed group. There are also three relative measures that can be used to calculate association between disease and exposure: risk ratio, rate ratio and odds ratio.

The risk ratio, also commonly referred to as relative risk, is calculated as the ratio between the cumulative incidence in the exposed group and the cumulative incidence in the unexposed group. Risk ratio equation is derived using 2 × 2 table:

Using the 2 × 2 table, the risk ratio can be calculated by:

Risk factor	Exposed unexposed total	Disease		
		Yes	No	Total
		a	b	$a + b$
		c	d	$c + d$
		$a + c$	$b + d$	$a + b + c + d$

$$Riskratio = \frac{\frac{a}{(a+b)}}{c(c+d)}$$

The risk ratio is used as a measure of aetiological strength (i.e. the strength of association between risk factor and outcome). A value of 1.0 will be obtained if the incidence of disease in the exposed and unexposed groups is identical, and therefore indicates that there is no observed association between the exposure and the disease, according to the given data. A value greater than 1.0 indicates a positive association or an increased risk among those exposed to the factor. A value less than 1.0 means that there is an inverse association a decreased risk among those exposed, or in other words, the exposure is protective.

The odds ratio is calculated in a similar way to the risk ratio and the rate ratio, in that the odds in the exposed group are compared with the odds in the unexposed group. Using the 2 × 2 table, the odds ratio can be calculated by:

$$Odds\ ratio = \frac{ad}{bc}$$

Odds ratios are usually used in studies where the incidence of the disease of interest is not known or if the study participants are selected on the basis of their disease status rather than because of their exposure status.

5 Results and Discussion

The relationship between welding and occupational dermatitis was studied using an explanatory study design. Work-related dermatitis was assessed in 183 welders involved in metallurgical industry and repair shops. A standardized dermatitis questionnaire (NOSQ-2002) was used and compared the outcomes to an internal reference group (74 personnel) consisting of those who work in the welding industry, but are not exposed or have only very low exposures to welding and/or other potential risk factors. The researchers also measured the severity and frequency of the occupational dermatitis both in welders and the internal reference group using Eczema Area and Severity Index (EASI) and Patient-Oriented Eczema Measure (POEM).

5.1 Demographic Characteristics

The survey respondents are composed primarily of 183 welders and 74 reference workers (including security guards, porters and clerical workers) working on different metal-related establishments around Bulacan, Pampanga and Quezon City.

By means of Nordic Occupational Skin Questionnaire (NOSQ-2002), the researcher assessed the respondents' skin condition. The results revealed that at most half (51%) of the respondents (47% welders and 4% from the reference).

The researcher determined the prevalence of occupational dermatitis in the welding industry. In welding industry, the prevalence of occupational dermatitis is 50.58%.

Whereas prevalence is the frequency of existing cases of disease in a population, incidence is the frequency of new cases of disease in a defined population during a specified time period. Cumulative incidence (CI) refers to the occurrence of risk events (disease, injury or death) in a group of people studied over time. The researcher computed the CI of the respondents. It can be observed a continual increasing in terms of number of workers who are disease-free at the start of the period but now have disease (occupational dermatitis).

To know what risk factors affect the prevalence of OD among the welders, the researcher used ordinal and binary logistics regression and Cochran-Mantel-Haenszel Method in risk and odds ratio. Risk ratio estimate the size of an association between exposure and disease, and indicate how much more likely people in an exposed group are to develop the disease than those in an unexposed group. Odds ratio, on the other hand, measures the possibility that the disease will develop among the unexposed group. The Cochran-Mantel-Haenszel method was a technique that generates an estimate of an association between an exposure and an outcome after adjusting for or taking into account confounding. The method was used with a dichotomous outcome variable and a dichotomous risk factor while a binomial logistic regression predicts the probability that an observation falls into

Table 1 Interpretation of risk ratio

Risk factor	*p*-value	Risk ratio	Odds ratio (%)
Age	0.612	0.7591	0.64
Skin condition:			
Cuts and skin abrasion	0.040	4.1077	12.83
Dry skin	0.000	3.4592	11.63
Itchy feeling when sweating	0.000	1.1194	5.94
Working hours	0.043	2.9093	6.64
Years of working	0.236	2.3157	4.13
PPE	0.033	2.3803	5.07
Location of workplace:			
Inside	0.048	6.3786	23.93
Outside	0.054	3.7710	7.38
Materials	0.018	4.8607	12.25
Welding process:			
GTAW	0.043	2.9365	4.01
SMAW	0.008	5.4070	13.87
GMAW	0.023	5.7286	16.59
FCAW	0.000	2.6265	3.39
Welding fumes	0.032	22.0299	62.42
Daily routines	0.010	2.6896	6.78
Chemical products	0.025	4.8769	10.53

one of two categories of a dichotomous dependent variable based on one or more independent variables that can be either continuous or categorical. Table 1 summarizes the p-value, risk ratio, odds ratio and interpretation of result for each risk factor.

Since the researcher was not allowed to observe inside the workplace due to concern in safety and hazards, the researcher used a subjective judgement scale for thermal condition. This implies that most of the workplace has hot to very hot thermal condition inside due to insufficient ventilation material and heavy protective clothing worn inside the workplace.

6 Conclusion

Skin can be protected at exposed workplaces if protective measures and programs are applied sufficiently and consistently. This is sometimes difficult.

The main aim of this study is to prevent the occurrence of OD on the basis of accurate risk assessment, in order to use and maintain appropriate control measures, information, instruction and training in the form of a program promoting skin protection.

In conclusion, the researcher's findings support the hypothesis that welding work is a risk factor for OD and that professional welding workers may not be sufficiently protected from dermal hazards at work.

7 Recommendation

7.1 Recommendation #1 Welding Fume Control

To lower the direct exposure of workers to welding fumes, some metal processing companies used welding fume extractor installation which is designed to ventilate large construction halls in compliance with the statutory provisions for welding fumes extraction.

Types of Welding Fume Extractor

- Source Capture
- Exhaust System Method
- General Air Cleaning Method
- Mechanical Filter Removal Equipment
- Electrostatic Precipitator
- Molecular Filtration.

7.2 Recommendation #2 Employee Health Service

Larger employers began providing on-site healthcare in the 1980s, usually as a way to treat occupational injuries. These employers were generally in the heavy industry or manufacturing trades and saw a significant number of on-the-job injuries that needed to be treated quickly similar to welding industry (Fig. 4).

Figure 5 shows a typical layout of an onsite clinic consisting on wash room, examination area and a single-bed rest area.

Advantages:

1. Lower costs—Employers have more negotiating power and therefore more control over the actual healthcare delivery costs at on-site clinics.
2. Increased preventive care - The convenience and lower costs of on-site or near-site clinics encourage employees to focus more on their preventive care, such as diagnostic screenings or flu shots. Through collaboration with the employee's independent primary care physician, on-site clinics can also help manage chronic conditions effectively. It can also help in earlier treatment of illnesses and injuries.

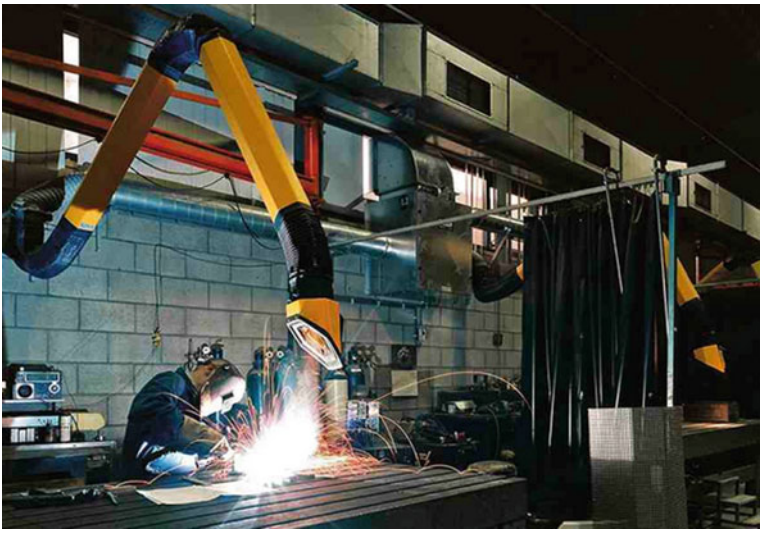


Fig. 4 Welding fume extractor

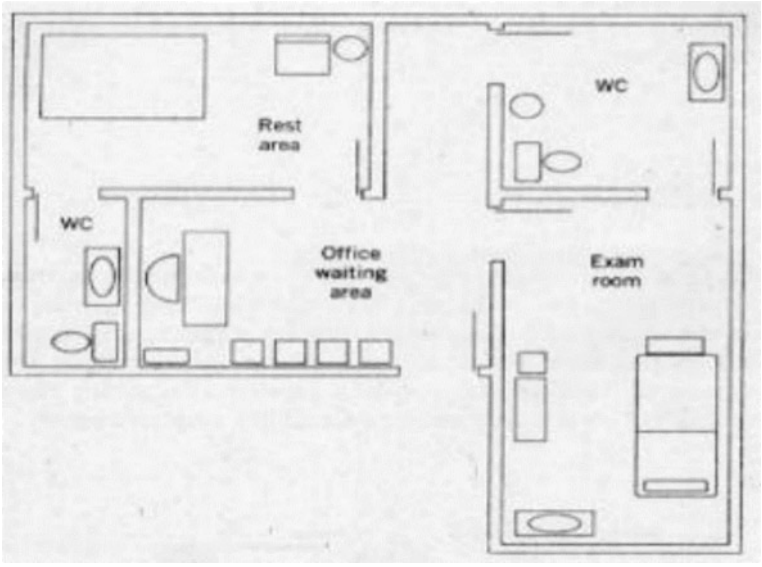


Fig. 5 Sample layout of an onsite clinic

3. Reduced absenteeism—When employees have better access to convenient, lower cost healthcare, they are less likely to develop illnesses that keep them away from work.
4. Recruiting benefit—Potential employees recognize the availability of an on-site or near-site clinic as an added benefit when considering whether to accept a position with a company.

7.3 Recommendation #3 Using Personal Protective Equipment

Using personal protective equipment at all times inside the workplace even though PPE devices alone should not be relied on to provide protection against hazards (based on the data gathered and related studies) but should be used in conjunction with guards, engineering controls, and sound manufacturing practices.

PPE description	Area of protection	Characteristics	Cost
Clothing	Upper and lower limbs	– Treated with non-durable flame-retardant materials – Made from heavyweight, tightly woven, 100% wool or cotton Woollen clothing is preferable to cotton because it is not so readily ignited and helps protect the welder from changes in temperature – Outer clothing can be jumpers or overalls – Sleeves and collars be kept buttoned to prevent from sparks to lodge in – Pockets be eliminated from the front of clothing – Fire-resistant leggings or other equivalent means should be used for heavy work	P 450/each
Gloves	Hands	– Made of durable flame-resistant cotton (light work) – Made of leather or other suitable durable flame-resistant (heavy work). Leather is a good electrical insulator – Insulated linings should be used to protect areas exposed to high radiant energy – Gauntlet-type cuff leather to protect wrists and forearms	P 300/pair
Apron	Lower limbs	– Made of leather or other suitable flame-resistant materials – Used for additional protection against sparks and radiant energy	P 300/each

(continued)

(continued)

PPE description	Area of protection	Characteristics	Cost
Boots	Feet	– Made from steel and capped – Must be high top boots fully laced to prevent sparks from entering into the boots	P 1,300/pair
Welding mask/ helmet	Face and neck	Made of: Helmet shell—must be opaque to light and resistant to impact, heat and electricity Outer cover plate made of polycarbonate plastic which protects from radiation, impact and scratches Filter lens made of glass containing a filler which reduces the amount of light passing through to the eyes. Filters are available in different shade numbers ranging from 2 to 14. The higher the number, the darker the filter and the less light passes through the lens Clear retainer lens made of plastic prevents any broken pieces of the filter lens from reaching the eye Gasket made of heat insulating material between the cover lens and the filter lens protects the lens from sudden heat changes which could cause it to break. In some models the heat insulation is provided by the frame mount instead of a separate gasket	P 500/each
Total cost: P 2850/welder			

7.4 Recommendation #4: Ventilation

Temperature is one of the major risk factor that affects the frequency and severity of the dermatitis. Well-designed and well-maintained ventilation systems remove toxic vapors, fumes, mists or airborne dusts from the workplace before workers are exposed.

7.5 Recommendation #5: Highlight the Law Subjected to Welding

The regulation entitled "Insist on Safe Working Conditions Before You Weld" should be strictly implemented in the company. It covers many aspect of welding works, including welding safety, welding in confined space, handling of compressed gases, fire and electrical safety, ventilation, protective equipment, and workers training (AFSCME 2011).

8 Area for Future Work References

Since this study focuses on subjective side of the epidemiology study, clinical test can be done to further verify the data gathered by the researchers. Larger population, welding companies and location of the workplace can also be considered. Personal factors like gender, smoking habits, and genetic background can be added to the long list of risk factor that can affect the prevalence of occupational dermatitis. Skin products used can also be added. These factors can help in further analysis of the study, arriving at new problems and developing new objectives and recommendations. This study can also be applied in a certain welding industry to check if the recommendations are feasible and viable.

References

1. Agency for Safety and Health at Work (2008) Retrieved 19 Nov 2010, from: http://osha.europa.eu/en/publications/reportsTE7007049ENC_skin_diseases
2. Kanerva L, Elsner P, Wahlberg JE, Maibach HI (eds) (2000) Handbook of occupational dermatology. Springer, Berlin, Heidelberg, New York, pp 1–1300
3. Kanerva L, Elsner P, Wahlberg JE, Maibach HI (2004) Condensed handbook of occupational dermatology. Springer, Berlin
4. Last JM (ed) (1995) A dictionary of epidemiology, 3d edn. Oxford University Press, New York, p 55
5. Leshem YA, Hajar T, Hanifin JM (2015) What the Eczema area and severity index score tells us about the severity of atopic dermatitis: an interpretability study. British J Dermatol 172 (5):1353–1357
6. Wilkinson S, Beck M (2004) Occupational dermatitis In: Burns T, Breathnach S, Cox N, Griffiths C (eds) Rook's textbook of dermatology

The Study of Optimal Illumination for Long Term Memory Activation: Preliminary Study

Chung Won Lee, Won Teak Kwak and Jin Ho Kim

Abstract This study is a preliminary study to verify the effect of long-term memory on light illumination. Experimental conditions of this study were 200, 400, 1000 lx. The color temperature was set to 5000 k in all three conditions. The participants consisted of four adults aged 24 years. The illumination was on the stand type engothh-8100 and the long-term memory was measured by the WFC task. The results showed that long-term memory was superior at 200 and 400 lx, which are relatively low illuminance than high illuminance of 1000 lx. In particular, long-term memory was found to be most active at 400 lx.

Keywords Long-term memory · LED lights · WFC task · Illumination

1 Introduction

The development of lighting technology is increasing the exposure of artificial light to people living in modern society. In particular, LED lighting has been reported to directly or indirectly affect human emotions [1, 2]. Therefore, LED lighting is often called emotional lighting. Emotional lighting can be defined as 'lighting technology that can change the space by applying the color temperature and brightness to the psychological state and the biorhythm of the person'.

In recent years, cognitive disorders such as dementia have become social problems, and interest in the effects of light on memory is increasing. In response to this trend, various studies on light and memory have been carried out. Most studies

C. W. Lee (✉) · W. T. Kwak · J. H. Kim
Department of Industrial and System Engineering, Kongju National University, 1223-24 Chenan-daero, Seobuk-gu, Cheonon-si, Chungcheongnam-do, Korea
e-mail: ChungWonLeekjh@kongju.ac.kr

W. T. Kwak
e-mail: WonTeakKwakkjh@kongju.ac.kr

J. H. Kim
e-mail: JinHoKimkjh@kongju.ac.kr

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_41

between light and memory have been conducted focusing on human activity, attention, cognitive performance, and working memory [3–9]. Among them, most of the studies related to memory focus only on working memory. However, memory is not composed of working memories.

The currently accepted memory model is the modal model, which is divided into sensory memory, short-term (working memory) memory, and long-term memory [10]. e sensory register receives the information that reaches the senses and moves to a short-term memory storage by paying attention to the information. The information in the storage of short-term memory is stored as a long-term memory through rehearsal. Sensory memory and short-term memory can be extinct according to the retention time because of the limit of capacity, whereas long-term memory is accepted as having no capacity limit [11, 12]. Recently, the concept of working memory is used more often by emphasizing the active aspects of short-term memory through various studies related to retrieval. The working memory is controlled by the central executive system, which is a phonological loop that mainly deals with auditory and linguistic information and a visuo-spatial sketchpad which mainly deals with visual data (Fig. 1).

In fact, long-term memory may be more important in the structure of memory. Because sensory memory and working memory quickly disappear. However, it is long-term memory to remember for a long time. But, few studies have examined the effect of light on long-term memory associated with long-term memory. The study of light and long-term memory performed by Jeong et al. (2017) is a representative study [13]. They verified the long-term memory through the WFC (Word Fragment completion) task at 400, 700, and 1000 lx for the light illumination, which was the best at 400 lx conditions. This result is in contradiction with the result that the cognition performance is excellent in the condition of relatively high light illumination. In addition, there is still some doubt about the results of the study that the long-term memory is excellent at relatively low illumination, because it has not been re-verified. Therefore, this study was carried out to re-examine whether long—term memory is effective under relatively dark conditions at 200, 400 and 1000 lx, and to search for optimal illumination conditions in long—term memory. In this paper, we will focus on the preliminary research stage for conducting the study.

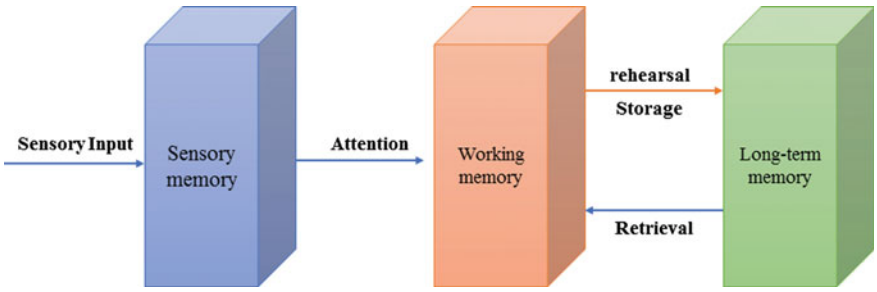


Fig. 1 Structure of memory

2 Experiment

2.1 Environment of Experiment

The laboratory environment was designed in the form of a learning space consisting of dark blocks, and the light was manipulated through LED lighting. The LED lighting used in this experiment is the engoth-8100 model and stands type. The height of the light source in this model is 42 cm, and the size of the light emitting portion is 3 cm in width and 38 cm in length. The color temperature was set to 5000 k in all three conditions. In order to satisfy the PMV condition recommended by the ASHRAE Standard, the temperature was maintained at 24 °C ±3 and the humidity was 50% ±10. In addition, prevent the noise which may affect the experiment, we installed a soundproofing system to completely block the noise (Fig. 2).

2.2 Experimental Participant

Preliminary studies were conducted on four adults without impairment of cognitive function. The average age of the participants was about 24 years old, and they instructed them to avoid food that affects cognitive function such as alcohol or caffeine drinks before the experiment day. In addition, the experiment was not conducted for the participants who felt sleepiness was insufficient considering that sleeping time would affect the results of the study.

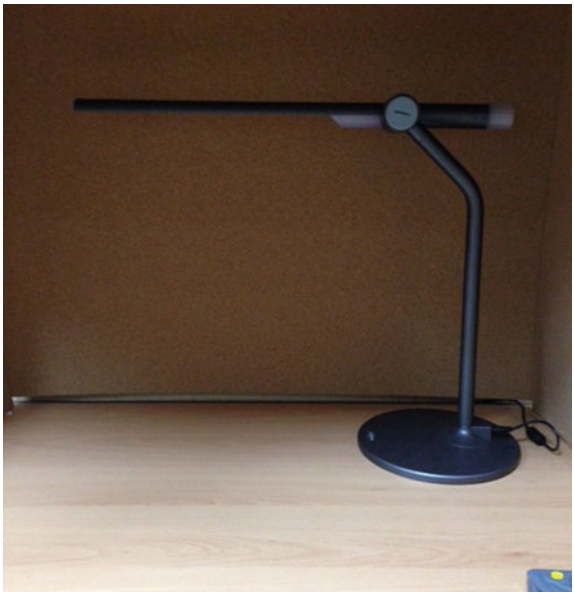


Fig. 2 LED stand lighting (engoth-8100)

2.3 Word Fragment Completion Task

Word-fragment completion task was used for long-term memory measurement. The Word-fragment completion task is a task to complete a word by filling in a missing letter based on the learned information and is one of the tools to indirectly measure long-term memory [14–18]. In this study, we used the meaningless spelling consisting of only 4 consonants. Long-term memory was measured by filling one digit. In addition, three types of forms with similar difficulty levels were produced and used randomly to exclude the effects of the tasks as much as possible.

2.4 Experimental Procedure

Experiments were assigned to three illuminance conditions (200, 400, 1000 lx) and the order was done randomly. In addition, three types of tasks were produced in accordance with the three conditions, and tasks were also assigned randomly to eliminate the effects of the tasks. Experiment participants were exposed to 2 min of dark adaptation and 2 min of light adaptation. Subsequently, participants were instructed to study 10 meaningless English words. Long-term memory is influenced by the number of rehearsal. Therefore, in order to equalize the number of rehearsal, participants were given only one word on A4 paper and read 10 times loudly. The speed of reading was controlled by reading 10 words over 5 min through preliminary practice. After completion of the learning task, the participant performed a five-minute interference task to prevent additional rehearsal. The interference task consists of a task of performing a simple subtraction operation. After the interference task, participants performed the WFC task. The WFC task consisted of 8 items excluding the first and last items in order to block the position effect (Fig. 3).

3 Result

3.1 Result of Analysis

The results of the preliminary study were analyzed based on the percentage of correct answers. The results are shown in Table 1.

The long-term memory of each participant is shown in Fig. 4.

The results showed that long term memory was the best at 400 lx (81.25%) and lowest at 1000 lx (46.88%). These results, although preliminary, seem to be somewhat consistent with the results of previous light and long-term memory study. In addition, 200 lx (71.88%) conditions were not better than 400 lx overall, but participant 1 (P1) was found to be more effective at 200 lx than 400 lx. The results of this study will be clearer if the number of participants increases.

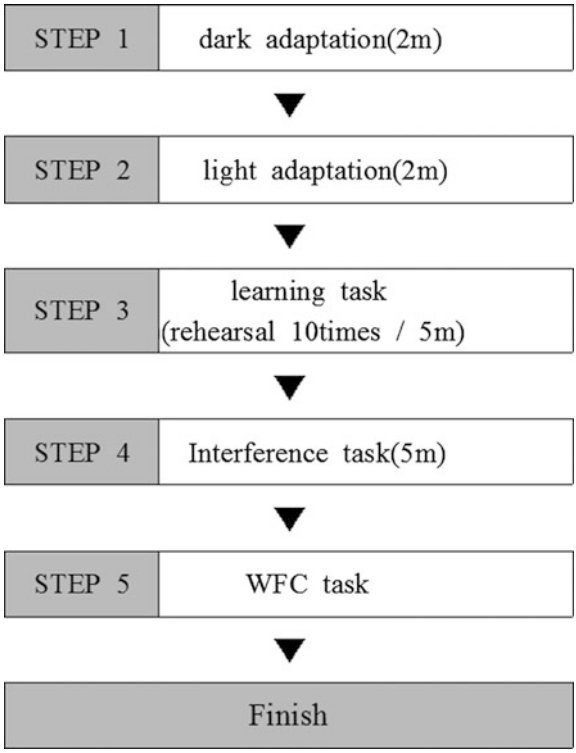


Fig. 3 Experimental procedure

Table 1 Mean and standard deviation of long-term memory

	Mean	sd	N
200 lx	71.88	23.66	4
400 lx	81.25	16.14	4
1000 lx	46.88	11.97	4

3.2 Retention of Memory After a Long Time

Whether or not this study measures actual long-term memory can be judged from the amount of memory retention after a long period of time. After 3 days, the retention of memory was re-measured. As a result, most of the participants were holding more than 50% of the memories, and one participant was holding 100% of the memories. As a result, most participants were retention more than 50% of the memories, and one participant retention of 100% of the memories. These results indirectly demonstrate that this study measures long-term memory.

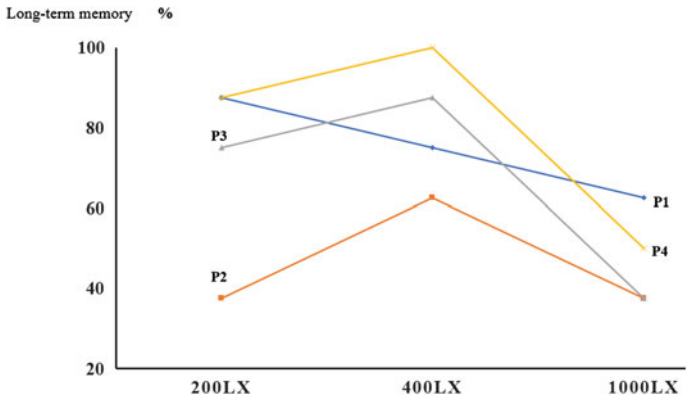


Fig. 4 Percentage of long-term memory per participant

4 Conclusion and Discussion

This study was conducted to systematically re-examine the effects of light illumination on long-term memory and to search for optimum illumination that activates long-term memory. The results of this study show that the 200 and 400 lx conditions have higher activation of long-term memory than the relatively high illuminance of 1000 lx. This result is in agreement with the results of Jung et al. (2017). In addition, the 400 lx condition was more effective than the 200 lx condition for long-term memory. However, the results of this study still have limitations in reaching definite conclusions due to the nature of the preliminary study. I think we will get more definite results when we draw the results through more experiment participants.

References

1. Enguo C, Tailiang G (2014) Modified Köhler illumination for LED-based projection display. Displays 35(2):84–89
2. Christian C (2007) Alerting effects of light. Sleep Med Rev 11(6):453–464
3. Smolders KCHJ, de Kort YAW, Cluitmans PJM (2012) A higher illuminance induces alertness even during office hours: findings on subjective measures, task performance and heart rate measures. Physiol Behav 107(1):7–16. https://doi.org/10.1016/j.physbeh.2012.04.028
4. Gilles V, Evelyne B, Christophe P, Christian D, Vincent M, Virginie S, Geneviève A, Annabelle D, Martin D, Thien TD, Philippe P, Andre L, Derk JD, Pierre M (2006) Daytime light exposure dynamically enhances brain responses. Curr Biol 16(16):1616–1621. https://doi.org/10.1016/j.cub.2006.06.031
5. Karin CHJS, Yvonne AWK (2014) Bright light and mental fatigue: effects on alertness, vitality, performance and physiological arousal. J Environ Psychol 39:77–91. https://doi.org/10.1016/j.jenvp.2013.12.010

6. Scott SC, Drew D (1990) Enhancement of nighttime alertness and performance with bright ambient light. Physiol Behav 48(2):317–320. https://doi.org/10.1016/0031-9384(90)90320-4
7. Kretschmer V, Schmidt K, Griefahn B (2011) Bright light effects on working memory, sustained attention and concentration of elderly night shift workers. Lighting Res Technol 44 (3):316–333
8. Boyce PR, Beckstead JW, Eklund NH, Strobel RW, Rea MS (1997) Lighting the graveyard shift: the influence of daylight-simulating skylight on the task performance and mood on night-shift workers. Lighting Res Technol 29(3):105–134
9. Huibert LM, Smolders KCHJ, deKort YAW (2015) Shining light on memory: effects of bright light on working memory performance. Behav Brain Res 294:234–245. https://doi.org/10.1016/j.bbr.2015.07.045
10. Atkinson RC, Shiffrin RM (1968) Human memory: a proposed system and its control processes. In: Spence KW, Spence JT (eds) The psychology of learning and motivation, vol. 2. Academic Press, New York. https://doi.org/10.1016/S0079-7421(08)60422-3
11. Premjit KS (2013) Attention and intended action in multitasking: an understanding of cognitive workload. Displays 34(4):283–291. https://doi.org/10.1016/j.displa.2013.09.001
12. Peterson L, Peterson MJ (1959) Short-term retention of individual verbal items. J Exp Psychol 58:193–198. https://doi.org/10.1037/h0049234
13. HeeChang J, JinHo K, Chungwon L (2017) The effect of the illuminance of light emitting diode (LED) lamps on long-term memory. Displays 49:1–5. https://doi.org/10.1016/j.displa.2017.05.001
14. Tulving E, Schacter DL, Stark HA (1982) Priming effects in word-fragment completion are independent of recognition memory. J Exp Psychol Learn Mem Cogn 8(4):336–242. http://dx.doi.org/10.1037/0278-7393.8.4.336
15. Nelson DL, Canas JJ, Bajo MT, Keelean PD (1987) Comparing word fragment completion and cued recall with letter cues. J Exp Psychol Learn Mem Cogn 13(4):542–552. https://doi.org/10.1037/0278-7393.13.4.542
16. Challis Bradford H, Brodbeck David R (1992) Level of processing affects priming in word fragment completion. J Exp Psychol Learn Mem Cogn 18(3):595–607
17. MacLeod CM, Kampe KE (1996) Word frequency effects on recall, recognition, and word fragment completion tests. J Exp Psychol Learn Mem Cogn 22(1):132–142. https://doi.org/10.1037/0278-7393.22.1.132
18. Maria JS, Juan CR, Martin V, Carmen D, Inma F (2011) Perceptual priming in schizophrenia evaluated by word fragment and word stem completion. Psychiatry Res 190:167–171. https://doi.org/10.1016/j.psychres.2011.08.008

Effect of the Shock Absorber of the Shock Absorption Benefits for Upper Arm

Yiyang Chen, Wenhsin Chiu and Yuting Chen

Abstract The aim of this study was to explore the effect of handlebars shock absorbers on the human upper body muscles. We recruited ten male adult subjects to join our experiment. Subjects need to respectively rode the bike with shock absorber and without the shock absorber on the two kinds of vibration conditions (40 and 50 Hz). The vibration generated by tire contact with the ground while riding was simulated through the self-made vibration simulation machine. The muscle activities of ulnaris, radialis, biceps brachii, triceps brachii of participant were collected by BTS Bioengineering® Wireless EMG meter. Descriptive statistics and one-way ANOVA were adopted to analyze the data. The results showed only ulnaris muscle has reached statistically significant level ($p < 0.05$) under low-frequency shock conditions, the ulnaris muscle, and triceps brachii muscles has reached statistically significant level ($p < 0.05$) under high-frequency shock conditions.

Keywords Bike · Shock absorbers · Muscle activity · Upper arm

Y. Chen (✉)
Department of Physical Education, National Taiwan Normal University, Taipei, Taiwan
e-mail: hc7022709@gmail.com

W. Chiu
Department of Physical Education, National Tsing Hua University, Hsinchu, Taiwan
e-mail: whchiu@mx.nthu.edu.tw

Y. Chen
Department of Industrial Management, Chung Hua University, Hsinchu, Taiwan
e-mail: p2f3020@gmail.com

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_42

1 Introduction

When riding the bike, the vibration will through the tire to the handlebar, pedal and chair then pass to the human body, the body need withstand a lot of vibration. Because the shock of handlebar more largest the upper arms were more uncomfortable and easily cause damage, the common injury is the cyclist's palsy or ulnar neuropathy. Therefore, for the long time riding, it is necessary to install the good shock absorber to prevent the fatigue and injury of the upper arms [1].

Based on the literature reviewed, riding the bicycle which install the fork shock absorber can increase the efficiency and reduces the sore of upper arms, when the vibration was larger, the shock absorber must have sufficient damping factor and the buffer space (riser length) to achieve the effect of shock absorption [2]. Shock absorber is a kind of machinery, the main function is absorbed the vibration caused by the bumpy road and minimize the vibrations as much as possible before the impact reaches rider. A good shock absorber can help riders to maintain good handling. It can also help the bicycle have good handling and absorb more impact and increase the grip of the tire when riding on the bumpy road [2, 3].

The vibrations caused by tire contact with the ground may increase the chance of the hand injury. As a result, riders need to choose the better shock absorber. However, the development of sports equipment must be continually tested to realize its effect, so the aim of this study was investigate the effect of newly designed of faucet shock absorber and understand about the shock absorption benefits of this product for the human body.

2 Research Approach

2.1 Participants

Ten male adults who without upper arm disease history in a recent year and have more than one-year bike riding experience and the used hand is the right hand was be recruited as a subject. Participants signed an informed consent form prior to join the experiment. The basic data of the subjects are shown in Table 1.

Table 1 Subjects basic information (n = 10)

	Age (year)	Height (cm)	Weight (kg)	90% leg length (cm)
Mean	25	174	69	82
Std.	3	6	12	1

2.2 Experimental Control

This study focuses on the measurement of muscle activation on the used upper arm, but, in order to measure the effect of shock absorber cushioning accurately and the simulated the actual situation of riding a bicycle. Therefor, need to control body pressure, riding rhythms and rider's seat height. Solid tire was used in this experiment, riding speed was set at 50 rpm (revolutions per minute), 50 rpm is the rhythm of the general public stampede, also usually used in physical experiments [4] and the height of seat was on the basis of every participants' 95% leg length [5]. Subjects need to naturally place their hand on the handlebar, cannot standing when biking and keep their body as stable as possible [6]. We installed the metronome on the bike to fixed the riding rhythms of subjects in the experiment.

2.3 Apparatus

The shock absorber used in this experiment is pneumatic shock absorbers which volume is smaller than general shock absorbers. Height is 10 cm, weight is 300 g and caliber is 25.4 mm, it suits any bike which caliber is 28.6 mm.

The bike was set on the self-made vibration simulation machine (see Fig. 1) and the shock absorber installed on the handle Stem. The BTS Bioengineering® Wireless EMG meter (see Fig. 2) was placed on the table.

2.4 Experimental Procedure

There were two factors, vibration conditions and shock absorbers pattern. The vibration conditions included low-frequency shock (40 Hz) and high-frequency shock (50 Hz)

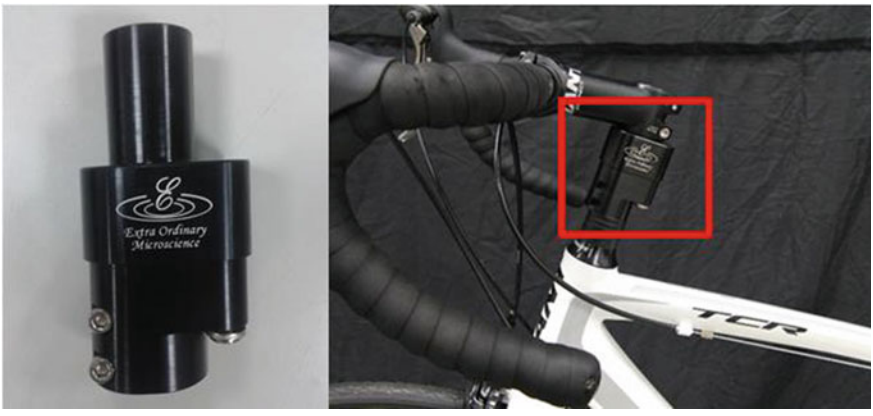


Fig. 1 Shock absorber and installed location

Fig. 2 BTS Bioengineering® Wireless EMG meter

conditions. The shock absorbers pattern included with shock absorbers and without shock absorbers. There were a total of 4 experimental conditions. Each subjects do 10-min warm-up before the experiment, remove the corneum and hair of skin and stick the EMG receiver on their body, then measure the maximal voluntary isometric contraction (MVC) [6].

Participants have ten seconds to achieve a steady tempo (50 rpm), and they were riding the bicycle respectively which with shock absorbers and without shock absorbers under two vibration conditions through the self-made vibration simulation machine (see Fig. 3), and recorded 30 s of muscle activation parameters. The BTS Bioengineering® Wireless EMG meter collected the muscles activity of ulnaris, radialis, biceps brachii, triceps brachii of the participant. Figure 4 shows the layout of this experiment.

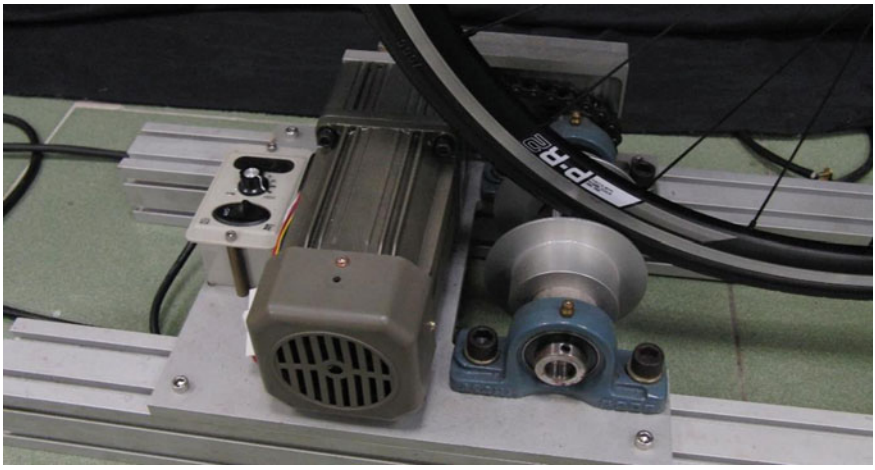


Fig. 3 Vibration simulation machine

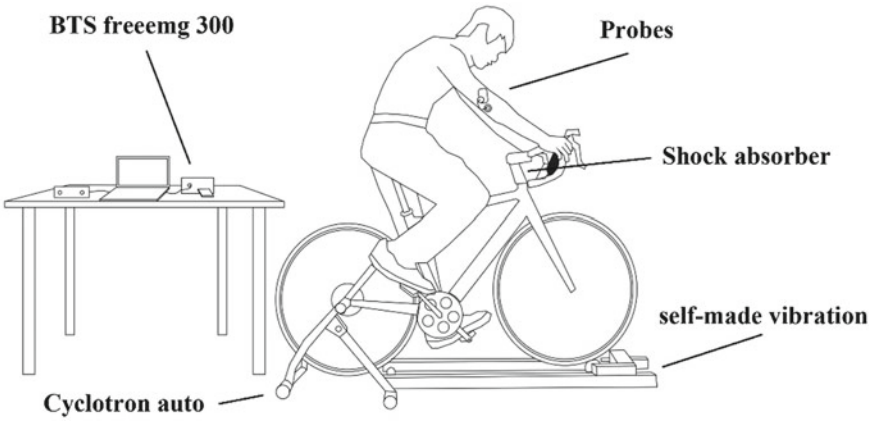


Fig. 4 Experimental layout

2.5 Data Analysis

The raw EMG data collected was used band pass filter processing of 20–500 Hz, next do the full-wave rectification process, finally, a low pass filter with a cut-off frequency of 10 Hz was used for getting EMG signal linear envelope. Further, use the root mean square (RMS) calculation procedures to get root mean square-EMG (RMS-EMG). RMS-EMG divided by time and then Standardize the RMS-EMG of MVC to available four muscle activation RMS value of myoelectric (%). The data collected were analyzed using descriptive statistics and one-way ANOVA. Microsoft® Excel 2010 and SPSS® 23 were adopted for data processing and statistical analyses, the significant level set at $p < 0.5$.

3 Results

Figure 5 shows the MVC under low-frequency shock condition(40 Hz). Both the MVC of ulnaris and radialis with shock absorber are lower than without shock absorber and the MVC of biceps brachii and triceps brachii with shock absorber are higher than without shock absorber. The ANOVA results of MVC were shown in Table 2. The muscle activity of ulnaris have reached statistically significant level ($p < 0.05$).

Table 3 and Fig. 5 show the MVC under high-frequency shock condition (50 Hz), Both the MVC of ulnaris and radialis with shock absorber are lower than without shock absorber. But the MVC of biceps brachii and triceps brachii with shock absorber are higher than without shock absorber. The two-way ANOVA results of the MVC were shown in Table 2. Both the muscle activity of ulnaris and triceps brachii have reached statistically significant level ($p < 0.05$) (Fig. 6).

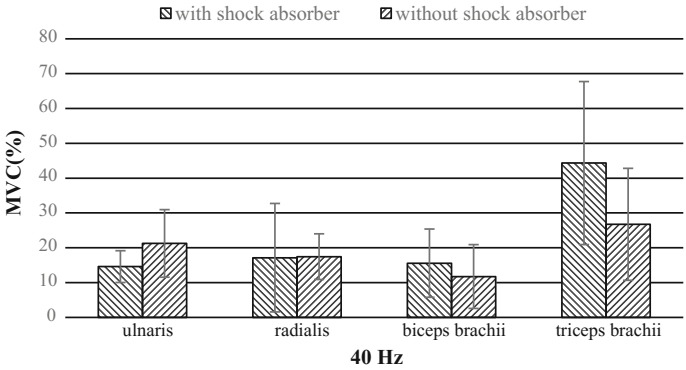


Fig. 5 The MVC under low-frequency shock condition

Table 2 Muscle activity in the low-frequency shock condition (n = 10)

	With	Without	*F*	*p*
Ulnaris	14.6 ± 4.5	21.2 ± 9.7	5	0.034*
Radialis	17.1 ± 15.5	17.4 ± 6.5	0.004	0.948
Biceps brachii	15.6 ± 9.8	11.7 ± 9.1	1.071	0.314
Triceps brachii	44.3 ± 23.4	26.7 ± 16	1.317	0.262

*$p < 0.05$

Table 3 Muscle activity in the high-frequency shock condition (n = 10)

	With	Without	*F*	*p*
Ulnaris	17.5 ± 3	23.3 ± 4.8	5	0.001*
Radialis	18.2 ± 13.2	22.7 ± 17.9	0.004	0.471
Biceps brachii	14.8 ± 13.3	12.1 ± 11.9	1.071	0.517
Triceps brachii	45.3 ± 22.2	25.3 ± 24.3	1.317	0.037*

*$p < 0.05$

According to those experiment results, we can find that muscle activity of ulnaris with the shock absorber in two vibration condition have obviously decreased, but the biceps brachii and triceps brachii have no reduced. Besides, the muscle activity of triceps brachii was higher than without the shock absorber. Therefore, the effects of this shock absorber can not cover whole upper limb muscle.

It is speculated that the larger of vibration, the shock absorber must have sufficient damping factor and the buffer space (riser length) to achieve the effect of shock absorption [2]. If the volume of shock absorber was contraction may cause the limited of effect. In addition, we can find that compare with not install the shock absorber, the muscle activity of triceps brachii in the dominant hand was greater with install the shock absorber. It is speculated that the handlebar fluctuates greatly

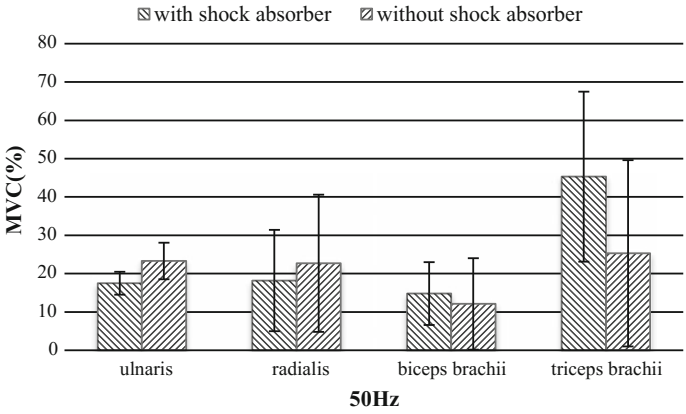


Fig. 6 The MVC under high-frequency shock condition

due to the buffering effect of the shock absorber when the shock absorber is installed, As the degree of shaking of the handlebar increases, the triceps muscle needs more power to keep the body stable.

4 Discussion and Conclusion

Some studies have shown that the forearm can feel the vibration directly during riding the bike and it is easy to increase the muscle activation. When riding with flat handlebars, the muscle activity of extensor is higher than flexors [7]. Installed the shock absorber can reduce hand fatigue, increase ride comfort and stability [8]. This result is consistent with our study.

Another research considered that the weight of the bicycle which installed the shock absorber will increase and need to spend more power, it will affect the speed and performance of rider. Besides, when the shock absorber is buffering, it may cause the amplitude of vibration increased and also increases the muscle activity [9]. with the increase in the shaking degree of the handlebar, the triceps will need more strength to keep the stable of the body, controlled the direction, delivered force and shared the pressure in the same time, may cause more muscle activity [10]. This result is consistent with our study.

The muscle activity of triceps brachii has increase may because of the main muscle group will do the self-contraction to achieve the buffering when vibration occurs, that human body will not be exposed to the dangerous or harsh environment [6].

As a Whole, the design of shock absorber must retain the vibration buffer space. If the size is reduced, the vibration buffer space will be reduced too. When the degree of bumpy road increases, it is bound to affect the effectiveness of shock absorbers. Therefore, it is recommended in the design of shock absorber must retain sufficient buffer space to increase the effectiveness of shock absorption.

References

1. Akuthota V, Plastaras C, Lindberg K, Tobey J, Press J, Garvan C (2005) The effect of long-distance bicycling on ulnar and median nerves: an electrophysiologic evaluation of cyclist palsy. Am J Sports Med 33(8):1224–1230
2. Levy M, Smith GA (2005) Effectiveness of vibration damping with bicycle suspension systems. Sports Eng 8:99–106
3. Raddy S, Thontaraj Urs TS (2014) Comparative study of static structural analysis of a shock absorber for different material. Int J Eng Sci Innovative Technol 3(6):632–641
4. American College of Sports Medicine (2006) ACSM's guidelines for exercise testing and prescription, 7th edn. Lippincott Williams & Wilkins, Baltimore, MD
5. Nordeen-Snyder KS (1977) The effect of bicycle seat height variation upon oxygen consumption and lower limb kinematics. Med Sci Sports Exerc 9(2):113–117
6. Arpinar-Avsar P, Birlik G, Sezgin ÖC, Soylu AR (2013) The effects of surface-induced loads on forearm muscle activity during steering a bicycle. J Sports Sci Med 12(3):512–520
7. Arpinar-Avsar P, Birlik G, Sezgin ÖC, Soylu AR (2013) The effects of surface-induced loads on forearm muscle activity during steering a bicycle. J Sports Sci Med 12(3):512–520
8. Raddy S, Thontaraj Urs TS (2014) Comparative study of static structural analysis of a shock absorber for different material. Int J Eng Sci Innovative Technol 3(6):632–641
9. Wang EL, Hull ML (1996) A model for determining rider induced energy losses in bicycle suspension systems. Veh Syst Dyn 25:223–246
10. Duc S, Bertucci W, Pernin JN, Grappe F (2008) Muscular activity during uphill cycling: effect of slope, posture, hand grip position and constrained bicycle lateral sways. J Electromyogr Kinesiol 18(1):116–127

Design and Development of Maternity Pillow: An Ergonomic Approach

Klea Charmaine Quintero, Viah Mae Condez, Siena Marie Plaza, Mark Joseph Bedolido and Philip Ermita

Abstract In today's generation, there are a lot of new inventions and innovations of the existing products with the use of technology and soon it will be all automated. Different studies were conducted in order to identify things that hold back one thing to explore. With this, customization of different products was important to address all the pains and discomforts and for them to be more productive in life. In different variances of pillows, maternity pillow is one thing the people need to focus on. Pillows can further eliminate the pains and discomforts. Musculoskeletal is the primary discomforts of pregnant women while sleeping. This paper aimed to identify the pains and discomforts with respect to factors such as the age, periods of pregnancy and number of birth of pregnant women. Pearson's Correlation, Quadruple Visual Analogue Scale (QVAS) and weighted mean were used in making the design and development of the maternity pillow.

Keywords Pregnant · Discomforts · Maternity pillow · Quadruple Visual Analogue Scale

K. C. Quintero (✉) · V. M. Condez · S. M. Plaza · M. J. Bedolido · P. Ermita
Department of Industrial Engineering, University of Perpetual Help System DALTA
Calamba, Calamba, Laguna, Philippines
e-mail: klea.quintero@gmail.com

V. M. Condez
e-mail: v.m.condez@gmail.com

S. M. Plaza
e-mail: sienaplaza22@gmail.com

M. J. Bedolido
e-mail: markjosephbedolido@gmail.com

P. Ermita
e-mail: philip.ermita@perpetualdalta.edu.ph

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_43

1 Introduction

Pregnancy is a time of great joy, excitement, anticipation and one of the most important events in most women's life [1]. But there are also problems that most women encounter while in the process of pregnancy. One of the most worldwide problems experience by a pregnant woman is getting enough sleep, uninterrupted sound sleep. In fact, according to the National Sleep Foundation's Women and Sleep poll, 79% of women report more disturbed sleep during pregnancy than at other times [2]. In an article, they said that they had conducted a survey, wherein 8 out of 10 women had suffered from sleep problems during their pregnancy [3]. And conforming from the pregnant women, getting to sleep is very hard when they are not comfortable and experiencing sleep disorders. Sleep is one of the three pillars of health that a pregnant woman needs to have. Good sleep is an essential component of health and wellbeing [4]. It consumes one-third of human existence. The regular sleeping position of a woman may no longer work during pregnancy.

Sleep disorders, frequently prevalence for pregnant, such as nocturnal awakenings, snoring, low back pain, restless leg syndrome, leg cramps, abdominal and pelvic pain [5]. Hormonal and physical changes affect the sleep quality and patterns of pregnancy that may cause sleep disorders. The regular pillow has no help for those women who experience discomforts such as low back, pelvic pain, insomnia, and restless leg syndrome or RLS [6].

Maternity pillow is a product that can enhance sleep quality of pregnant women [7]. There was a study on how they came up with the design of the pillow for a pregnant woman. They used Kansei words and analyzed the optimum design with Taguchi method. Uncomfortable maternity pillow, being too soft or too hard, too thick or very thin, can affect the sleep of the pregnant women; it can cause stress and frequent awakening when sleeping [7].

However, in this research, an Ergonomic Maternity Pillow was developed in order to lessen the pains/discomforts of the pregnant women while sleeping. It aimed to determine the demographic profile and physical discomforts of the pregnant women using the current type of their pillow and if they are experiencing pain from different parts of their body due to the changes every semester, ascertained the significant relationship of the pains with the demographic profile of the pregnant women, measured and analyzed the anthropometric data of the pregnant women, and selected the most comfortable and effective materials, design and thickness, developed an ergonomically designed maternity pillow, and determined the effectiveness of the design of the maternity pillow that will benefit the pregnant women while they are sleeping. Hence, this paper specifically intends to determine the factors that mainly affect the pregnant woman's convenience during sleeping and to have a wonderful morning awakening.

2 Research Methodology

2.1 Interview, Survey and Questionnaire

In the direction of obtaining the required number of sample size for the survey, the researchers of this study gathered the latest population of pregnant women of the two barangays which were Barangay Hornalan and Baranggay Laguerta, respectively. Then, the researchers determined who among them were in the second to third trimester of their pregnancy and eventually became the respondents of the study. The primary data were gathered through interviews and surveys of the researchers directly on pregnant mothers on selected household and health centers in the selected Barangays. The duration of gathering data was from August 2017 to February 2018 both for the survey and validation. Mainly, the purpose of the survey questionnaire was necessary to develop needed measurements, dimension, and design in the fulfillment of the design and development of maternity pillow from the gathered data.

Also, another way of gathering data was by asking the pregnant woman to answer the questionnaires that were used for evaluating the pain that they usually experience during pregnancy was Quadruple Visual Analogue Scale (QVAS). These have 4-time guides that were labeled the Von Korff rating scale: (1) pain level at the time of the current visit; (2) typical or average pain since the last visit; (3) pain level at its best since the last visit; and (4) pain level at its worst since the last visit. The scores from the rating scale number 1, 2, and 4 are averaged and then multiplied by 10 to yield a score from zero to 100. The final score is then categorized as "low-intensity" (pain < 50) or "high-intensity" (pain >= 50) [8].

Moreover, the level of satisfaction scale was the basis of the magnitude to confirm the satisfaction of the pregnant women with the design of the maternity pillow. Ratings were divided into 4 levels. Level 1 means that the pregnant woman was unsatisfied while level 4 means that they were highly satisfied with the maternity pillow.

2.2 Analysis with Statistical Tool

In order to achieve the effectiveness of the designed maternity pillow, a validation was held at the same place and person where the data on the sizes of the mothers were gathered. This strategy was used to integrate the design of the maternity pillow. Hence, the application of statistical tool was indeed analyzed in this study. Those are Pearson's Correlation which is commonly used for testing relationships between categorical variables, weighted mean, and Standard Deviation.

3 A Proposed Maternity Pillow in an Ergonomic Approach

3.1 Generating Design of Maternity Pillow

The data towards determining the most common physical discomforts or pains that the pregnant women from Barangay Hornalan and Laguerta mostly experiencing. In the survey questionnaire, the possible parts of the body that might suffer from pains while sleeping were indicated. The pregnant women would have to choose from it, wherein they can choose more than one of the body parts. And they can specify if they have any pains or discomforts from other parts of the body.

Table 1 shows the parts of the body that feel pain during their pregnancy. Based on the data gathered, it shows that the pregnant women primary feel pain in their lower back, followed by their upper back, then neck and pelvic area, next is a shoulder, then the legs and hips lastly, the thigh.

Quadruple Visual Analogue Scale (QVAS) is a more reliable and valid method for pain measurement. Consequently, the researchers used this as their instrument for the body parts to assess the concentration of the pains [8].

Based on Table 2, the result of QVAS Questionnaire Analysis, most of the body parts that experienced pain are the Lower Back, Upper Back, Neck, Pelvic Area,

Table 1 Body parts of the pregnant woman that experienced pain

Body parts	Frequency	Percentage (%)
Upper back	23	19
Neck	20	17
Shoulder	18	15
Arms	0	0
Lower back	25	21
Hips	12	10
Pelvic area	20	17
Abdomen	0	0
Thigh	9	7
Legs	17	14
Total	121	100

Table 2 Level of pains experienced by the pregnant women

Rank	Result	Body parts
1	High intensity pain	Lower back
2	High intensity pain	Upper back
3	High intensity pain	Pelvic area, neck
4	High intensity pain	Shoulder
5	High intensity pain	Legs

Shoulder, and Legs are on the high level of pain. Those body parts which were on the highest level of pain became the focus of creating the maternity pillow, but the body parts that got the no intensity pain would not be given attention to the support of the maternity pillow. The lower Back pain got the first rank, which means lower back pain is the most body part that experienced a high intensity of pain.

Table 3 shows the results that on their period of pregnancy, there is a moderate to substantial correlation to the pain that the pregnant women are experiencing, the other two on the demographic profile which is the age and number of birth mostly has a negligible correlation.

Table 4 shows the preferences of the pregnant women with regard to the materials, design, and thickness. Through the thickness of the pillow was changing depending on their period or the intensity of pain. Thus, the maternity pillow should come up with different features on the maternity pillow. First is the adjustable thickness of the maternity pillow to possibly adjust the height of the pillow by the means of blowing or air pumping. And the second one, detachable parts of the maternity pillow wherein each part is connected with magic tape that will serve as the binder of the various parts of the pillow. So, the pregnant women can remove a part of the pillow if it is not needed.

In addition, the materials and design were included in the survey questionnaire. For the aesthetics, the blue color was all about tranquility, calm, depth of feeling, harmony and it attracts loner rest; it is also good for relaxation. For the stuffing of the pillow, latex foam was one of the most used today which is also similar to

Table 3 Pearson's correlation of demographic profile with body pains

Body pains	Age	Period of pregnancy	Pregnancy order
Upper back pain	Negligible correlation	Moderate correlation	Negligible correlation
Neck pain	No correlation	Substantial correlation	Low correlation
Shoulder pain	Negligible correlation	Moderate correlation	Negligible correlation
Lower back pain	Negligible correlation	Moderate correlation	Negligible correlation
Hips pain	Negligible correlation	Low correlation	Negligible correlation
Pelvic pain	Negligible correlation	Substantial correlation	Negligible correlation
Thigh pain	Negligible correlation	Negligible correlation	Low correlation
Legs pain	Negligible correlation	Substantial correlation	Negligible correlation

Table 4 Elements of the maternity pillow

Design elements		Percentage (%)
Features	Adjustable thickness	70
	Detachable parts	63
Materials	Latex foam	43
	Polyester fiber	33
	Cotton fabric	100
Design	Blue color	30
	Plain design	77

Table 5 Design parameter for maternity pillow

Feature	Anthropometric measurement	Design measurement (cm)
Pillow length (head part)	Based on journals	76
Pillow width (head part)	Based on journals	51
Pillow depth (head part)	Head breadth	25
Neck rest	Nape length	18
Shoulder rest	Shoulder width	20
	Shoulder length	28
Upper back rest	Bust breadth, side view	25
Lower back rest	Below the bust width, side view	9
	Waist breadth, side view	26
Pelvic area rest	Hip breadth, side view	20
Leg rest	Leg length up to knees	23
	Leg width	20
Full body rest	From shoulder to hips	66

memory foam, but it is lighter and more breathable. These pillows have some of the highest satisfaction ratings of all pillows and can offer the user a very comfortable pain-free night. For the fabric, cotton was ideal for those with skin sensitivities, it is also breathable, soft and comfortable.

Table 5 shows the design parameter of the pillow. These data were based on the anthropometric measurement of the respondents.

The Maternity Pillow's design is shown in Fig. 1 with its corresponding measurements that came from the average anthropometric measurement of the pregnant women as the basis of the design parameters. And its features and design were the ones getting the highest frequency of pregnant women from the options given in the survey questionnaire (Fig. 2).

Consequently, the design of a maternity pillow is divided into 5 parts: (A) Upper Head Part, it is a wedge type of pillow. It can be folded to adjust the preferred height of the head of the pregnant woman. This will help the pregnant woman to breathe properly while sleeping. Also, this pillow serves as a backrest while sitting; (B) Lower Head Part, it is to support the head while sleeping, it is ergonomically designed to improve posture, and relieve back and neck pain, the pillow should be neutral align to spine [9]; (C) Connected by magic tape, this will serve as the binder of the various parts of the pillow; (D) Left I-Shaped pillow, it will support the back part of the body; I Right I-Shaped pillow, it will support the tummy and pelvic area; and (F) Pillow for between the legs, it helps to relieve the tension and reducing water retention in legs and ankles.

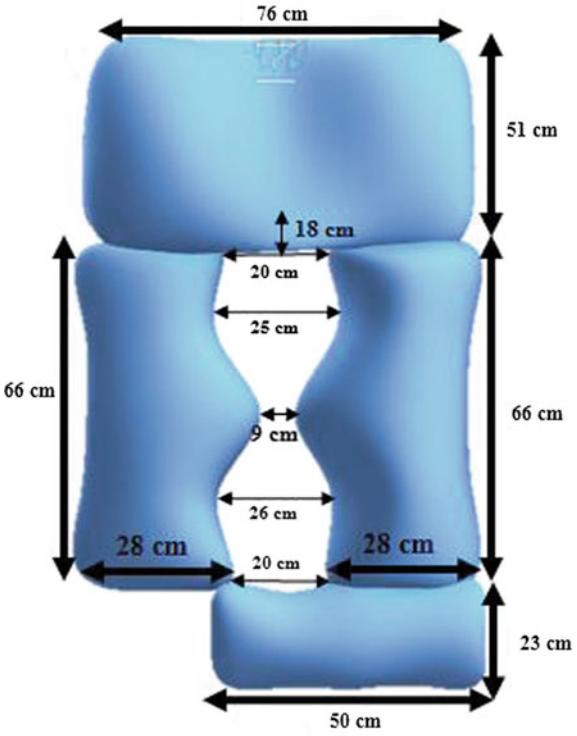


Fig. 1 The design of ergonomically designed maternity pillow

3.2 Validation of Proposed Maternity Pillow

Then, the result of the validation of the pain assessment per body parts and the validation using the QVAS questionnaire was compared. To determine if the design has a huge help reduce or disappear the intensity of pain that the pregnant women from Barangay Hornalan and Laguerta were experiencing.

Table 6 shows the comparison of the result of the QVAS where it shows that before the validation, the respondents rated the pains that they were experiencing in every part of their body to be on the high intensity of pain. While after the validation, where the respondents already experienced the maternity pillow, it shows that the pains they were experiencing before were now reduced since they rated the pain to be on the low intensity of pain (Table 7).

It shows how satisfied the respondents are towards the materials used in the pillow such as the filling and the fabric. It shows that they were highly satisfied on it.

Table 8 shows the assessment of thickness per body parts. It shows that most of the respondents preferred the thickness of their pillow to be on the high level, but there were also some of them who wanted it to be on the medium level.

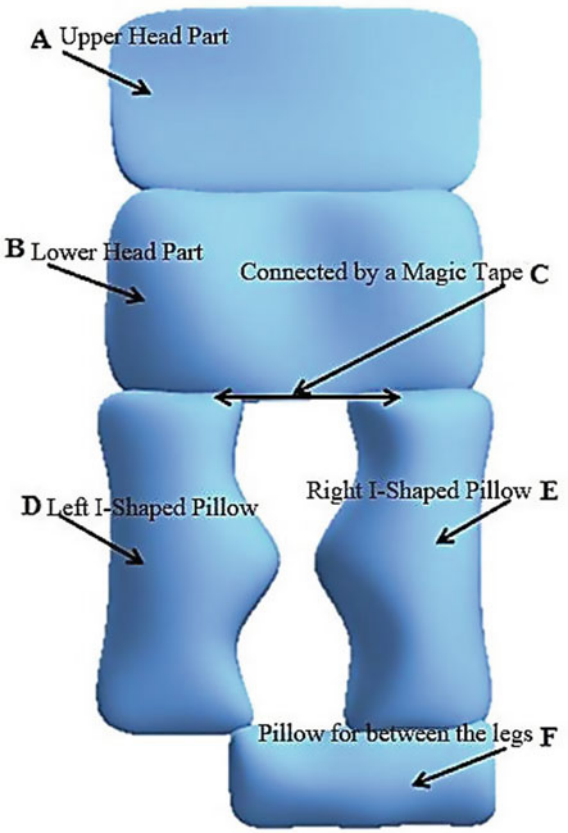


Fig. 2 Parts of the proposed design maternity pillow

Table 6 Comparison of QVAS result from before and after validation

Body part	Before validation		After validation	
	Percentage (%)	Interpretation	Percentage (%)	Interpretation
Neck	43	High intensity pain	53	Low intensity pain
Shoulder	47	High intensity pain	67	Low intensity pain
Upper back	63	High intensity pain	73	Low intensity pain
Lower back	77	High intensity pain	73	Low intensity pain
Pelvic area	60	High intensity pain	60	Low intensity pain
Legs	53	High intensity pain	53	Low intensity pain
Shoulder	47	High intensity pain	67	Low intensity pain

Table 7 Pillow's materials

Materials	Mean	Scale
Latex foam filling	3.67	Highly satisfied
Cotton fabric	3.87	Highly satisfied

Table 8 Pillow's thickness

Body parts	Medium level (13–15 cm)		High level (17–20 cm)	
	Mean	Scale	Mean	Scale
Neck	2.83	Fairly satisfied	3.37	Highly satisfied
Shoulder	2.87	Fairly satisfied	3.47	Highly satisfied
Upper back	2.8	Fairly satisfied	3.3	Highly satisfied
Lower back	3.03	Fairly satisfied	3.57	Highly satisfied
Pelvic area	3.17	Fairly satisfied	3.4	Highly satisfied
Legs	2.93	Fairly satisfied	3.23	Highly satisfied

Table 9 Over-all experience with the maternity pillow

Overall experience	Mean	Scale
Overall pillow's appearance	3.63	Highly satisfied
Overall pillow's comfort	3.53	Highly satisfied

Table 9 shows how satisfied the respondents were about the overall experience they have using the pillow. They rated their experience using it as highly satisfied.

The effectiveness of the pillow regarding the thickness, design, appearance, and comfort was rated by most of the respondents by being highly satisfied with it.

4 Conclusion

As the study went along, the researchers determined the demographic profile of the respondents such as their age, periods of pregnancy and number of birth. The researchers also identified that there are multiple of pains that the pregnant women are experiencing aside from the lower back pain. Periods of pregnancy and number of birth, especially. If they were on their first pregnancy have a meaningful relationship to the pain that they were experiencing. The researchers also found out that the anthropometric measurement has also a significant relationship to the design of maternity pillow. This has been clear with the use of different statistical tools that made the output of the data as stable as it is.

Therefore, the initial data gathered before the validation the pains were identified, such as neck, shoulder, upper back, lower back, pelvic area, and legs all of

them were identified if they have a relationship to the demographic profile (age, the period of pregnancy and number of births) of the respondents using the Pearson's Correlation. The pains that they were experiencing were rated as being on the high-intensity pain based on the QVAS questionnaire. The effectiveness of the pillow regarding the thickness, design, appearance, and comfort was rated by most of the respondents by being highly satisfied with it.

Accordingly, ergonomically designed maternity pillow that will fit the user and the support provided has a good benefit to them, which can lessen the pains that they are experiencing. Towards preventing sleep disorders in the pregnant women that may generate some unwanted events. Essentially, having good and comfortable sleep for pregnant women may lead to an awesome morning awakening and healthy baby.

References

1. Almeida D et al (2016) Risk of miscarriage in women receiving antidepressants in early pregnancy, correcting for induced abortions. Epidemiol J 1–4
2. National Sleep Foundation, https://sleepfoundation.org/sleep-polls-data/sleep-in-america-poll/2007-women-and-sleep
3. Baby Centre Homepage, https://www.babycentre.co.uk/x1047808/is-it-safe-to-sleep-on-my-back-now-that-im-pregnant
4. O'Brien LM et al (2015) Maternal sleep position: what do we know where do we go? O'Brien and Warland BMC pregnancy and childbirth, pp 1–2
5. Bat-Pitault F et al (2015) Sleep pattern during pregnancy and maternal depression: study of Aube cohort. J Sleep Dis Manag 1–8
6. Hintze JP et al (2010) Diagnosis and treatment of restless leg syndrome (RLS) during pregnancy. Nat Healthy Sleep Project 1–4
7. Rahmillah et al (2017) Design of maternity pillow by using Kansei and Taguchi methods. In: IOP conference series: materials science and engineering, pp 1–9
8. Christensen K Outcomes assessment in chiropractic practice. Dynamic chiropractic. http://www.dynamicchiropractic.com/mpacms/dc/article.php?id=52325&no_paginate=true&p_friendly=true&no_b=true. Last assessed 2007/09/10
9. Wong et al (2016) Effect of pillow height on the biomechanics of the head-neck complex: investigation of the cranio-cervical pressure and cervical spine alignment. Peer J journal

Disaster Relief Model in Laguna Using Integer Linear Programming

Apollo Panzo, Mary Rose Ann Sangalang, Karl Anthony Ibali, Noemi Paulen Lizardo and Philip Ermita

Abstract This paper presents a disaster relief model that determined the areas with high to low susceptibility of flooding and gave an optimal route for vehicles delivering relief goods to the flood prone areas in the Province of Laguna, Philippines. This optimization problem is characterized as a capacitated vehicular routing problem. Data from 2012 to 2017 typhoons from the Provincial Risk Reduction Management Office (PDRRMO) were used to analyze the typhoons and generate specific routes of each vehicle from the warehouses to the affected areas. Microsoft Visual Studio was used to give the optimal routes to be covered and the minimum total travel time to be taken by the delivery vehicles. The solution to this problem can be used by PDRRMO in efficiently distributing the relief goods to different areas in the Province of Laguna and as a result, utilizing the use of government's resources to its full potential.

Keywords Disaster relief model · Integer linear programming · Microsoft visual studio

A. Panzo (✉) · M. R. A. Sangalang · K. A. Ibali · N. P. Lizardo · P. Ermita
Department of Industrial Engineering, University of Perpetual Help System DALTA Calamba, Calamba, Laguna, Philippines
e-mail: apollopanzo@gmail.com

M. R. A. Sangalang
e-mail: mrosesangalang07@gmail.com

K. A. Ibali
e-mail: cedie1917@gmail.com

N. P. Lizardo
e-mail: chineedelosantos11@gmail.com

P. Ermita
e-mail: philip.ermita@perpetualdalta.edu.ph

K. J. Kim and H. Kim (eds.), *Mobile and Wireless Technology 2018*, Lecture Notes in Electrical Engineering 513, https://doi.org/10.1007/978-981-13-1059-1_44

1 Introduction

Disaster is a sudden, calamitous event, bringing great damage, loss, destruction and devastation to life and property. The damage caused by disasters is immeasurable and influences the mental, socioeconomic, political, and cultural state of the affected area. Disasters are events that inflict great damage and human suffering. Their origin can be natural, such as earthquakes, floods and hurricanes, or of human origin: accidents and terrorists acts.

Laguna is a province in the Philippines located in the CALABARZON region in Luzon. Its capital is Santa Cruz and the province is situated southeast of Metro Manila, south of the province of Rizal, west of Quezon, north of Batangas and east of Cavite. Laguna hugs the southern shores of Laguna de Bay, the largest lake in the country.

Provincial Disaster Risk Reduction Management Office (PDRRMO) is an establishment institutionalized by RA 10121 or "an act strengthening the Philippines disaster risk reduction and managements system, providing for the national disaster risk reduction and management plan." They help with the distribution of relief to evacuation centers. Shortage of relief goods is one of the major issues in such events. In one of their operations, they have only served a total of 165 out of 251 barangays which is about 65.7% of the areas that are prone to flooding. The department, thus, needs an effective distribution model to address this issue. Having such effective delivery system is vital in the distribution of relief.

Due to the urgent need of relief goods, an optimal route for the delivery vehicles is needed. These relief goods must satisfy the demand of each area [1].

This research focused on the delivery of disaster relief done by Provincial Risk Reduction Management Office (PDRRMO) of the province of Laguna from its main warehouse in Sta. Cruz, Laguna and its satellite office in Barangay Halang, Calamba City, and Laguna to the evacuation areas. PDRRMO are having problems on how to distribute their relief goods fast to affected areas when disaster strikes. Drivers of delivery trucks have many choices of routes to follow, hence, choosing the best route to get to their destination can be difficult.

2 Research Methodology

2.1 Data Gathering

The researchers used an oral interview—it is an open-ended question which allows the development of both researchers and respondents to ask and answer in their own words. Mr. Rex Maunahan, staff of PDRRMO, was interviewed to acquire knowledge about the process of distributing the relief goods. PDRRMO also provided the following information:

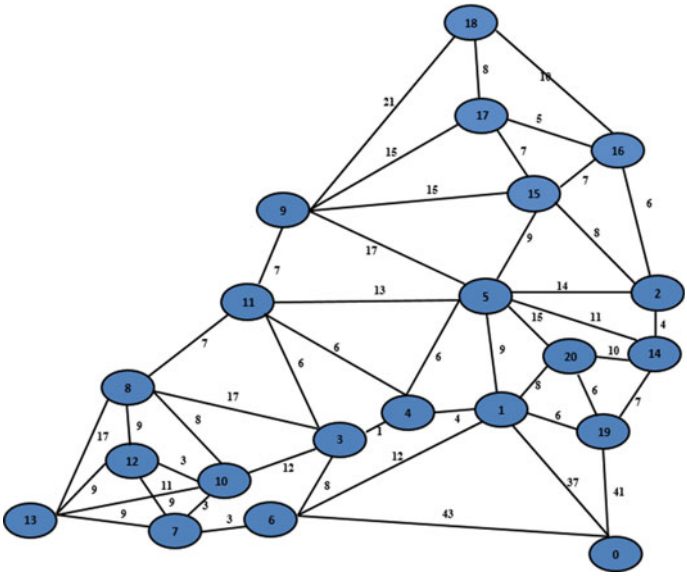


Fig. 1 Network representation of San Pedro City, Laguna

- data of typhoons from 2012 to 2017;
- number of vehicles they use to distribute the relief and their respective capacities; and
- the susceptibility to flood of the towns and barangays in the province of Laguna.

Google Maps was used to establish the distances between barangays in every town. The relative travel time between two barangays was computed using the distance given by Google Maps. Figures 1, 2 and 3 show the map and their corresponding network representation created for each town in the 1st District of the province of Laguna. The nodes represent the barangays of each town. Node 0 is the source node or the warehouse which the relief goods will come from and in this case it is from their satellite office in Brgy. Halang, Calamba City, Laguna. Google was also used to determine the household population of each barangay which was used to estimate the number of families in the affected area. It was also used to describe the demographic profile of Laguna (Table 1).

2.2 Method of Analysis

This research is an applied research which is a type of research that involves seeking new applications of scientific knowledge to the solution of a particular problem, such as the development of a new system or produce, new device, or new method to solve the problem. The study entitled "Disaster Relief Model in Laguna

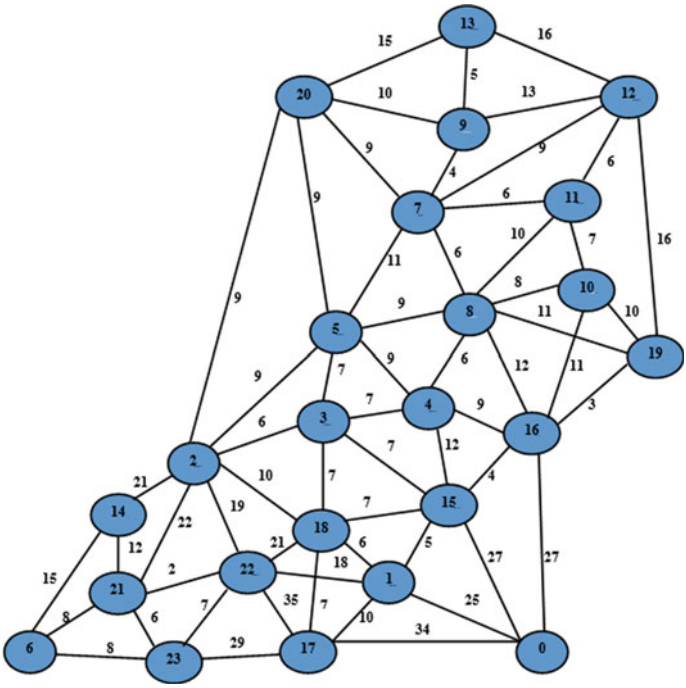


Fig. 2 Network representation of Biñan City, Laguna

using Integer Linear Programming" utilized an applied research design for the reason that it attempted to develop a new distribution route system of relief goods.

3 Results and Discussion

3.1 Data and Analysis

The historical data were provided by PDRRMO main office located at Sta. Cruz, Laguna. These were evaluated and interpreted in order to answer the problems. The researchers made a network representation of every town with their corresponding barangays to determine the distance from the main warehouses.

Based on the data from the Mines and Geosciences Bureau, portions of Calamba City, Bay, Famy, Lumba, Mabitac, Pangil, Pakil, Siniloan, Sta. Maria and Sta. Rosa are highly susceptible to flood hazard. Biñan, Cabuyao, Calauan, Kalayaan, Liliw, Los Baños, Magdalena, Majayjay, Rizal, Paete, Pagsanjan, Pila, San Pedro, and Victoria are moderately susceptible. San Pablo City, Nagcarlan, Alaminos, Liliw, Cavinti, and Luisiana on the other hand are not susceptible to flooding so they were not used in this research.

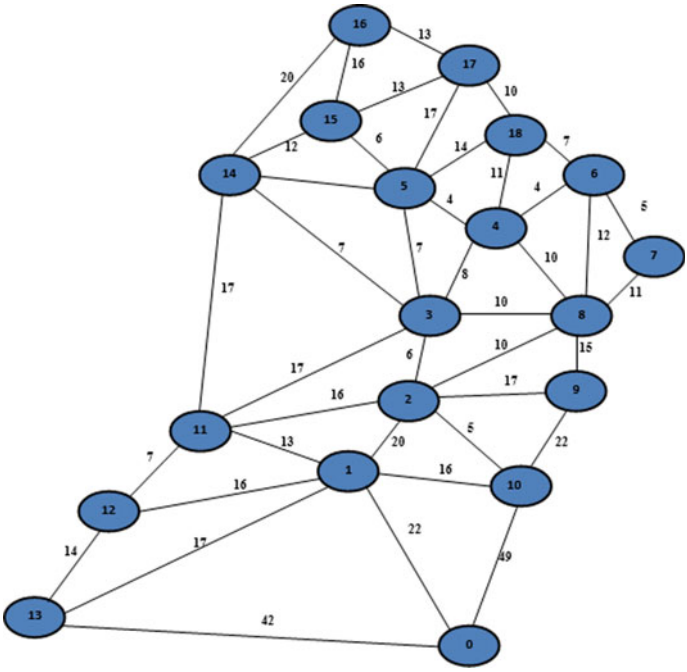


Fig. 3 Network representation of Sta. Rosa City, Laguna

Out of the 674 barangays in the province, 10% or 67 barangays are highly susceptible, 37% or 251 barangay are moderately susceptible and 53% or 356 are not susceptible (Fig. 4).

Table 2 shows the population of every barangay in the cities of San Pedro, Biñan and Sta. Rosa which is a factor in distributing of relief goods. Because of ever-changing demand, unstable relief operation environment, and the unavailability of accurate information on the amount of necessary supplies of relief goods to meet the needs of the affected population, the government and humanitarian actors participating in the relief operations rely on rough estimates to procure the goods [2].

The number of vehicles used in delivering goods is limited. The vehicles need to go back and forth to the affected city to satisfy all the demand. The department thus, needs an efficient distribution system. Developing such system is crucial in any organization dealing with multiple distribution points. Several constraints must be taken into consideration [1].

In order to serve the affected areas, PDRRMO has utilized the use of their delivery vehicles. As of today, they have 4 delivery trucks at their disposal. Each truck has capacity of 4000 packs. The availability of these trucks will be crucial to the delivery of relief goods to affected areas.

Table 1 Node labels for the cities of San Pedro, Biñan and Sta. Rosa

San Pedro City		Biñan City		Sta. Rosa City	
Node	Barangay	Node	Barangay	Node	Barangay
0	Halang	0	PDRRMO-Halang	0	PDRRMO-Halang
1	United Bayanihan	1	Santo Tomas	1	Malitlit
2	Landayan	2	San Francisco	2	Balibago
3	Narra	3	Soro-Soro	3	Macabling
4	Riverside	4	San Vicente	4	Kanluran (POB)
5	Calendola	5	Tubigan	5	Malusak (POB)
6	Laram	6	Binan Poblacion	6	Ibaba
7	Bagong Silang	7	Poblacion	7	Labas
8	GSIS	8	Santo Domingo	8	Pook
9	San Antonio	9	San Jose	9	Dila
10	Sampaguita Village	10	San Antonio (Halang)	10	Dita
11	United Better Living	11	Casile	11	Pulong Santa Cruz
12	Estrella	12	Malaban	12	Don Jose
13	Langgam	13	De la Paz	13	Santo Domingo
14	Nueva	14	Langkiwa	14	Tagapo
15	Poblacion	15	Bungahan	I5	Market Area (POB)
16	San Roque	16	Zapote	16	Sinalhan
17	Santo Nino	17	Ganado	17	Aplaya
18	Cuyab	18	Mamplasan	18	Caingin
19	Magsaysay	19	Platero		
20	San Vicente	20	Canlalay		
		21	Timbao		
		22	Loma		
		23	Malamig		

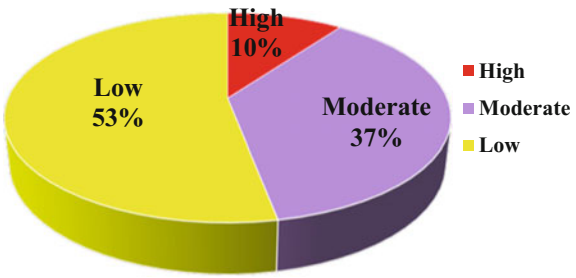


Fig. 4 Flood susceptibility in Laguna

Delivery of the humanitarian aids from suppliers to shelters must be done within certain time limits. Though helicopters are used for transporting a fraction of the daily requirements, the capacity of small aircrafts limits the throughput of this mode. Ground transportation modes are still playing a major role in the humanitarian logistics [3].

Table 2 Household population of the cities of San Pedro, Biñan and Sta. Rosa

San Pedro City, Laguna household population		Biñan City, Laguna household population		Sta. Rosa City, Laguna household population	
Bagong Silang	1385	Biñan(Poblacion)	1037	Aplaya	3880
Cuyab	5433	Bungahan	427	Balibago	5149
Estrella	1870	Santo Tomas (Calabuso)	10770	Caingin	5399
G.S.I.S.	717	Canlalay	4850	Dila	7716
Landayan	7852	Casile	1037	Dita	7065
Langgam	7406	De La Paz	7844	Don Jose	4718
Laram	1571	Ganado	1313	Ibaba	2073
Magsaysay	3043	San Francisco (Halang)	7167	Labas	4849
Nueva	1242	Langkiwa	9454	Macabling	4858
Poblacion	1473	Loma	3001	Malitlit	6125
Riverside	807	Malaban	6628	Malusak (Pob.)	666
San Antonio	15357	Malamig	1016	Market Area (Pob.)	3528
San Roque	1753	Mampalasan	1728	Kanluran (Pob.)	1196
San Vicente	6532	Platero	2605	Pook	10696
Santo Niño	1043	Poblacion	741	Pulong Santa Cruz	5638
United Bayanihan	1772	Santo Niño	1389	Santo Domingo	1023
United Better Living	1630	San Antonio	8953	Sinalhan	5449
Sampaguita Village	1433	San Jose	1494	Tagapo	8415
Calendola	1182	San Vicente	2133		
Narra	607	Soro-Soro	1580		
		Santo Domingo	1526		
		Timbao	3373		
		Tubigan	1685		
		Zapote	1507		

The researchers used Microsoft Visual Studio to validate the data they have gathered. Figure 5 shows the software the researchers used to interpret the data they have gathered and analyze.

This study started from analyzing the data given by Provincial Risk Reduction Management Office. Hence, after proposing a distribution model, the researchers were able to identify the areas that need to be prioritized and the optimal route from the main warehouses to the affected areas.

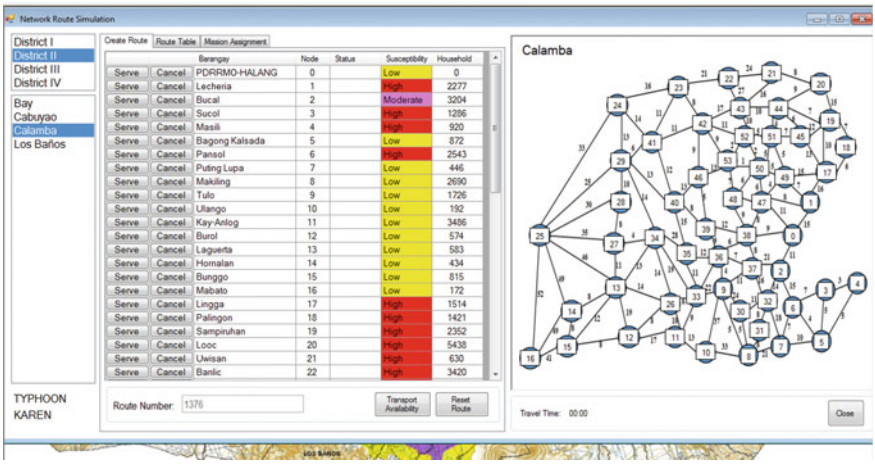


Fig. 5 Microsoft visual studio network route simulation

4 Conclusion

Throughout the year, Philippines experiences the devastating effects of typhoons. This paper presents a methodology to come up with an ideal relief model of delivering relief goods in the affected areas of the Province of Laguna. The researchers have concluded that Laguna is a flood prone area, thus, having an effective distribution model will be a big help to mitigate the effects of such calamitous event. The historical data of typhoons that hit the Province of Laguna helped a lot in analyzing the areas that need to be prioritized when distributing relief goods. Analyzing the significant factors that affect the distribution of relief goods has helped the researchers find a way to solve the problem. The areas that need to be prioritized must always be taken into consideration when distributing relief goods. Determining the optimal route to the distribution of relief goods in Laguna will help the government utilize its full potential to aid victims of such catastrophic event.

References

1. De Lara ML et al (2013) Determining the optimal route of vehicles delivering relief goods to the calamity-prone areas in region IV-A. J Nat Stud 13(2):13–24
2. Fritz Institute (2003) Enabling disaster response. [White Paper]. Retrieved from http://www.fritzinstitute.org/pdfs/whitepaper/enablingdisasterresponse.pdf
3. Hamedi M et al (2012) Reliable transportation of humanitarian supplies in disaster response: model and heuristic. Procedia Soc Behav Sci 54:1205–1219