Lecture Notes in Electrical Engineering 391

Kuinam J. Kim
Naruemon Wattanapongsakorn
Nikolai Joukov
Editors

Mobile and Wireless Technologies 2016

Springer

Lecture Notes in Electrical Engineering

Volume 391

About this Series

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

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

LNEE publishes authored monographs and contributed volumes which present cutting edge research information as well as new perspectives on classical fields, while maintaining Springer's high standards of academic excellence. Also considered for publication are lecture materials, proceedings, and other related materials of exceptionally high quality and interest. The subject matter should be original and timely, reporting the latest research and developments in all areas of electrical engineering.

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

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

Kuinam J. Kim · Naruemon Wattanapongsakorn
Nikolai Joukov
Editors

Mobile and Wireless Technologies 2016

Springer

Editors
Kuinam J. Kim
Kyonggi University
Seongnam-si, Gyeonggi
Korea, Republic of (South Korea)

Nikolai Joukov
IBM T.J. Watson Research Center
New York, NY
USA

Naruemon Wattanapongsakorn
King Mongkut's University of Technology
Thonburi
Thung Khru, Bangkok
Thailand

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-981-10-1408-6 ISBN 978-981-10-1409-3 (eBook)
DOI 10.1007/978-981-10-1409-3

Library of Congress Control Number: 2016940128

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Science+Business Media Singapore Pte Ltd.

Preface

This LNEE volume contains the papers presented at the International Conference on Mobile and Wireless Technology (ICMWT2016) which was held in Jeju Island, South Korea, during May 23–26, 2016.

ICMWT2016 received over 100 paper submissions from various countries. After a rigorous peer-review process, 26 full-length articles were accepted for presentation at the conference. This corresponds to an acceptance rate was very low and is intended for maintaining the high standards of the conference proceedings.

ICMWT2016 will provide an excellent international conference for sharing knowledge and results in Mobile and Wireless Technology. The aim of the Conference 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 Mobile and Wireless Technology related fields.

On behalf of the Organizing Committee, we would like to thank Springer for publishing the proceedings of ICMWT2016. 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 also spending count-less days in preparing the final program in time for printing. We would also like to thank our organization committee for their hard work in sorting our manuscripts from our authors.

We look forward to seeing all of you next year at ICMWT.

Kuinam J. Kim
Nikolai Joukov

Organizing Committee

General Co Chairs

Nikolai Joukov, New York University and modelizeIT Inc., USA
HyeunCheol Kim, NamSeoul University, Republic of Korea

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

Cheong Ghil Kim, Namseoul University, Republic of Korea
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

Donghwi Lee, University of Colorado, USA
Jinsul Kim, Chonnam National University, Republic of Korea

Program Chair

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

Organizers & Supporters

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

Program Committee

Yadigar Imamverdiyev, Azerbaijan National Academy of Sciences, Azerbaijan
Kok-Seng Wong, Soongsil University, South Korea
Jitender Grover, M.M.U., Sadopur, India
Mohammed Awadh Ahmed Ben Mubarak, Infrastructure University Kuala Lumpur (IUKL), Malaysia
Yin-Fu Huang, National Yunlin University of Science and Technology, Taiwan
Seung Yeob Nam, Yeungnam University, South Korea
Muhammad Atta-ur-Rahman, PMAS Arid Agriculture University, Pakistan
Zied Chtourou, Research Center for Computer Science and Multimedia of Sfax, Tunisia
Fu-Chien Kao, Da-Yeh University, Taiwan

Nikolaos (Nikos) Petrellis, Technological Educational Institute (TEI) of Thessaly, Greece
Ljiljana Trajkovic, Simon Fraser University, Canada
Lukasz Krzywiecki, Wrocław University of Technology, Poland
Seyed Hossein Siadat, Shahid Beheshti University, Iran
Mohd Faizal Abdollah, University Technical Malaysia Melaka, Malaysia
Vasco N.G.J. Soares, Polytechnic Institute of Castelo Branco, Portugal
Xiaodi Li, Shandong Normal University, China
Afizan bin Azman, Multimedia University, Malaysia
Jinhong Kim, Sunkyunkwan University, South Korea
Susan Mengel, Texas Tech University, USA
Hyunsung Kim, Kyungil University, South Korea
Rajib Kar, NIT Durgapur, West Bengal, India
Harikumar Rajaguru, Institute of Technology Sathyamangalam, India
Ghada Al-Hudhud, King Saud University, Saudi Arabia
Mohammad Faiz Liew Abdullah, Universiti Tun Hussein Onn Malaysia, Malaysia
Srinivas Sethi, IGIT Sarang, India
Houcine Hassan, Universitat Politecnica de Valencia, Spain
Partha S. Mallick, VIT University, India
Azeddien M. Sllame, University of Tripoli, Libya
Robert S. Laramee, Swansea University, UK
Sattar B. Sadkhan Al-Maliky, Chairman of IEEE ComSo Iraq Chapter, Iraq
Anton Setzer, Swansea University, UK
Mainguenaud Michel, Institut National des Sciences Appliquées, France
Hejab Maazer Al fawareh, Northern Border University, Saudi Arabia
Shakeel Ahmed, King Faisal University, Saudi Arabia
Goi Bok Min, Universiti Tunku Abdul Rahman (UTAR), Malaysia

Contents

Part IV Mobile Software

Part I
Mobile, Wireless Networks and Applications

Adaptive Resource Allocation in LTE Downlink Transmission Systems

Ruiyi Zhu, Jian Yang, and Peng Si

Department of Automation
University of Science and Technology of China
Hefei, China 230027
Email: zhuruiyi@mail.ustc.edu.cn, jianyang@ustc.edu.cn, daisys@mail.ustc.edu.cn

Abstract. Dynamic resource allocation in 3GPP Long Term Evolution (LTE) for satisfying Quality-of-Service (QoS) requirement as well as improving the system throughput has gained significant attentions. In this paper, we proposed a buffer-aware adaptive resource allocation scheme for LTE downlink transmission to improve the overall system throughput while providing statistic QoS guarantee and keep certain fairness among users. Specifically, each users' queue priority in the base station is ranked according to its remaining life time or its queue overflow probability which is estimated by applying large deviation principle. Based on the queue priorities, we dynamically allocate the Resource Blocks (RBs) for avoiding buffer overflow and guaranteeing the statistic QoS. The simulation results demonstrate that the proposed algorithm improves the system throughput and fairness while considerably reduces the average bit loss rate.

Keywords: Dynamic resource allocation; LTE; QoS; Large deviation principle

1 Introduction

3GPP Long Term Evolution (LTE) is one of the major steps in mobile communication to provide high-data-rate, low-latency, packet-optimized radio-access and flexible bandwidth deployments for next-generation mobile broadband networks [1]. Orthogonal Frequency-Division Multiple Access (OFDMA) is adopted in LTE downlink transmission systems which allows high flexibility in the resource allocation among the potential users. Due to the diverse channel qualities of the users as well as Quality-of-Service (QoS) requirement, designing a multi-user resource allocation scheme in LTE downlink transmission systems to improve the total system throughput, guarantee QoS requirements and achieve fairness still be an interesting and challenging problem.

Resource allocation in wireless data systems has been discussed in several related works. A scheduling algorithm in [2] for downlink multiuser systems considers both channel state and buffer status in order to reduce average packet delay, while maintaining the stability condition of the networks. In [3], a resource

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_1

allocation scheme in LTE systems is presented for reducing the waste of system resource and improving the total throughput by utilizing the knowledge of buffer status and channel conditions. However, these algorithms have not considered the system fairness which plays an important role in the system performance. In fact, providing fairness among users is an essential design consideration, although it will usually sacrifice the system throughput and/or violate QoS requirements.

PF algorithm achieves a tradeoff between system throughput and fairness. In [4], S. Lee proposes a sub-optimal method, i.e., PF metric(2), which introduces the status of queues into PF metric(1). However, it is pointed out that PF metric(2) may result in an inefficient Resource Block (RB) assignment, because a single RB is considered in isolation one-by-one regardless of other RB assignment status. By considering both the constraint of finite buffer space and fairness, in [5], Yan Lin proposes a Channel-Adapted and Buffer-Aware (CABA) packet scheduling algorithm which applies the user priority in the resource allocation to avoid buffer overflow. However, it is very different to choose the empirical parameters in the priority function appropriately. It may induce excessive resource allocated to the users, which reduces the system utility.

In this paper, we propose an adaptive resource allocation scheme by jointly considering the user scheduling and RBs allocation to improve overall system throughput, provide QoS guarantee and keep certain fairness among users in LTE downlink transmission systems. For user scheduling, each user's queue priority is ranked according to its remaining life time or its queue overflow probability which is estimated by applying large deviation principle. In the aspect of RBs allocation, an online measurement based algorithm for adjusting the service rates is proposed based on the user queues' priorities and dynamically allocates RBs in order to avoid buffer overflow and satisfy QoS requirement.

The rest of the paper is organized as follows: Section 2 describes the resource allocation system model and the problem statement. Section 3 is devoted to describing the user priority determination algorithm. Section 4 we present an online measurement based algorithm for dynamic resource allocation. Section 5 shows the simulation results. Section 6 concludes the paper.

2 System Model and Problem Statement

2.1 System Model

We consider LTE system architecture with downlink resource allocation as shown in Fig. 1. The eNode B (eNB) controls the bit service rate through dynamically allocating RBs to users. According to the time and frequency domain, the structure of RBs can be known from [7]. The total number of the data bits within a RB is referred to as RB capacity. The better channel condition of an RB implies a higher achievable RB capacity. Different RBs may have distinct channel conditions [6]. The smallest resource unit that can be allocated to a user is a Scheduling Block (SB), which consists of two consecutive RBs [7]. In each time slot, several SBs may be allocated to a single user, but each SB is uniquely assigned to a user.

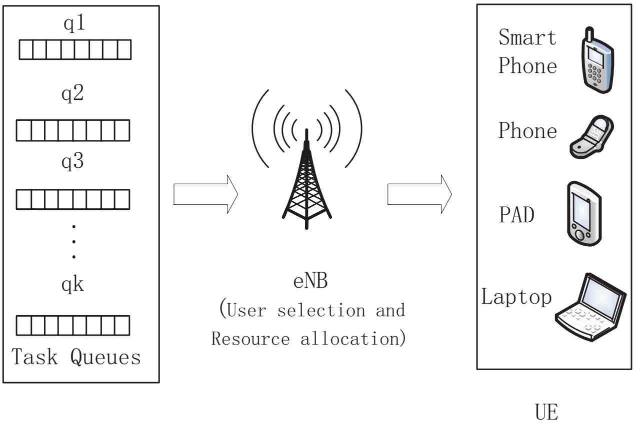


Fig. 1. System Model

Let K and N, respectively, denote the number of the users and the number of all SBs. One user is assumed to have a single queue. Then, the kth queue length can be updated as

$$Q_k(t+1) = max\{Q_k(t) - D_k(t), 0\} + A_k(t), \tag{1}$$

where $A_k(t) \in A = \{0, 1, \ldots, m_A\}$ denotes the bit arrival number of the kth user queue during the tth slot , and m_A is the maximum number of bits arriving in a single slot. Here, we assume that the bit arrival process of the kth user queue $A_k(t)$ to be an $i.i.d$ sequence. $D_k(t) \in D = \{0, 1, \ldots, m_D\}$ represents the bit number transmitted in the tth slot, where m_D is the maximum number of bits served in a single slot. Define $Q_k(t)$ as the length of the kth queue in terms of bits at the beginning of the slot t, Q_k^{max} as the the maximum length of the kth user queue. If the kth queue length is higher than Q_k^{max}, it implies that an excessive number of bits is buffered, and bit loss may be likely to occur. Naturally, any queue length exceeding Q_k^{max} is undesirable. Hence, the problem may be described as that of selecting an appropriate service rate to keep each queue length lower than Q_k^{max}. We plan to apply the large deviation algorithm to calculate the queue overflow probability to reduce the average Bit Loss Rate (BLR). Accordingly, we define overflow probability of the kth queue as

$$P_{k_{overflow}} = P(Q_k(t) > Q_k^{max}). \tag{2}$$

For a given slot t, the increase of the kth queue length is characterized by $I_k(t) = A_k(t) - D_k(t)$.

We define remaining life time $R_k(t)$ to denote the remaining time of the kth user queue to be fullness, which can be calculated by

$$R_k(t) = \frac{Q_k^{max} - Q_k(t)}{E[I_k(t)]}. \tag{3}$$

However, the distribution of $I_k(t)$ is not available. Hence, we use the sample mean of queue variations in the last N slots to estimate $E[I_k(t)]$, i.e., $\frac{1}{N}\sum_{t_0=t-N-1}^{t-1}(A_k(t_0)-D_k(t_0))$. Then, $R_k(t)$ can be calculated by

$$R_k(t)=\frac{Q_k^{max}-Q_k(t)}{\frac{1}{N}\sum_{t_0=t-N-1}^{t-1}(A_k(t_0)-D_k(t_0))}. \tag{4}$$

2.2 Problem Statement

Channel quality indicator (CQI) reporting procedure is a fundamental feature of LTE networks since it enables the estimation of the downlink channel quality at the eNB. UE reports a CQI value of each RB to the eNB, and the eNB uses CQI for the resource allocation [8]. Let $r_k^n(t)$ denote instantaneous data transmission rate when the nth SB is assigned to the kth user queue at the slot t. According to Channel quality indicator (CQI) information, $r_k^n(t)$ can be calculated using the Adaptive Modulation and Coding(AMC) module or simply estimated via the well-known Shannon's formula for the channel capacity [8], i.e., $r_k^n(t)=log_2(1+\gamma_{k,n})$, where $\gamma_{k,n}$ is the Signal-to-Interference-plus-Noise-Ratio (SINR) for the kth user on the nth SB. Let $\rho_k^n(t)$ indicate whether the nth SB is assigned to the kth user at the slot t. If the nth SB is assigned to the kth user at the slot t, we have $\rho_k^n(t)$=1. Otherwise $\rho_k^n(t)$=0. Then, the resource allocation problem can be defined as improving the system throughput as large as possible, i.e.,

$$\max\sum_{k=1}^{K}\sum_{n=1}^{N}\rho_k^n(t)r_k^n(t), \tag{5}$$

In this paper, we apply *Jain's* fairness index [11], $F(t)$, to indicate the fairness to denote the system fairness at time t. The constraints are listed as follows:

$$\sum_{k=1}^{K}\rho_k^n(t)=1, n\in\{1,2,\ldots,N\}. \tag{6}$$

$$1-F(t)\le\xi. \tag{7}$$

Eq. (6) indicates that each SB can be only assigned to one user in the tth slot. Equation (7) indicates that the difference between 1 and the system fairness should be kept less than ξ.

The resource allocation problem (5-7) is complicated and intractable to obtain the optimal solution by exhaustive search. Here, we propose an adaptive resource allocation scheme which considers both the different priorities of user queues and RB capacity, in order to achieve a better performance tradeoff among throughput, fairness and average BLR.

3 User Priority Determination Algorithm

In this section, we determine the priorities of the users' queue by their remaining life time or their queue overflow probability which is calculated by applying large deviation principle.

3.1 Estimation Model for the Queue Overflow Probability

In this section, we apply large deviation theory to propose an estimation model for the queue overflow probability. Let $I_k(t) = A_k(t) - D_k(t)$, which characterizes the mismatch between the bit service rate and the bit arrival rate of the kth user queue during the slot t, where $I_k(t) \in \{-m_D, \cdots, 0, 1, \cdots, m_A\}$. Let $\pi_i^k = P(I_k(t) = i)$ denote the corresponding kth user queue-length variation probability distribution.

The kth user queue length increment during the period spanning from the tth slot to the $(t+T)$th slot can be formulated as

$$I_k^{t+T} = \sum_{i=1}^{T} I_k(t+i). \tag{8}$$

where T is called prediction interval. Then, the length of the kth user queue at the beginning of the $(t+T)$th slot can be expressed as

$$Q_k(t+T) = Q_k(t) + I_k^{t+T}. \tag{9}$$

Let $P_{k_{overflow}}^{t+T}$ denote the overflow probability of the kth user queue during the slot $(t+T)$, which is defined as

$$\begin{aligned} P_{k_{overflow}}^{t+T} &= P(Q_k(t+T) > Q_k^{max}) \\ &= P(Q_k(t) + I_k^{t+T} > Q_k^{max}). \end{aligned} \tag{10}$$

Define the achievable average queue growth of the kth user queue during the future T slots as

$$a_k = \frac{Q_k^{max} - Q_k(t)}{T}. \tag{11}$$

and the expected average queue growth of the kth user queue in each slot during the T slots as

$$b_k = E[\frac{\sum_{i=1}^{T} I_k(t+i)}{T}]. \tag{12}$$

where $E[\cdot]$ denotes expectation operator. $b_k \geq a_k$ implies that there is high overflow possibility of the kth user queue after T slots. Eq. (11) can be further written as

$$\begin{aligned} P_{k_{overflow}}^{t+T} &= P(Q_k(t) + I_k^{t+T} > Q_k^{max}) \\ &= P(I_k^{t+T}/T > (Q_k^{max} - Q_k(t))/T) \\ &= P(\frac{\sum_{i=1}^{T} I_k(t+i)}{T} > a_k). \end{aligned} \tag{13}$$

The larger value of $P_{k_{overflow}}^{t+T}$ means that queue overflow is more likely to occur and the corresponding user queue should have the higher priority of resource allocation, thus reducing the bit loss rate and providing QoS guarantee. This is why the proposed method jointly considers the queue priority and RB capacity.

Since $A_k(t)$ is an i.i.d process, $I_k(t)(t = 1, 2, \ldots)$ are also i.i.d random variables with a finite moment generating function $G(\theta) = E\{e^{\theta I_k(t)}\}$. According to $Cram\acute{e}r's$ Theorem [9], if $b_k < a_k$, the sequence $I_k(t)(t = 1, 2, \ldots)$ obeys the large deviation principle, and we have

$$\lim_{T\to\infty} \frac{1}{T} logP(\frac{\sum_{i=1}^{T} I_k(t+i)}{T} > a_k) = -l(a_k). \tag{14}$$

where

$$l(a_k) = sup_{\theta>0}\{a_k\theta - logG(\theta)\}. \tag{15}$$

and

$$logG(\theta) = log\{\sum_{i=-m_D}^{m_A} \pi_i^k e^{i\theta}\}. \tag{16}$$

For a sufficiently large value of T, according to (14) the overflow probability can be approximated by

$$P_{k_{overflow}}^{t+T} \approx e^{-Tl(a_k)}. \tag{17}$$

If we want to estimate the overflow probability according to (17), the values of a_k, b_k and π_i^k are required. It is easily to calculate a_k according to (11). But we can not derive analytical expressions about b_k and π_i^k directly due to the unknown about probability distribution of $I_k(t)$ in prior. However, we can use the mean of the historical observations to estimate the b_k and π_i^k.

3.2 User Priority Determination Algorithm

In the case of $\hat{c}_k \geq g_k$, it implies that if keeping the current queue configuration unchanged with the bit service rate $V_k(t)$, after T slots, the queue will be more likely to have an overflow situation. In this case, we use remaining life time $R_k(t)$ to rank the queue priority of resource allocation, which can be calculated by (4). However, in the case of $\hat{c}_k < g_k$, the queue overflow remains a rare event, and the queue overflow probability in the $(t+T)$ slot, $P_{k_{overflow}}^{t+T}$, can be approximated by (17). The buffer-aware priority determination algorithm determines a priority value for each user, where the user in the case of $\hat{c}_k \geq g_k$ is more emergent than in the case of $\hat{c}_k < g_k$. The smallest value of $R_k(t)$ indicates the highest priority of the kth user, the smallest value of $P_{k_{overflow}}^{t+T}$ indicates the lowest priority of the kth user. According to the user queues' different priorities, in the next section, we show how to dynamically allocate the RBs to adjust the service rate for each user queue.

4 Dynamic RBs Allocation algorithm

Suppose there are K user queues indexed by the set $\Phi = \{1, 2, \ldots K\}$ and N SBs indexed by the set $\Omega = \{1, 2, \ldots N\}$. The detail of the strategy is presented in Algorithm 1.

Algorithm 1 Dynamic RBs Allocation algorithm

```
while Φ ≠ ∅ or Ω ≠ ∅ do
    if the value of R_k(t) exists then
        Choose the user k_1 = arg min_{k∈Φ}{R_k(t)}
    else
        Choose the user k_1 = arg max_{k∈Φ}{P^{t+T}_{k_overflow}}
    end if
    Seek the SB n_1 = arg max_{n∈Ω}{γ_{n,k_1}} for the k_1th user
    while r^{n_1}_{k_1}(t) ≤ A_{k_1}(t) do
        Ω = Ω\{n_1}
        Seek the SB n_1 = arg max_{n∈Ω}{γ_{n,k_1}}
    end while
    after the data of user k_1 has been transmitted, set Φ = Φ\{k_1}
end while
```

5 Performance Evaluation

In this section, we characterize the performance of our adaptive resource allocation algorithm, and provide performance comparisons with three algorithms, namely PF metric(1) algorithm, PF metric(2) algorithm [4], CABA algorithm [5] and MaxWeight algorithm [12]. We first describe our simulation setup, and then the metrics used for performance evaluation are presented.

5.1 Experiment Setup and Performance Metrics

We simulated a multiuser scenario, where the maximum number of communicating users was set to $K = 10$. Here, the bit arrival rate for each user is assumed to obey the Poisson distribution with $\lambda > 0$. CQI is discretized into 15 levels, which results in 15 different pairs of modulation choice and code rate. This implies that there may be 15 possible transmission rates. A mapping between SINR ranges and CQIs is presented in [10]. The obtained CQIs are then used, together with the number of allocated RBs, to determine transmission rates that are used by the proposed optimization algorithm to improve the overall throughput.

To evaluate the performance of the proposed dynamic resource allocation, we define three metrics as follows:

- Average bit loss rate: This metric indicates QoS of K users. It is defined as time average bit loss rate during a period of $\varDelta$, i.e., $\bar{C}_k = \frac{1}{\varDelta+1}\sum_{t=T_0}^{T_0+\varDelta} \frac{B_k(t)}{A_k(t)}$, where $B_k(t)$ denotes the number of bit loss during the slot t for the kth user. Obviously, smaller $\bar{C}$ is preferred.
- Fairness: This metric is measured using *Jain's* fairness index [11], which is widely applied for evaluating the system fairness. It is described as $F(t) = \frac{(\sum_{k=1}^{K} D_k(t))^2}{K\sum_{k=1}^{K} D_k^2(t)}$, where $F(t)$ denotes the fairness at time t. Then, the system fairness can be calculated as $F = \frac{1}{\varDelta+1}\sum_{t=T_0}^{T_0+\varDelta} F(t)$.
- Average throughput: The larger average system throughput implies better performance.

5.2 Experimental Results

Performance Comparison for Different User Index We used Matlab for implementing the simulations. The corresponding simulation parameters are listed in **Table I**.

Table 1. SIMULATION PARAMETERS.

Parameter	Setting
Carrier Frequency	2GHz
System Bandwidth	10MHz
Transmission Time Interval	1ms
Subcarriers per Resource Block	12
Resource Block Bandwidth	180 KHz
Number of Resource Blocks	50
Type of System	Single Cell
Channel Model	Urban
Simulation Time	1000 TTIs

In Fig. 2, we show that the average BLR and the average throughput corresponding to 10 users for the five resource allocation schemes. The X axis denotes the user index. As we can see, in Fig. 2a, the proposed algorithm achieves better performance with average BLR of 2.12×10^{-3}, which is lower than those of PF metric(2) (about 2.13×10^{-3}), CABA (about 2.29×10^{-3}) and PF metric(1) (about 2.57×10^{-3}). In Fig. 2b, the average throughput for each user in the proposed algorithm significantly outperforms others except for MaxWeight-Alg. The reason for these is that, we calculate the priority of each user queue by using the remaining life time or queue overflow probability, and then allocate RBs dynamically. Consequently, it achieves a lower value of the average BLR, improves all users' service rate and keeps a high fairness among all users. By contrast, MaxWeight-Alg performs better than the other algorithms, but it does not consider the fairness, the curve of the average BLR and the average throughput for MaxWeight-Alg is unstable compared with other algorithms. For CABA, the empirical parameters in the priority function may influence the performance. PF metric(1) does not consider the queue length at all, which leads to the lowest performance. PF metric(2) suffers from the isolated RB assignment strategy, and thus it fails to obtain the achievable system throughput.

Performance at Different SINRs This section intends to investigate the effect of different SINR conditions, on the performance of the proposed algorithm and other three compared algorithms. The average SINR recorded varies from 11 dB to 20 dB with a step-size of 1 dB. The average BLR for all users is calculated by $\bar{C} = \frac{1}{K}\sum_{k=1}^{k=K}\bar{C}_k$.

As shown in Fig. 3, the average BLR, the fairness and the average throughput versus average SINR for four resource allocation schemes with 10 users were

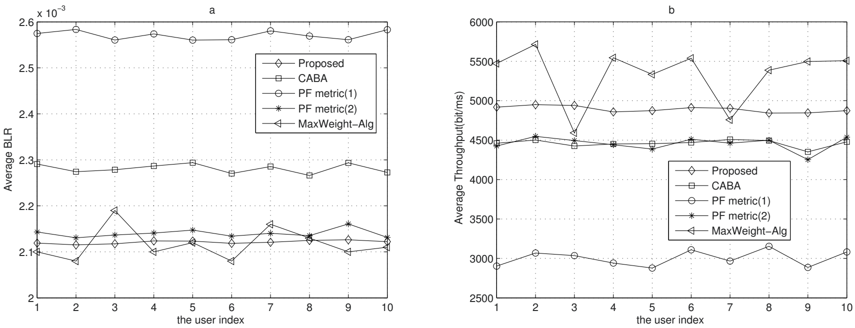


Fig. 2. The performance for different user index

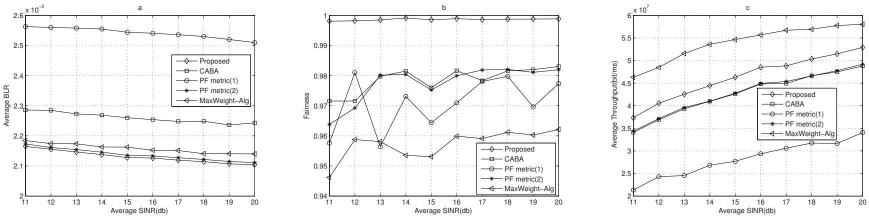


Fig. 3. The performance for different average channel SINR

plotted. In Fig. 3a, as the value of SINR increases, the average BLR decreases, and when the average SINR is low, the average BLR are high for all resource allocation algorithms. The proposed algorithm achieves the lowest BLR. This is because that at low SINR region, the bit service rate is not sufficient, and the current queue may have a shorter remaining life time or a larger overflow probability, which leads to a large number of bits lost. As the average SINR raises, the bit service rate is increasing. Thus, the remaining life time is prolonged as well as the overflow probability decreases, which may reduce the average BLR. While other algorithms fails to consider the priorities of user queues based on the buffer status and the RBs capacity. It is shown in Fig. 3b that the fairness index of the proposed algorithm is the highest among these algorithms, which is approximately 0.998. This indicates that the queue priority assists the proposed algorithm to balance the resource allocation among the users, thereby achieving superior fairness. Also we can see that the proposed algorithm may be insensitive to the changing of the average SINR, but the other four algorithms undergo a relatively large variation for the different average SINRs. This also implies that their performance is subject to the channel quality. As shown in Fig. 3c, the average SINR increases, the average system throughput is increasing. The results demonstrate that MaxWeight-Alg performs better than the other strategies in terms of the overall throughput, but it has a lowest fairness level as shown in Fig. 3b.

6 Conclusion

In this paper, we develop an adaptive resource allocation scheme in LTE downlink transmission systems by jointly considering user queue priority and the RBs capacity. We define the remaining life time of a queue, and its estimation model was presented. Also, we derived an analytical formula based on the large deviation principle invoked for estimating the overflow probability as a function of the buffer variance. We use the remaining life time and the queue overflow probability to determine the queue priority. According to the users' queue priority, an online measurement based algorithm was proposed to schedule RBs dynamically for adjusting the service rate of the user queues. Numerical results show that compared with traditional scheduling schemes, the proposed algorithm has a better tradeoff among throughput, QoS and fairness. It improves the total system throughput while guaranteeing certain fairness among users and reducing the average BLR.

References

1. A. Toskala, H. Holma, K. Pajukoski, and E. Tiirola: UTRAN Long Term Evolution in 3GPP. In Proc. IEEE PIMRC, Helsinki (2006) 1-5.
2. Y. Wang, G. Yue, S. Rangarajan, R. Sankar, and S. Morgera: Buffer-aware packet scheduling in downlink multiuser systems. In Proc. IEEE PIMRC, London (2013) 1495-1500.
3. C. Wang and X. Li: A buffer-aware resource allocation scheme for 4G LTE systems. In Proc. IEEE ISCE, Hsinchu (2013) 157-158.
4. S. Lee: Swap-based frequency-domain packet scheduling algorithm for small-queue condition in OFDMA. IEEE commun. letters, vol. 17. (2013) 1028–1031.
5. G. Y. Yan Lin: Channel-adapted and buffer-aware packet scheduling in LTE wireless communication system. In Proc. IEEE WICOM, Barcelona (2008) 239-243.
6. I.-F. Chao and C.-S. Chiou: An enhanced proportional fair scheduling algorithm to maximize QoS traffic in downlink OFDMA systems. In Proc. IEEE WCNC, Shanghai (2013) 239-243.
7. E. U. T. R. Access: Further advancements for E-UTRA physical layer aspects. 3GPP TR 36.814, Tech. Rep. (2010).
8. F. Capozzi, G. Piro, L. A. Grieco, G. Boggia, and P. Camarda: Downlink packet scheduling in LTE cellular networks: Key design issues and a survey. IEEE Commun. Surveys and Tutorials, vol. 15. no. 2. (2013) 678–700.
9. M. Mandjes: Large deviations for Gaussian queues: modelling communication networks. John Wiley & Sons, (2007).
10. E. U. T. R. Access: Physical layer procedures (release 10). mphTS, vol. 36. (2011).
11. A. Bin Sediq, R. H. Gohary, and H. Yanikomeroglu: Optimal tradeoff between efficiency and jain's fairness index in resource allocation. In Proc. IEEE PIMRC, Sydney (2012) 577-583.
12. Andrews, Matthew, and Lisa Zhang: Scheduling algorithms for multicarrier wireless data systems. IEEE Trans. Networking , vol.19. no. 2. (2011) 447-455.

Energy-Efficient Power Allocation for Decode-and-Forward OFDM Relay Links

Guangli Zhou[1], Tao Wang[2], Yating Wu[3], Guoxin Zheng[4], and Guangli Yang[5]

Key Laboratory of Specialty Fiber Optics and Optical Access Networks, Shanghai University, 149 Yanchang Road,Shanghai,China
[1]mugua@shu.ehu.cn,[2]twang@shu.edu.cn,[3]ytwu@shu.edu.cn,[4]gxzheng@staff.shu.edu.cn, [5]guangli.yang@shu.edu.cn

Abstract. This paper investigates the power allocation aiming to optimize the energy efficiency for an OFDM two-hop relay link based on DF protocol. First, an equivalent single-hop link is used to replace the two-hop link, and the model of the energy efficiency is established. When modeling the energy efficiency of the link, the circuit power consumption is not only taken into account, but also modeled as a linear increasing function of the rate. Subsequently, the energy efficiency optimization problem is solved by convex optimization, and the optimal power allocation is obtained under rate constraint. After that, the energy efficiency performance is also analyzed. Finally, the correctness of the theoretical analysis is verified by simulation results which adopt the spatial channel model conforming to the 3GPP standard, and the influence of circuit parameters on the performance of the system is also studied.

1 Introduction

Due to the sharp increase of the carbon emission and operating cost of wireless communication systems, energy efficiency(EE) has become a new design goal[1].For a communication system, its EE can be evaluated by either the total energy consumption for transmitting per message bit(TEPB),or the number of message bits transmitted with per-Joule total energy consumption(NBPE). A higher EE is represented by either a smaller TEPB or a greater NBPE.

Relaying was expected as an energy saving strategy by breaking a long distance transmission into several short distance transmissions. Cooperative relay transmission can expand the coverage of the network, improve the throughput and enhance the reliability of communication link. With the combination of relay and OFDM technology, optimal performance can be achieved via a reasonable allocation of subcarriers and power. In [2],under the premise of the optimal subcarrier pairing and the total rate constraint, the resource allocation algorithm is designed for the selective decode-and-forward(DF) relay OFDM link, which makes the total power consumption minimum. However, most of current studies don't take the circuit power consumption into account.

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_2

In practical systems, the energy is not only consumed by transmitting data, but also by various circuits and signaling. Nevertheless, early works studying the EE of communication systems only considered transmission energy but ignored circuit energy consumption. For instance, weighted sum power minimization problems were addressed for a source-relay-destination communication link subject to a minimum rate requirement in [3].For high-EE design of short-distance communication systems, the circuit energy consumption cannot be ignored, because the small transmission energy consumption is comparable to the circuit energy consumption[4]. Thus, the circuit energy must be taken into account especially for short-distance communication systems.

In view of the above fact, the circuit energy was taken into account to optimize the EE of communication systems in recent works. For instance, modulation schemes were optimized in [5] for communication links operating over flat-fading channels, and link adaptation algorithms were developed in [6]-[9] for multi-carrier systems transmitting over frequency-selective channels. In [6]-[12], the circuit power was assumed to remain fixed independently of the bit transmission rate. In[4] and [5], the circuit power was assumed to be linear with the transmission rate. In general, the circuit power is an increasing function of the transmission rate, since a greater bit rate indicates that a bigger codebook is used which usually incurs higher power for encoding and decoding on baseband circuit boards.

In this paper, we investigate the power allocation for high-EE in DF three-node OFDM relay link in order to optimize the system performance. The EE is evaluated by TEPB. Both the transmit power and various circuit power consumption are taken into account, where the circuit power is modeled as a linear increasing function of the bit transmission rate.

The rest of this paper is organized as follows. The system model is described in the next Section. After that, the power optimization function is formulated and the theoretical analysis is made in Section 3. The simulation results are shown in Section 4 to illustrate the obtained insight. Finally, we conclude the paper in Section 5.

2 System Models and Problem Formulation

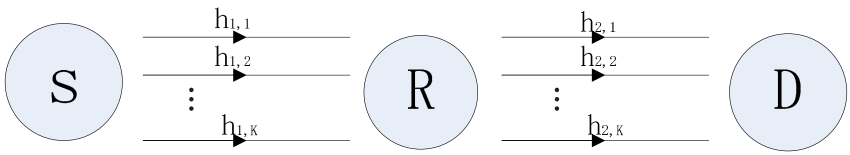


Fig. 1. Two-hop OFDM relay link

2.1 System Models

Consider a simple two-hop OFDM relay link with K orthogonal sub-carriers in each hop link, as shown in Figure 1.Each of the subcarriers can be approximated as flat-fading and quasi-static channel, so the channel gain of individual link is kept constant within each frame. Assume that the destination (D) is out of reach of the source(S) and therefore does not listen when the source is transmitting, such that direct link is ignored in our model, and the signal must be transmitted with the help of the relay(R).The relay is also supposed to be half-duplex and work in time division duplex (TDD) mode. Specifically, every data-transmission session takes two consecutive equal-duration time slots. During the first slot, the source emits signals to the relay node. The relay decodes and reencodes those message bits, then forwards the symbols(same as those emitted by the source) to the destination at every subchannel during the second slot. Actual circuit diagram of DF relay is shown in figure 2 in detail.

The channel coefficients of the kth subcarrier on the two hops are denoted by $h_{1,k}$ and $h_{2,k}$, respectively. Moreover, the power allocated for the kth subcarrier on the two hops is $P_{s,k}$ and $P_{r,k}$, respectively. Therefore, in the two time slots, the signals received by the relay node and the destination node are expressed as:

$$y_r = \sum_{k=1}^{K} \sqrt{P_{s,k}}\, h_{1,k} x_s + n_{sr}, \tag{1a}$$

$$y_d = \sum_{k=1}^{K} \sqrt{P_{r,k}}\, h_{2,k} x_r + n_{rd}, \tag{1b}$$

where x_s and x_r are the transmitted symbols by the source and relay node, respectively, and that n_{sr} and n_{rd} denote additive white Gauss noise (AWGN) with zero mean and variance σ^2 in the S-R and R-D link, respectively. Therefore, the received SNR for the kth subcarrier on the two hops can be given by $\Gamma_{1,k} = P_{s,k} g_{1,k}$ and $\Gamma_{2,k} = P_{r,k} g_{2,k}$, where $g_{1,k} = \frac{|h_{1,k}|^2}{\sigma^2}$ and $g_{2,k} = \frac{|h_{2,k}|^2}{\sigma^2}$, respectively.

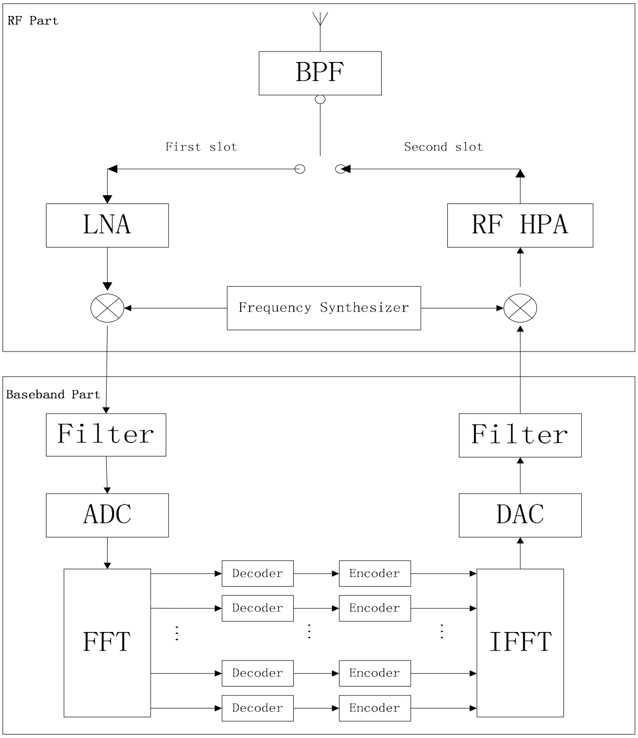


Fig. 2. Actual circuit diagram of DF relay

2.2 Power Allocation of a Single Link

Because the rate on the kth subcarrier depends on the small rate between the two hop links, the instantaneous rate of the kth subcarrier pair in the DF OFDM relay link is expressed as:

$$\begin{aligned} R_k &= \frac{1}{2}\log_2(1+\min\{\Gamma_{1,k},\Gamma_{2,k}\}) \\ &= \frac{1}{2}\log_2(1+\min\{P_{s,k}g_{1,k},P_{r,k}g_{2,k}\}), \end{aligned} \tag{2}$$

Given the instantaneous rate of the kth link r_k, the optimal power allocation of this link must be obtained under the condition of known rate in order to minimize the power consumption and to save energy of the communication system. Therefore, the problem can be formulated as(P1):

$$\min_{\{P_{s,k},P_{r,k}\}} \quad P_{s,k}+P_{r,k}$$
$$s.t. \quad \frac{1}{2}\log_2(1+\min\{P_{s,k}g_{1,k},P_{r,k}g_{2,k}\})=r_k, \tag{3}$$
$$P_{s,k},P_{r,k}\geq 0$$

We can get the optimal solution of (P1) if and only if $P_{s,k}g_{1,k}=P_{r,k}g_{2,k}=2^{2r_k}-1$, which indicates that the optimal power allocation of the kth link can be obtained when the rates of the two hops are the same. Assume that the optimal power of the kth link is P_k, namely $P_{s,k}+P_{r,k}=P_k$, then

$$P_k=\frac{(g_{1,k}+g_{2,k})}{g_{1,k}g_{2,k}}.(2^{2r_k}-1). \tag{4}$$

Let $G_k=\dfrac{g_{1,k}g_{2,k}}{g_{1,k}+g_{2,k}}$ denotes the equivalent channel gain. At this time, the two-hop relay link can be seen as a one-to-one paired single hop link[13],as shown in Figure 3.Then,the optimal power in the kth link can be expressed as:

$$P_k=\frac{2^{2r_k}-1}{G_k}. \tag{5}$$

Furthermore, the power allocation of each hop is

$$P_{s,k}=\frac{g_{2,k}}{g_{1,k}+g_{2,k}}P_k \ , P_{r,k}=\frac{g_{1,k}}{g_{1,k}+g_{2,k}}P_k, \tag{6}$$

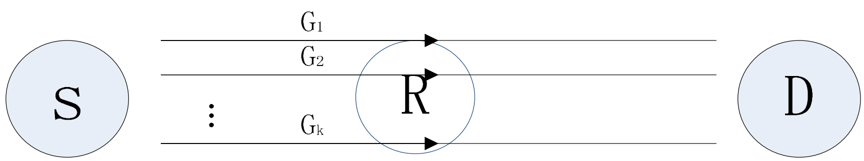


Fig. 3. Equivalent single hop link

2.3 EE Formulation

In last part,we analysis the power of a single link. Therefore, the total power of communication system can easily be expressed as:

$$P_{total} = \sum_{k=1}^{K} P_k + P_c$$
$$= \sum_{k=1}^{K} \frac{2^{2r_k} - 1}{G_k} + \beta \sum_{k=1}^{K} r_k + \alpha, \tag{7}$$

where P_c denotes the circuit power, including the relay and the base station circuit consumption. moreover, P_c is related to the rate instead of a simple constant, and can be modeled as linear increasing function of the rate [4][5], namely $P_c = \beta \sum_{k=1}^{K} r_k + \alpha$, where α, β are constants depending on the circuit properties.

In this paper, the EE of the system is defined as the total energy consumption for transmitting per message bit(TEPB).The smaller the EE is, the better the performance will be. We can express the EE as:

$$\eta(\mathbf{r}) = \frac{P_{total}}{R_{total}} = \frac{\sum_{k=1}^{K} \frac{2^{2r_k} - 1}{G_k} + \beta \sum_{k=1}^{K} r_k + \alpha}{\sum_{k=1}^{K} r_k} (J / bit), \tag{8}$$

where $\mathbf{r} = [r_1, r_2, ..., r_K]$ represents a set of link rates. In order to enable the system to achieve the best performance, the best EE must be obtained.

3 Power Allocation and Theoretical Analysis

3.1 Power Allocation Based on EE Optimization

As can be seen from the previous analysis, we can minimize the power consumption under the premise that the total rate is equal to a fixed value in order to reach the best EE. Therefore, we can draw the following problem(P2):

$$\begin{aligned} \min_{\{r_k\}} \quad & \sum_{k=1}^{K} \frac{2^{2r_k} - 1}{G_k} + \beta \sum_{k=1}^{K} r_k + \alpha \\ s.t. \quad & \sum_{k=1}^{K} r_k = R \\ & r_k \geq 0 \text{ , } k = 1,2..., K, \end{aligned} \tag{9}$$

We can see that (P2)is a convex optimization problem as the objective function P_{total} is a convex function, and the solution which subjects to the KKT conditions will be the global optimal solution[14].Using KKT conditions to solve the above problem, the optimal rate of each subcarrier is obtained:

$$r_k^{\ *} = [\frac{1}{2}\log_2(\frac{(\lambda^* - \beta)G_k}{2\ln 2})]^+, \tag{10}$$

where λ^* is the optimal Lagrange multiplier meeting all the constraints, and derived from $\sum_{k=1}^{K}[\frac{1}{2}\log_2(\frac{(\lambda^* - \beta)G_k}{2\ln 2})]^+ = R$ by bisection method. Then, the corresponding optimal power allocation is:

$$P_k^* = \frac{2^{2r_k^*} - 1}{G_k}, \tag{11}$$

Return to original two-hop relay link, the power allocation for every subcarrier in each slot can be expressed as:

$$P_{s,k} = \frac{g_{2,k}}{g_{1,k} + g_{2,k}} P_k^* \ , P_{r,k} = \frac{g_{1,k}}{g_{1,k} + g_{2,k}} P_k^* . \tag{12}$$

3.2 Theoretical Analysis

Lemma 1: Total power $P_{total}(R)$ is a convex function of the total rate R.

Proof :Let $P_{total} = f(r)$, $\sum_{k=1}^{K} r_k = g(r)$, $R = t$,then the original problem (P2) can be expressed as:

$$\begin{aligned} &\min_r \quad f(r) \\ &s.t. \quad\ g(r) = t, \\ &\qquad\quad r \geq 0 \end{aligned} \tag{13}$$

Assume r_{t_1} is the independent variable value of $\min f(r)$ when $g(r) = t_1$, while r_{t_2} is the independent variable value of $\min f(r)$ when $g(r) = t_2$.

To this end, the following just needed to be proved:

$$\theta P_{total}(t_1) + (1-\theta)P_{total}(t_2) \geq P_{total}(\theta t_1 + (1-\theta)t_2), \theta \in (0,1),$$

that is to prove: $\theta f(r_{t_1}) + (1-\theta)f(r_{t_2}) \geq f(r_{\theta t_1 + (1-\theta)t_2}), \theta \in (0,1)$.

Due to the fact that P_{total} is convex function, $f(r)$ is a convex function with r. Therefore, $\theta f(r_{t_1}) + (1-\theta)f(r_{t_2}) \geq f(\theta r_{t_1} + (1-\theta)r_{t_2})$ holds.

Because the constraint condition of the original problem is linear, then $g(\theta r_{t_1} + (1-\theta)r_{t_2}) = \theta g(r_{t_1}) + (1-\theta)g(r_{t_2})$ .Moreover, $g(r_{t_1}) = t_1, g(r_{t_2}) = t_2$, so $g(\theta r_{t_1} + (1-\theta)r_{t_2}) = \theta t_1 + (1-\theta)t_2$ can be obtained. We can see that $\theta r_{t_1} + (1-\theta)r_{t_2}$ is the feasible solution of $f(r)$ when $t=\theta t_1+(1-\theta)t_2$ while $f(r_{\theta t_1+(1-\theta)t_2})$ is the minimum of $f(r)$ .Thus, $f(\theta r_{t_1}+(1-\theta)r_{t_2})\geq f(r_{\theta t_1+(1-\theta)t_2})$ holds. Therefore, $\theta f(r_{t_1})+(1-\theta)f(r_{t_2})\geq f(r_{\theta t_1+(1-\theta)t_2}),\theta\in(0,1)$, meaning that the total power $P_{total}(R)$ is a convex function of the total rate R.

Lemma 2: EE η is a quasi-convex function of the total rate R.

According to Lemma 1, we can express the EE as follows:

$$\eta(R)=\frac{P_{total}}{R_{total}}=\frac{P_{total}(R)}{R}. \tag{14}$$

Proof: $\eta(R)$ is strictly quasi-convex of R if $S_\varepsilon=\{R\in dom\,\eta(R)\,|\,\eta(R)\leq\varepsilon\}$ is a strictly convex set for any real value ε .

Note that $\eta(\mathbf{r})\geq 0$, which means that S_ε is empty if $\varepsilon\leq 0$, hence S_ε is strictly convex since no point lies on the contour of S_ε .

When $\varepsilon>0$, $S_\varepsilon=\{R>0\,|\,f(R,\varepsilon)=P_{total}(R)-\varepsilon R\leq 0\}$,where $f(R,\varepsilon)$ is a strictly convex function of R. Suppose R_1 and R_2 are any two points on the contour of S_ε ,and also $R_1<R_2$ .We can see from the strict convexity of $f(R,\varepsilon)$ with respect to R that $f(R,\varepsilon)\leq\max\{f(R_1,\varepsilon),f(R_2,\varepsilon)\}\leq 0$ holds for $\forall R\in(R_1,R_2)$, which means that any R between any two points on the contour of S_ε must lie in the interior of S_ε .Thus, S_ε is a strictly convex set when $\varepsilon>0$ .Therefore, η is strictly quasi-convex of R.

From the previous proof, we denote the minimum TEPB as η^* ,i.e., $\eta^*=\min_{R\geq 0}\eta(R)$,and the corresponding optimum rate as R^* ,i.e., $R^*=\arg\min_{R\geq 0}\eta(R)$ .According to the strict quasi-convexity of $\eta(R)$,the following properties are satisfied:

Lemma 3: 1) There exits a unique R^* and it satisfies

$$\frac{\partial\eta(R^*)}{\partial R}=0;$$

2) $\eta(R)$ is strictly decreasing with $R\in(0,R^*]$ and

$$\forall R\in(0,R^*],\frac{\partial\eta(R)}{\partial R}<0;$$

3) $\eta(R)$ is strictly increasing with $R\in[R^*,+\infty)$ and

$$\forall R\in[R^*,+\infty),\frac{\partial\eta(R)}{\partial R}>0.$$

Proof: Suppose there exits R_1 and R_2 satisfying $R_1<R_2$, and $\eta(R_1)=\eta(R_2)=\eta^*$. Because of the strict quasi-convexity of $\eta(R)$, $\eta(R)\leq\max\{\eta(R_1),\eta(R_2)\}=\eta^*$ holds for $\forall R\in(R_1,R_2]$,leading to a contradiction with $\eta^*=\min_{R\geq 0}\eta(R)$ .Therefore, there must

exit a unique R^* satisfying $\eta(R^*)=\eta^*$. Moreover, R^* must satisfy $\forall R \geq 0, \frac{\partial \eta(R^*)}{\partial R}(R-R^*) \geq 0$ [15].As said earlier, $R^*>0$ must hold, thus $\frac{\partial \eta(R^*)}{\partial R}=0$ must hold to satisfy the condition. This proves the first claim.

For any R_1 and R_2 satisfying $0<R_1<R_2<R^*$, $\eta(R_2)<\max\{\eta(R_1),\eta(R^*)\}=\eta(R_1)$ follows the strict quasi-convexity. This means that $\eta(R)$ is a strictly decreasing with $R \in (0,R^*]$.Therefore, $\frac{\partial \eta(R)}{\partial R}<0$ must hold for $\forall R \in (0,R^*]$.This proves the second claim.

Likewise, we can prove the third claim easily.

4 Simulation Results

In this section, we will present simulation results to verify the correctness of the properties proposed in last section via Matlab. In this paper, we adopt the spatial channel model (SCM) conforming to the 3GPP standard model[16] in order to simulate the practical communication scenarios. The distance of the source-relay and the relay-destination are fixed to 500m, respectively, and the parameters of the circuit power are set as $\alpha=0, \beta=1$.The other system parameters are listed in Table 1.

Table 1. Main simulation parameters

System parameters	Values
Number of subcarriers	100
Noise power σ^2	10^(-4)
Number of paths	6
Number of subpaths per path	20
Carrier frequency	2e9 Hz
Number of time samples	100
Height of base station	32m
Height of relay station	32m
Height of user station	1.5m
SCM scenario	Urban_macro

4.1 Relationship between Total Power and Total Rate Limiting

We randomly generate 100 sets of channels to verify the correctness of Lemma 1, and record the variation of the total power when total rate increases from zero. Then, the curve of the relationship between system power and total rate limiting is plotted as shown in Figure 4 by averaging the data recorded above. We can see from Figure 4 that the total power increases with rate increasing, meeting the characteristics of convex function, which verifies Lemma 1.It is not difficult to understand that the rate allocated to each subcarrier becomes larger when the total rate increases. Moreover, $P_k = \frac{2^{2r_k}-1}{G_k}$ and then the overall system power consumption increases.

4.2 Relationship between EE and Total Rate Limiting

The relationship between system energy efficiency and total rate under different channel information (CSI) is analyzed in Figure 5.We verify the correctness of Lemma 2 and Lemma 3 by randomly selecting three channel conditions to simulate. The results illustrate that EE decreases first and then increases with the increase of the rate under given CSI. Moreover, there is only one specific rate corresponding to minimum EE. In addition, the similar trend of EE exits under different channel conditions, which shows that EE is a quasi-convex function with total rate. Therefore, the rate must be chosen reasonably in practical communication system in order to reach the best EE.

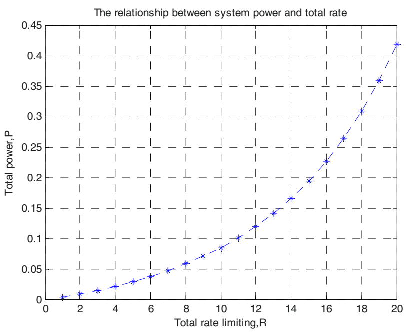


Fig. 4. The relationship between system power and total rate

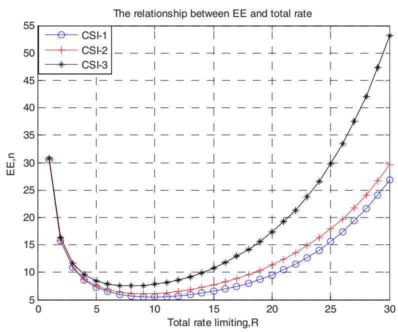


Fig. 5. The relationship between EE and total rate under different CSI

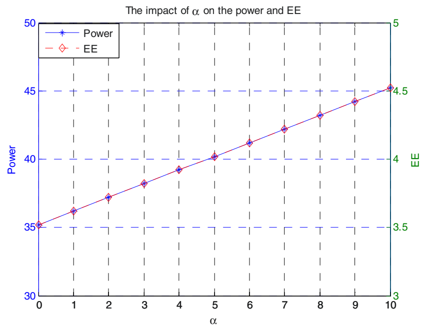


Fig. 6 The impact of α on the system when $\beta = 1, R = 10$

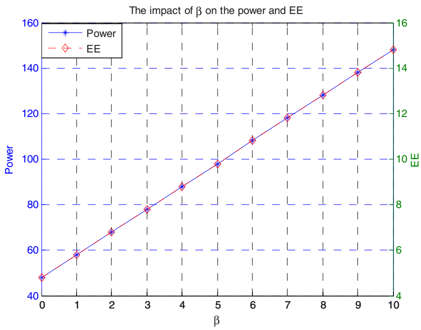


Fig. 7 The impact of β on the system when $\alpha = 0, R = 10$

4.3 Impact of Circuit Parameters on the System

We have done simulation experiment above setting the circuit parameters as fixed value. However, as we can see from (7) and (8) that the circuit parameter is also one of the factors affecting the system performance. The following will study the influence of circuit parameters on communication system:

In order to analyze the influence of circuit parameters on the system, the total rate is fixed to R=10, the results are shown in Figure 6 and Figure 7.It can be seen that the total power and EE of the system increase linearly with one of parameters increasing from zero when the other parameter is fixed. Moreover, the growth rate of the power is ten times as much as that of EE. There are reasons for this phenomenon: as can be seen from (7), the system power is linear with circuit parameters α, β under the given rate, respectively. What's more, EE is only the ratio of power with a fixed rate, meaning that EE is also a linear increasing function. In addition, the fixed rate is exactly the ratio of them.

5 Conclusion

In this paper, we have studied the power allocation in three-node OFDM two-hop relay link based on DF protocol in order to minimize EE. We use an equivalent single hop link to replace the two hop link, and model the circuit power as a linear increasing function of the total rate. Subsequently, optimization problem is solved by convex optimization, and the optimal power allocation is obtained under rate constraint. Finally, the theoretical analysis is verified by simulation results.

In the actual system, the real efficiency of the power amplifier is not a constant. Therefore, in the future work, we will consider the situation and continue to study it in depth.

Acknowledgments. This work was supported by NSFC(61401266,61571282),ChenGuang project and Innovation Program(14ZZ096) by Shanghai Municipal Education Commission and Education Development Foundation, Specialized Research Fund for the Doctoral Program of Higher Education of China(20133108120015),and NSF of Shanghai(14ZR1415100).

References

[1]Li G Y, Xu Z, Xiong C, et al. Energy-efficient wireless communications: tutorial, survey, and open issues.[J]. IEEE Wireless Communications, 2011, 18(6):28-35.
[2]Wang T, Fang Y, Vandendorpe L. Power minimization for OFDM Transmission with Subcarrier-pair based Opportunistic DF Relaying[J]. Communications Letters IEEE, 2013, 17(3):471 - 474.
[3]Wang T. Weighted sum power minimisation for multichannel decode-and-forward relaying[J]. Electronics Letters, 2012, 48(7):410-411.
[4]Isheden C, Fettweis G P. Energy-Efficient Multi-Carrier Link Adaptation with Sum Rate-Dependent Circuit Power[C]// Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE. IEEE, 2010:1-6.
[5]Cui S, Goldsmith A J, Bahai A. Energy-constrained modulation optimization[J]. Wireless Communications IEEE Transactions on, 2005, 4(5):2349-2360.
[6]Miao G, Himayat N, Li G Y. Energy-efficient link adaptation in frequency-selective channels[J]. Communications IEEE Transactions on, 2010, 58(2):545-554.
[7]Xiong C, Li G Y, Zhang S, et al. Energy-Efficient Resource Allocation in OFDMA Networks[J]. IEEE Transactions on Communications, 2011, 60(12):1-5.
[8]Sun C, Cen Y, Yang C. Energy efficient OFDM relay systems[J]. IEEE Transactions on Communications, 2013, 61(5):1797-1809.
[9]Miao G, Himayat N, Li G Y, et al. Interference-Aware Energy-Efficient Power Optimization[J]. IEEE Transactions on Wireless Communications, 2011, 10(4):1-5.
[10]Chinaei M H, Omidi M J, Kazemi J. Circuit power considered energy efficiency in Decode-and-Forward relaying[C]// Electrical Engineering (ICEE), 2013 21st Iranian Conference on. IEEE, 2013:1-5.
[11]Wu Y, Wang T. Energy-efficient resource allocation for OFDM transmission with opportunistic DF relaying[C]// Communications in China (ICCC), 2014 IEEE/CIC International Conference on. IEEE, 2014:570-575.
[12]Huang G, Fang X, Yu C, et al. Optimal Energy-Efficient Path Selection Scheme in OFDM-based DF Relay Networks[J]. Xinan Jiaotong Daxue Xuebao/journal of Southwest Jiaotong University, 2015, 50(1):58-65.
[13]Li Y, Wang W, Kong J, et al. Subcarrier pairing for amplify-and-forward and decode-and-forward OFDM relay links[J]. Communications Letters IEEE, 2009, 13(4):209-211
[14]Boyd S, Vandenberghe L. Convex Optimization[M]. Cambridge University Press, 2004.
[15]Bertsekas D P. Nonlinear Programming[J]. Journal of the Operational Research Society, 1995, 4(1):67-110.
[16]3GPP,TR 25.996,V9.0.0,http://www.3gpp.org

Voice over LTE Performance Evaluation

Ahmed Amou El Arby and Ousmane Thiare

University Gaston Berger
Saint Louis, Senegal
Email: arby.amou@gmail.com
Email: othiare@ugb.edu.sn

Abstract. The 4G technology, known as Long Term Evolution (LTE), has been dominating the telecommunication market the last years offering a high data rate (up to 100Mbps) that will perfectly satisfy the users multimedia applications needs in term of bandwidth. Designed to be a fully packet switched network, LTE lack in offering the voice CS service. This paper propose an analytical model for simulating the performance of Voice over LTE (VoLTE) while presenting an overview of the latest up to date solutions for the voice over LTE such as Circuit Switched Fall Back, Voice over LTE via Generic Access and third party services for providing call continuity over LTE.

Keywords: LTE, SRVCC, VoLGA, CSFB

1 Introduction

During the past years, voice service was the primary revenue generator for cellular operators, though any migration in the network must support the transmission of the main service. Legacy network such as 2G/3G was basically relying on the Circuit Switched (CS) domain to offer the voice service and on the Packet Switched (PS) for transmitting data. Contrary to the previous cellular generations, LTE was designed to be All-IP network offering only the services based on the packet switching domain leaving the operators with a gap when it comes to serve the users who still have the 2G/3G phones.

Many solutions have been proposed to support the voice CS over LTE to ensure the service continuity. However, only one of them has been selected as a long term solution by 3GPP. This paper start by presenting an overview of the various up to date solutions in the telecommunication market that permit the voice over LTE such as the Circuit Switched Fallback (CSFB) [1]-[2], Voice over LTE via Generic Access (VoLGA) [3], Third Party Voice Over IP[6] and the Single Radio Voice Call Continuity (SRVCC) [4],[5]. Then, the paper also focuses on the long term solution SRVCC and proposes an analytical model for simulating the performance of SRVCC.

The rest of the paper is structured as follow: In Section II, we review the short-term solutions in LTE. Section III presents the SRVCC handover. In section IV, we

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_3

propose an analytical model that represents the handover delay interruption in SRVCC. Finally, we conclude the paper in Section VI.

2 Voice Over LTE short term solutions

2.1 Circuit Switched FallBack (CSFB)

The Circuit Switch Fallback is a short term solution that relies on the existing 2G/3G network. When a UE, camped on LTE, wants to make a CS call, it activate an Extended Service Request (ESR) to the MME. After receiving the ESR, the MME order the eNodeB to redirect the call to the legacy network 2G/3G. The UE receives the LTE Release Connection and camp on the legacy 2G/3G network and establishes the standard 2G/3G RRC Connection. The downside of this solution is the introduction of the delay due to the procedure execution. This delay varies based on the mobile originating the call, in this case the delay is around 4.7s [1], or terminating the call, in this case the delay is around 2.84s [1].

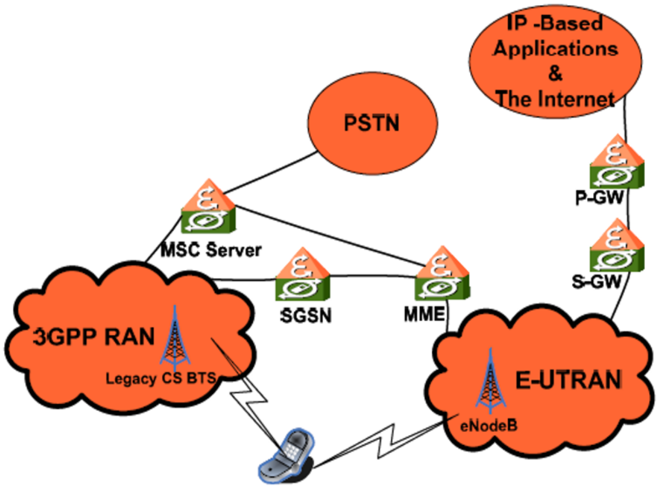


Fig. 1. CSFB network Architecture

Furthermore, the CSFB requires an introduction to new interfaces between LTE and legacy 2G/3G particularly the MME-MSC/SGSN and S-GW-SGSN [1] in order to perform the combined registration and paging process.

The benefit of using the CSFB is that the operators can use their existing 2/3G infrastructure to provide voice call.

2.2 The Voice Over LTE via Generic Acess (VoLGA)

The Voice over LTE via Generic Access (VoLGA) consists of having a second interfaces for Wi-Fi on the UE becoming dual mode devices that can connect to Wi-Fi over a 2/3G network when it is available and connects to the network operator through a GAN gateway. Similarly, the VoLGA use the LTE as a replacement of the Wi-Fi and introduce an entity named Access Network Controller (VANC) which is a mix between an IP based node and a base station controller node from the LTE and GSM/UMTS point of view respectively. The VANC connect to the 2G, 3Gand 4G via the same interfaces as per the figure 2.

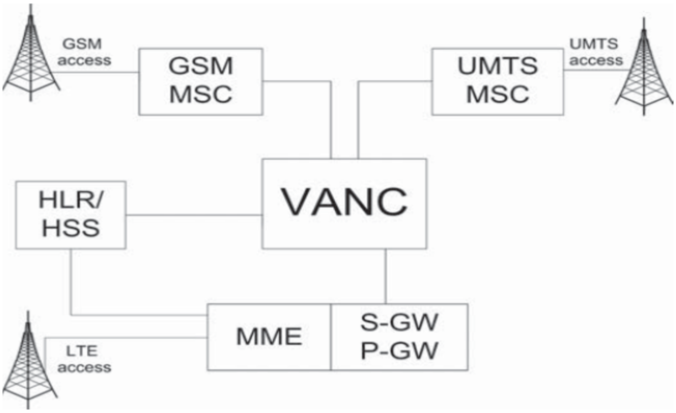


Fig. 2. VoLGA Network Architecture

The mobile start by registering to the LTE and then establishes a secure IPSec tunnel connection to the VANC over LTE [3]. Then, it registers to the 2G/3G through the VANC secure connection that has just been established. The translation between the PS LTE and CS legacy is performed into the VANC entity and is completely transparent to the user and the network.

The downside of VoLGA is that it has not been adopted by 3GPP and still requires all visited networks to support VoLGA in order to provide the roaming capability.

2.3 Third Party Software For VoLTE

Another solution that permits voice over LTE is enabled by using a third party software such as Viber or Skype, for providing calls over the LTE as presented in the figure3. No changes are required in the network [6], which makes the solution the least expensive among others VoLTE solutions. The downside of this solution is that there is no quality of service guaranteed and this solution doesn't support the CS calls.

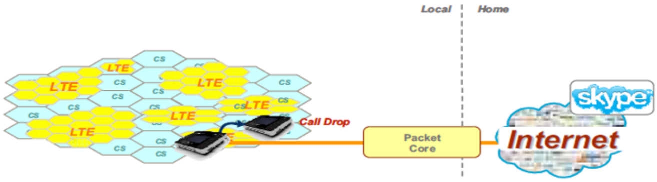


Fig. 3. Voice Over LTE via Skype

3 SINGLE RADIO VOICE CALL CONTINUITY

The Single Radio Voice Call Continuity (SRVCC) is the 3GPP target solution, based on IMS, for ensuring the voice continuity between LTE and 2/3G networks. The SRVCC requires that the UE initiates the call using IMS and the Application Server that needs to be implemented in the IMS [4].

The SRVCC uses a combination of the IMS session continuity procedure with a handover procedure. In detail, when a UE reaches the border of the LTE coverage, the E-UTRAN will send a SRVCC handover request to the MME, the latest will trigger the SRVCC procedure with the enhanced MSC (eMSC). The eMSC will perform the session transfer to the IMS and coordinates it with the CS handover to the target cell. Finally the eMSC will send the Forward Relocation Response to MME, which includes the handover command. The figure 4 and 5 presents the architecture and call flow of the SRVCC [4], [5].

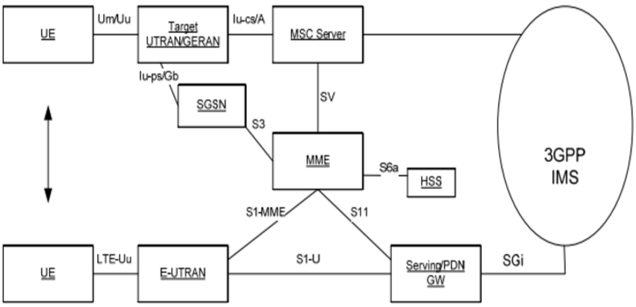


Fig. 4. SRVCC Architecture for E-UTRAN to UTRAN/GERAN

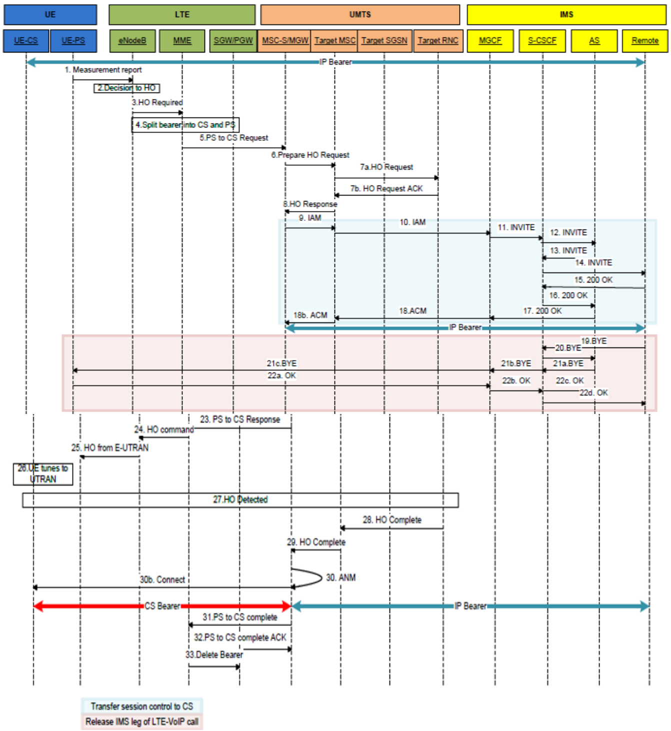


Fig. 5. SRVCC call flow from E-UTRAN to UTRAN/GERAN

4 Analytical Model for SRVCC

Many papers have studied the SRVCC handover and tried to come out with a standard model to simulate the performance of this long term solution for voice over LTE.

One of the mathematical models that have been developed is the model presented in [8], in which an analytical expression of the overall delay experienced by the call executing the handover is been presented. This delay was based on the sum of the delay caused by the radio link (RLC) and the delay caused by the remote and internet queuing which was assumed to be following the First In First Out queuing algorithm.

The results of [8] was more related to the radio link delay and has shown that the SRVCC handover interruption time will be reduce under a given Block Error Rate (BLER) with an increased data rate as presented in the figure6.

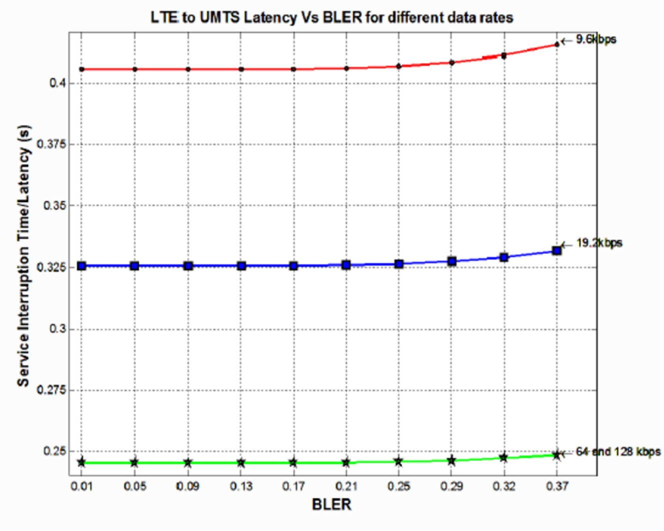


Fig. 6. LTE to UMTS SRVCC service interruption time Vs BLER

In this paper, we will be reevaluating the model while assuming that the Quality of Service is applied in the whole network since it is crucial for LTE to implement the QoS in order to serve fairly the different traffics that flow through the network.

We estimate that applying a more advanced queuing algorithm will minimize the delay of the SRVCC packets in the network nodes and there for reduce considerably the overall delay experienced by SRVCC handover which will enhance the performance of the handover.

In order to evaluate the performance of the SRVCC under priority queuing, it is necessary to analyze the interruption time experienced by the ongoing call. This delay, and based on the [9], can be split into three parts: Radio link delay, the network queuing delay and the remote network queuing delay. These delays can be represented by be mathematical models since they are following a unique mathematical behavior.

1. The radio link delay:

Based on several papers, the radio delay can be measured by analyzing the Radio Link Control (RLC). Assuming the same RLC model in LTE as UMTS, we can represent the delay in the radio part by the following equation1 [8]:

$$T_{RLC} = T_{Iub} + (k-1)TTI + \frac{k(P_s - (1-p))}{P_s^2} *$$
$$\left\{ \sum_{j}^{n} \sum_{i}^{j} \left[P(C_{ij}) \left(2jT_{Iub} + \left(\frac{j(j+1)}{2} + i \right) * TTI \right) \right] \right\}$$

Equation1: Radio Delay

Where:

K : number of frames to be transmitted
N: number of RLC retransmissions
Ps: probability of receiving RLC frame successfully after n transmission
P: probability of RLC frame received erroneously
TIub : latency of the Iub interface
TTI: transmission time interval at eNodeB
P(Cij): the first correctly received frame at destination

2. The network node and remote queuing delay:

The delay caused by the nodes queues relies heavily on the algorithm applied in the system. Many algorithms has been developed and globally used such as FIFO, Priority Queuing (PQ) and Weight Round Robin PQ.

In this research and since LTE is an ALL-IP network that must apply queuing algorithm that guarantee the Quality of Service (QoS) in the network in delivering the traffic, we, first, assume that PQ algorithm has been applied in order to priorize packet coming from the handover process and therefore reducing their treatment delay in the queue.

In order to estimate the average waiting time (delay) in the queues for both the network queuing delay and for the remote queuing delay, we assume that the packets priority is p and arriving with a Poisson distribution with parameter λ_p and their length has an exponential distribution with an average service time $\overline{x_p}$.

We can then calculate the average waiting time W_p of priority p packets in the following way [10]:

$$W_p = \frac{W_0}{(1-\sigma_p)(1-\sigma_{p+1})}$$

Equation2: Average waiting time for PQ

Where W_0 is the average delay caused by a packet, which is already serviced and calculated as in:

$$W_0 = \sum_{i=1}^{P} \frac{\lambda_i \overline{x^2}}{2}$$

Where $\overline{x^2}$ is the second moment of service time.

And

$$\sigma_p = \sum_{i=p}^{P} \rho_i = \sum_{i=p}^{P} \lambda_i \overline{x_i}$$

After identifying the delay in each part of the involved part of the network, we will deduct the expression of the whole interruption delay experienced by a user that executes the SRVCC handover.

Based on the signal flow in the SRVCC in the figure5, we can notice that the service interruption start when the call is released in order to establish a new connection with the 3G network. The figure7 [8] represent a zoomed part of the signal flow during the interruption of the service.

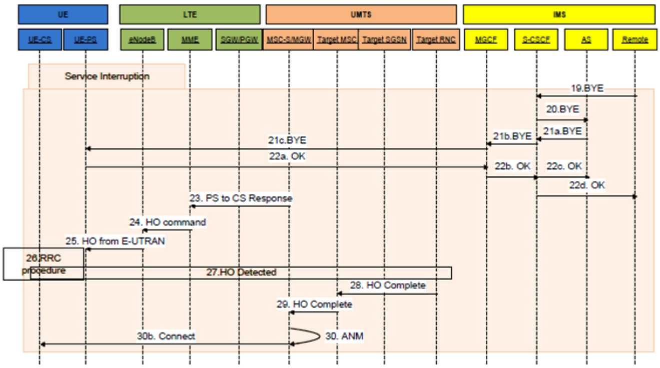


Fig. 7. Service Interruption

Based on the figure and message flow in each node and radio link, we can write the equation that represent the overall interruption delay as follow:

$$Dealy_{SRVCC} = Delay_{Remote} + Delay_{Internet} + 2 * Delay_{AS} + 3Delay_{CSCF} + 2Delay_{MGCF} + Delay_{UE} + 3Delay_{MSC} + Delay_{MME} + Delay_{eNodeB} + Delay_{RRC} + Dleay_{RNC}$$

Equation3: Overall SRVCC interruption time

Each value of the equation3 will be replaced by its expression that has already been presented in the section1 and 2. The equation3 will be used to estimate the delay of the interruption of service of the ongoing call until the SRVCC handover has reestablish the circuit with the new cell. The delay will be evaluated under different parameters values in order to study the behavior of the SRVCC performance.

The simulation results and performances evaluation for different priority queuing will be presented in separate paper in the near future.

5 Conclusion

Providing voice calls in LTE has been a challenge and a necessity for the operators since LTE is designed for only PS and the significant incomes generated by the voice calls. Many temporary solutions has been proposed and used, however the operators still seeking for implementing the long term solution SRVCC that will enable the transmission of voice over LTE and supporting the CS over LTE. This solution has not been studied in depth in order to evaluate it performance. This paper has presented the different temporarily solution that are up to date and has focused on the long term solution while proposing a new analytical model that could be used to evaluate the performance of SRVCC. This model will be simulated via Matlab and that for different priority queuing in order to evaluate the performance of SRVCC under these conditions.

References

[1] J. E. Vargas Bautista, S. Sawhney, M. Shukair, I. Singh, V. K. Govindaraju, and S. Sarkar, "Performance of Fallback from LTE to UMTS" Qualcomm Corporate Engineering. IEEE Communications Magazine, Vol. 51, Issue: 9, pp. 136-143, 2013.

[2] CS fallback in Evolved Packet System; Technical Specification 3GPP TS 23.272; 2010.

[3] Voice over LTE with Generic Access, Technical Specification VoLGA Stage 2; 2010.

[4] Single Radio Voice Call Continuity; Technical Specification 3GPP TS 23.216; 2008.

[5] Single Radio Voice Call Continuity; Technical Specification 3GPP TR 23.856; 2010.

[6] S.Gavrilovic, "Standard Based Solutions for Voice SMS Services over LTE"; IEEE/MIPRO; pp. 334-339, 2010.

[7] Cristopher Cox, "Introduction to LTE", Wiley, 2012.

[8] Namakoye, J.; Van Olst, R "A Performance evaluation of a voice call handover scheme between LTE and UMTS" IEEE Conference Publications, 2011

[9] Kim, Kyungmin. "A seamless voice call handover scheme for next generation cellular network". Shanghai : IEEE, 2009. IEEE 15th Asia-Pacific Conference on Communications, 2009.

[10] Tomáš Balogh, Martin Medvecký, "Comparison of Priority Queuing Based Scheduling Algorithms" ElektroRevue, 2010

[11] Enderle N., Lagrange X. "Radio link control-acknowledged mode protocol performance modeling in UMTS". IEEE, 4th International Workshop on Mobile and Wireless Communications Network, 2002

Receiver-Based Adaptive Signal Control for Enhancing VoIP Speech Quality

Hye-Seung Cho[1], Qing-Fu Han[1], Hyoung-Gook Kim[1], Jin Young Kim[2]

[1] Kwangwoon University, Seoul, Rep. of Korea
[2] Chonnam National University, Gwangju, Rep. of Korea
{hye_seung401,qingfu33,hkim}@kw.ac.kr,
beyondi@chonnam.ac.kr

Abstract. This paper proposes a receiver-based high performance adaptive signal control for enhancing Voice over Internet Protocol (VoIP) speech quality. In the proposed method, the buffering time is minimized by way of playing out normally, expanding or compressing each packet according to adaptive network jitter estimation. And recursive linear prediction-based packet loss concealment using an adaptive muting factor delivers high voice quality by concealing consecutive packet loss. Experimental results show that the proposed algorithm delivers high voice quality by pursuing an optimal trade-off between buffering delay and packet loss rate.

1 Introduction

VoIP uses packetized transmission of speech over the Internet. However, a number of factors, such as packet loss, packet delay, and network delay variation [1], [2], degrades the quality of speech at the receiving end of a voice transmission system in an IP network. For concealing packet loss or recovering lost packets, several methods including waveform substitution, pattern matching and model-based repetition have been developed. To reduce the influence of the jitter, a playout buffer can be employed by the receiver to hold a VoIP packet until its scheduled playout time. In adaptive playout, the buffer size can be minimized by expanding or compressing each packet according to the predicted network delay and variations [3].

In this paper, we focus on enhancing VoIP speech quality only at receiving portion of a mobile Internet phone. The important functionality to be implemented at the receiver is an adaptive signal control with two main functions, such as playout scheduling (PS) and packet loss concealment (PLC).

2 Proposed Adaptive Signal Control for Enhancing VoIP Speech Quality at the Receiver

The receiving system employs adaptive signal control containing packet loss concealment and playout scheduling for enhancing voice quality over IP networks.

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_4

2.1 Packet Loss Concealment Based on Recursive Linear Prediction

On the receiver side of mobile VoIP, when an i^{th} packet arrives at the receiver, the receiver strips the packet information and places the packet in the jitter buffer. In the jitter buffer, the network jitter is adaptively estimated and used to assign each packet a scheduled playout time. Each signal frame decoded from the jitter buffer is stored in a signal frame buffer and is classified into one of the three classes: voiced, unvoiced, and silence or background types. Our PLC algorithm is based on recursive linear prediction analysis and synthesis (LPAS) using soft estimated pitch period to improve the G.722 Appendix IV PLC algorithm [4].

If the number of lost packets is smaller than δ (δ ≥ 1), the short PLC generates synthesized speech using the excitation signal repetition of the previous received speech with a periodicity equal to the soft pitch period.

For example, if the $(i+1)^{th}$ future signal frame is not found in the signal frame buffer and in the jitter buffer for reconstruction of lost packets, the concealment algorithm works only on the $(i-1)^{th}$ previous signal frame stored in the history buffer. The i^{th} lost signal frame is substituted by direct repetition of the LPAS-based reconstructed signal.

If the $(i+1)^{th}$ future signal frame is available, the concealment algorithm uses the $(i+1)^{th}$ future packet information and the $(i-1)^{th}$ past signal frame information from the history buffer. Signal frame R is generated by LPAS of the $(i+1)^{th}$ future signal frame, while LPAS of the $(i-1)^{th}$ previous signal frame produces the signal frame T.

The refined estimated signal frame U for substituting the lost signal frame results in a smooth interpolation using peak alignment overlap-add (PAOLA) [5] between the signal frame R and the signal frame T. period.

In the consecutive PLC, the reconstructed signal frames for recovering the missing packets are generated using smooth excitation signal repetition based on the recursive LPAS with a periodicity equal to the soft pitch period. The smooth excitation signal is constructed frame by frame as the natural periodic extension.

The reconstructed signal is gradually muted for the duration of the loss period to alleviate the metallic artifacts and improve the voice quality. The adaptive muting factor using a sigmoid function reduces the error between the desired signal and the reconstructed signal by using the optimal sloping parameters of the sigmoid function when packet losses occur.

2.2 Adaptive Playout Scheduling Using Jitter Estimation

To play each packet at the scheduled time, our algorithm tries to manage the effect of transmission jitter

First, current buffering delay $b_{i,k}$ is calculated by first-order recursive averaging given by

$$b_{i,k} = \alpha_b \cdot b_{i-1,k} + (1-\alpha_b)\cdot(p_{i,k} - a_{i,k}) \tag{1}$$

where α_b ($0 < \alpha_b < 1$) is a smoothing parameter and $p_{i,k}$ and $a_{i,k}$ are the playout time and arrival time of the k^{th} packet in the i^{th} talk interval frame, respectively (k being the time index when each packet arrives at the receiving host).

Second, the subprocess decision logic for the playout scheduling is performed using the estimated network jitter and the calculated current buffering delay. For this, the following three subprocesses are handled:

- ***Time expansion subprocess***: Let $C_R = L_a/J_{i,k}$ denote the ratio between total length L_a of the remaining signal frames in the signal frame buffer to the active network jitter $J_{i,k}$ of the k^{th} arriving packet estimated from the jitter buffer. The equation $E_R = T_a/J_{i,k}$ represents the ratio between time length T_a that the packet did not arrive at the receiver to the active network jitter $J_{i,k}$. If C_R is smaller than the compression threshold Th_c ($0 < Th_c < 3$) and E_R is larger than the expansion threshold Th_e ($0 < Th_e < 1$), the time expansion subprocess is entered.
- ***Time compression subprocess***: If C_R is larger than the compression threshold $Th_{c,}$ and $| J_{i,k} - b_{i,k} |$ is larger than length L_f of one frame, the time compression subprocess is entered. For the signal frames classified as voice, time compression is not processed in order to maintain the voice quality.
- ***Normal subprocess***: If C_R is smaller than the compression threshold Th_c and E_R is smaller than the expansion threshold Th_e, the normal subprocess is entered. Additionally, if C_R is larger than the compression threshold $Th_{c,}$ and $| J_{i,k} - b_{i,k} |$ is smaller than the L_f of one frame, the normal subprocess is entered.

3 Experimental Results

In order to evaluate the adaptive PS and PLC algorithm, a test bed is set up [6]. The test-bed consists of three modules: session initiation protocol (SIP) signaling, audio data transport, and a network traffic emulator. The VoIP application was installed in clients. In the network traffic emulator module, a traffic generator is used in order to simulate WLAN connections with different traffic loads, such as delay, jitter and packet loss.

The four network delay traces that we collected from the Internet links for the performance evaluation are listed in Table 1.

Table 1. Statistics of network traces.

Trace	End-to-End Network Delay (ms)	STD of Network Delay (ms)	Maximum Jitter (ms)	Network Packet Loss (%)
1	27.41	7.48	63	2.56
2	36.26	13.42	156	4.23
3	48.31	17.93	162	2.13
4	80.35	31.35	375	4.07

STD, standard deviation

In Table 1, the average network delay and the standard deviation (STD) of the network delay are depicted. To further characterize the delay jitter, the maximum jitter and the network packet loss is included in Table 2. To evaluate the quality degradation, objective voice quality testing is performed using PESQ [7], total buffering delay (TBE), jitter estimation error (JER), and late packet loss (LPL).

Table 2. Experimental results of the four methods.

Trace	Method	Wide Band (Sampling Rate: 16 kHz)			
		TBE (ms)	**JER (ms)**	**LPL (%)**	**PESQ score**
1	M1	29.83	26.32	1.39	2.312
	M2	41.43	15.36	0.72	2.946
	M3	33.37	21.55	0.10	3.263
	PM	**26.78**	**12.56**	**0.05**	**3.531**
2	M1	36.35	35.25	3.21	2.015
	M2	38.65	28.53	3.39	2.149
	M3	44.73	28.45	2.01	2.632
	PM	**39.25**	**21.24**	**0.74**	**3.205**
3	M1	38.15	36.20	3.39	2.242
	M2	40.27	28.58	3.54	2.376
	M3	47.65	27.35	2.31	3.056
	PM	**43.65**	**20.52**	**1.29**	**3.654**
4	M1	56.78	100.34	15.44	1.716
	M2	45.12	76.85	8.65	1.842
	M3	77.81	76.59	7.53	2.349
	PM	**61.51**	**67.84**	**2.72**	**2.961**

TBE, total buffering delay; JER, jitter estimation error; LPL, late packet loss late; PESQ, perceptual evaluation of speech quality;

The performance of our proposed method (PM) is compared with three methods that have been modified from the contents of reference papers and then implemented. Method 1 (M1) is based on the adaptive gap-based algorithm incorporated with spike detection [8], while Method 2 (M2) is based on an adaptive normalized least mean square playout algorithm with delay spike detection [1] and packet loss concealment [9]. In Method 3 (M3), the time-scale modification and loss substitution method [10] are combined with modeling the statistics of the inter-arrival times with the K-Erlang distribution [3]. Our simulations are based on the G.722.2 voice codec for wideband.

Table 2 shows that our proposed method (PM) outperforms the reference methods M1, M2, and M3 in medium jitter, high jitter, 2% packet loss rate, and 4% packet loss rate. The performance difference becomes more significant, showing the clear advantage of the PM.

4 Conclusion

In this paper, a receiver-based adaptive signal control including adaptive playout scheduling and packet loss concealment method was proposed and evaluated. Compared to conventional methods, the proposed method has lower computational complexity and enables users to deliver higher quality voice. The experimental results encourage the use of the proposed fully receiver-based enhancing algorithm in many practical mobile VoIP applications.

Acknowledgments. This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2015R1D1A1A01059804).

References

1. Shallwani, A., Kabal, P.: An Adaptive Playout Algorithm with Delay Spike Detection for Real-Time VoIP. Electrical and Computer Engineering, IEEE CCECE 2003, Canadian Conference on, Montreal, Quebec, Canada, pp. 997–1000 (2003).
2. Sat, B., Wah, B. W.: Analyzing Voice Quality in Popular VoIP Applications," IEEE Transactions on Multimedia 16(1), pp.46–59 (2009).
3. Li, H., Zhang, G., Kleijn, W.: Adaptive Playout Scheduling for VoIP using the K-Erlang Distribution. The 2010 European Signal Processing Conference, Algborg, Denmark, pp. 1494–1498 (2010).
4. Kovesi, B., Ragot, S.: A Low Complexity Packet Loss Concealment Algorithm for ITU-T G.722. IEEE International Conference on Acoustics, Speech and Signal Processing 2008, pp. 3969–3972 (2008).
5. Dorran, D.: Audio Time-Scale Modification. Dublin Institute of Technology Doctoral Thesis, (2005).
6. Ng, S. L., Hoh, S., Singh, D.: Effectiveness of Adaptive Codec Switching VoIP Application Over Heterogeneous Networks. 2nd International Conference on Mobile Technology, Applications and Systems, Guangzhou, China, pp. 7–13 (2005).
7. ITU-T Recommendation P.862: Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-end Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs, (2001).
8. Pinto, J., Christensen, K. J.: An Algorithm for Playout of Packet Voice Based on Adaptive Adjustment of Talkspurt Silence Periods. 24th Conference on Local computer Networks, Lowell, USA, pp. 224–231 (1999).
9. Sanneck, H., Stenger, A., Younes, K. B., Girod, B.: A New Technique for Audio Packet Loss Concealment. Global Telecommunications Conference, London, UK, pp. 48–52 (1996).
10. Andrsen, S. V., Kleijn, W. B., Sorqvist, P.: Method and Arrangement in a Communication System. U.S. Patent 7 321 851, (2008).

AT-MAC: A Novel Full Duplex MAC Design for Achieving Asymmetric Transmission

Zhijie Ma1, Qinglin Zhao2, Yu Zeng3, Huan Zhang4, Hong-Ning Dai5

Faculty of Information Technology, Macau University of Science and Technology, Avenida Wei Long, Taipa, Macau, China

1. mazhijie0000@hotmail.com; 2.zqlict@hotmail.com; 3.cardwhen@yahoo.com; 4. huan@buu.edu.cn; 5.hndai@ieee.org

Abstract. Wireless Full-duplex technique can significantly improve system throughput of a wireless network. Most existing full-duplex-supporting wireless MAC schemes only achieve symmetric transmission (i.e., the uplink throughput of a node is equal to its downlink throughput). However, in reality, the uplink and downlink traffic is often asymmetric. In this paper, we propose a novel MAC scheme called AT-MAC to achieve asymmetric transmission. In AT-MAC, upon receiving a packet, a receiver determines whether to execute a reverse transmission to the peer or to initiate a new transmission to a third node, according to the peer's uplink and downlink requirements. Extensive simulations verify that the proposed design is feasible and efficient.

Keywords: Full Duplex, Asymmetric Transmission, WLANs

1 Introduction

With the rapid development of wireless techniques, people can download resources from Internet and upload some materials to Internet more easily and quickly. Full duplex technique makes it possible to upload and download at the same time, and significantly improves system efficiency [1,2]. However, since different people may have different throughput requirements for upload and download, the existing full duplex designs [3,4,5] would lead to unideal bandwidth allocation, and therefore lead to the waste of the system resource.

The related designs are summarized as follows. Authors in [3] proposed a simple distributed MAC protocol for a WLAN. In this protocol, each node initiates the transmission following CSMA/CA mechanism, and runs in the basic mode. When a receiver receives and decodes the header of a packet, it would immediately start the reverse transmission to the sender. In this way, the full duplex transmission occurs between a pair of sender and receiver nodes. Different from [3], in [4], the sender only sends the header of a packet. When the receiver decodes the header, it would reply the sender with bit information, where the bit value represents the transmission mode. For example, 0 denotes that the receiver would start a transmission to the send-

Corresponding author: Qinglin Zhao
Digital Object Identifier: xx

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_5

er, 1 denotes that the receiver would start a transmission to the other node, etc. In this manner, the scheme in [4] can build up a full duplex transmission between two nodes or three nodes. Authors in [5] proposed a full duplex MAC protocol, which is similar to [3], except that all nodes run in RTS/CTS mode. When nodes are in saturation (where a node always has packets to transmit and this situation frequently occurs in peer-to-peer networks [6]), however, all these designs will lead to a phenomenon: for a sender-receiver pair communication, the uplink throughput (from the sender to the receiver) is equal to the downlink throughput (from the receiver to the sender). In this paper, we call a transmission with equal uplink and downlink throughput a symmetric transmission; call it an asymmetric transmission otherwise. Obviously, for nodes which have different throughput requirement on their uplink and downlink, it is more desirable to achieve asymmetric transmission. As a result, all these existing designs would lead to the unreasonable resource allocation in some scenarios.

In this paper, in order to achieve a reasonable allocation of system resource, we propose a novel MAC scheme called AT-MAC for full duplex wireless LAN that aims at achieving asymmetric transmission for nodes. In our design, each node can obtain desirable uplink and downlink throughput according to its requirement.

The rest of this paper is organized as follows. In Section 2, we outline our AT-MAC design. Section 3 details our AT-MAC design. Section 4 presents the simulation results to verify our design. Section 5 concludes this paper.

2 AT-MAC Overview

In this paper, we outline the design idea of AT-MAC. We consider an infrastructure-based wireless LAN, which consists of an access point (AP) and some nodes.

The main goal of our proposed AT-MAC is to enable each node obtain the desired uplink/downlink throughput according to its own requirement. The transmission procedure of AT-MAC follows CSMA/CA [7], as shown in Fig. 1. In this figure, once the node has data to transmit, it first senses the channel for DIFS interval time. If the channel is idle, then it starts to contend for channel using 802.11 DCF binary exponential backoff (BEB) algorithm. When the backoff counter is decreased to 0, the asymmetric full-duplex transmission (marked with a dashed red border) is executed. In the asymmetric full-duplex transmission part, there are two concurrent transmissions: primary transmission (PTX) and secondary transmission (STX), where the PTX is initiated by sender, and the STX is initiated by the receiver. After that, the sender and receiver send ACK to each other after a SIFS interval time concurrently.

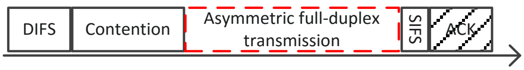


Fig. 1. The transmission procedure of the proposed AT-MAC.

In AT-MAC, according to the requirement of the node, we divide the nodes into two types: Uplink-dominant (UD) node and Downlink-dominant (DD) node, where we call the node a UD node if its desired uplink throughput is larger than the throughput from AP to this node, and we call the node a DD node if its desired uplink throughput is not larger than the throughput from AP to this node. In this paper, the node is either the UD node or the DD node. Therefore, there are three cases for the wireless LAN: 1) there only exists the UD nodes; 2) there only exists the DD nodes; and 3) UD nodes and DD nodes coexist. Here, we only focus on the system with a mix of UD and DD nodes. The other two cases will be considered in our future work.

3 AT-MAC design details

In this section, we detail our AT-MAC design in an infrastructure-based wireless LAN, where there are one AP and multiple nodes. Note that in full duplex context, the PTX can be initiated by either node, or AP, or a pair of node and AP (node sends data to AP, and AP also sends data to this node concurrently). In the following, we first focus on the asymmetric transmission procedure marked in Fig. 1, and introduce our design from three scenarios: node initiates the PTX, AP initiates the PTX, and a pair of node and AP initiates the PTX. Then we introduce another scenario when collision occurs.

3.1 Node initiates the PTX

When the node initiates the PTX to AP, there are two cases: the DD node initiates the PTX and the UD node initiates the PTX. In this subsection, we introduce our design under these two cases in turn.

DD node initiates the PTX: Suppose DD node A wins the channel and initiates the PTX to AP. In our design, AP would start the STX to node A immediately. Then the asymmetric full-duplex transmission occurs between DD node A and AP, as shown in Fig. 2(a).

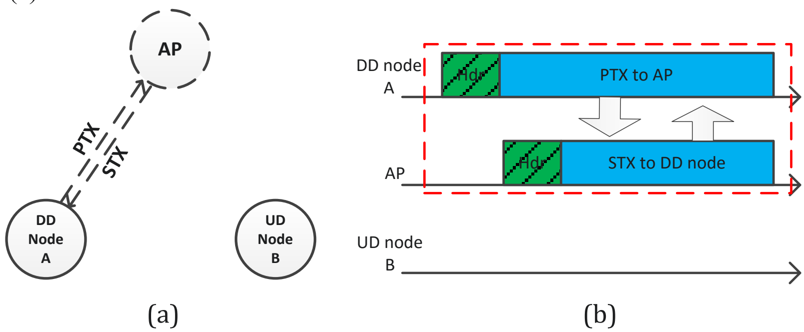


Fig. 2. The DD node initiates the PTX: (a) the topology (b) the transmission procedure.

We use Fig. 2 (b) to further explain the transmission procedure. In this figure, when the DD node A access the channel and initials the PTX through the uplink channel, all of the rest nodes, including AP, pause their backoff counters, and then try to decode the header of the incoming packet. After that, AP would immediately initiate the STX to the sender no matter whether its backoff counter has reduced to 0. On the contrary, because the uplink channel has been occupied by the DD node, the remained nodes continue suspending their backoff counters and do not transmit, then the full-duplex transmission operation occurs between DD node A and AP. After DD node A and AP finish the transmission, they send ACK frame to each other concurrently after a SIFS interval time.

UD node initiates the PTX: When UD node B wins the channel and initiates the PTX to AP, in our design, AP would start the STX to a random DD node (i.e., DD node A), then the asymmetric full-duplex transmission happens among DD node A, UD node B, and AP, as shown in Fig. 3(a).

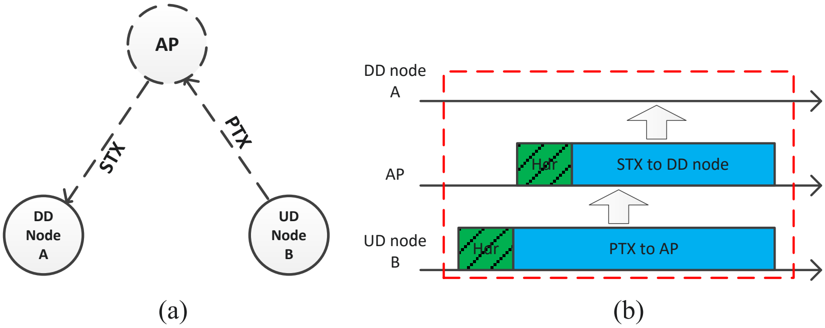


Fig. 3. The UD node initiates the PTX: (a) the topology (b) the transmission procedure.

We use Fig. 3(b) to further explain the transmission procedure. Unlike the case when the DD node initiates the PTX, there are two differences between them. Firstly, once the header of the incoming packet is decoded, AP would start the STX to one selected DD node immediately, instead of initiating the STX to the sender. This is because the DD node requires more downlink throughput, then AP should allocate resource to DD node as much as possible. Secondly, when UD node B initiates the PTX to AP and AP starts the STX to DD node A, there is an inter-node interference from node B to node A. To handle that, we need to guarantee the signal to interference ratio (SIR) at node A is above the minimum threshold γ_{th}, $SIR_A > \gamma_{th}$, so that node A can receive the data successfully.

Let P_{AP} and P_B be the transmission power of AP and UD node B, respectively. Let $h_{AP,A}$ be the channel gain between AP and node A, and $h_{A,B}$ be the channel gain between node A and node B. Then, the SIR at node A, SIR_A, can be expressed as

$$SIR_A = \frac{P_{AP} \, | h_{AP,A} |^2}{P_B \, | h_{A,B} |^2}.$$

Assume that AP has the information of the node's power and the channel gain between each node, then AP can determine the DD nodes which can satisfy the above threshold constrained condition. Finally, AP only needs to select one DD node, and then starts the STX to this DD node. After the transmissions from the UD node to AP and from AP to the DD node are finished, AP and the DD node respectively send the corresponding ACK frame to the UD node and AP after a SIFS interval time.

3.2 AP initiates the PTX

When AP initiates the PTX, there are also two cases: AP initiates the PTX to UD node, and AP initiates the PTX to DD node. The first case is similar to the case when DD node initiates the PTX to AP. Therefore, in this section, we only focus on the case when AP initiates the PTX to DD node.

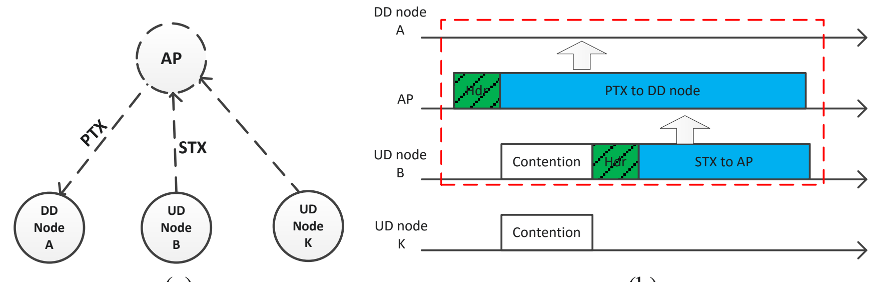


Fig. 4. AP initiates the PTX to the DD node: (a) topology (b) transmission procedure.

When AP initiates the PTX to DD node A, only UD nodes are allowed to contend for the STX, while DD nodes are forbidden to contend due to asymmetric requirement. When the counter is reduced to 0, then UD node starts the STX to AP, then the asymmetric full-duplex transmission can occur among the DD node, the UD node, and AP, as shown in Fig. 4(a).

We use Fig. 4 (b) to further illustrate the transmission procedure. In this figure, when AP starts the PTX to the DD node A, all the nodes pause their backoff counters and decode the header of the packet. When the header is decoded, all nodes know that the downlink channel is occupied, while the uplink channel is free, so all nodes can continue contending for the channel. In our design, since the DD nodes have less requirements for the uplink channel, and the UD nodes have more requirements, then we let the DD nodes proceed to suspend their backoff counters, and do not participate the contention until the asymmetric transmission is finished. The UD nodes, on the contrary, would first check if the incoming packet is for them through the address information in header. If the packet is for itself, it would start the STX to AP at once. Otherwise, the UD node continues decreasing its backoff counter if the uplink channel is still free. When the backoff counter is 0, the UD node, such as UD node B, starts the STX to AP.

Note that when the UD node starts the STX to AP, there will be interference from this UD node to node A. To handle that, we only let the UD node, which can satisfy $SIR_A > \gamma_{th}$, decrease its backoff counter, and the node, which cannot satisfy this condition, has to suspend its backoff counter. The SIR at node A can be expressed as $SIR_A = \dfrac{P_{AP}\,|h_{AP,A}|^2}{P_K\,|h_{K,A}|^2}$, where P_K denotes the power of UD node K, and $h_{K,A}$ denotes the channel gain between UD node K and DD node A.

When AP and the UD node finish the transmission, DD node and AP send ACK to each other concurrently.

3.3 A pair of AP and node initiate the PTX

In half duplex wireless communications, it would lead to a collision if two senders transmit data concurrently. In contrast, full duplex technique makes it possible if two senders are a pair of nodes.

In our design, when a pair of node and AP initiate the PTX to each other simultaneously, after decoding the header, all the other nodes would stop their backoff counter, since the uplink and downlink channel are all occupied. Once the transmission is finished, a pair of node and AP transmits ACK to each other in the interval time of SIFS.

3.4 Collision Problem

When more than two senders, or two-sender which are not a pair of node and AP, initiate the PTX at the same time, the collision would occur. At that time, AP would send the collision notification (CN) [8] signal to resolve the collision, as shown in Fig. 5. In this figure, we assume that two nodes initiate the PTX to AP concurrently, since AP and other nodes cannot decode the header of the incoming packet, AP and other nodes infer that there is a collision. To resolve the collision, AP directly broadcasts the CN signal which can be well-detected under low SNR. Once the transmitted node receives this CN signal, they would immediately stop their ongoing transmissions and release channel.

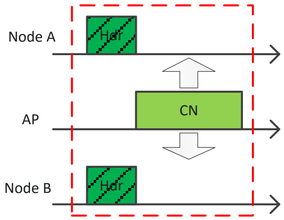


Fig. 5. The transmission procedure when the collision occurs.

4 Performance Evaluation

In this section, we simulate the proposed AT-MAC design through our C++-based simulator. The topology is shown in Fig. 6. To directly reveal the feasibility on achieving asymmetric transmission, we simplify the simulation and make the following assumption: 1) all nodes and AP have the same CW value, 2) when UD node initiates the PTX to AP, and AP starts the STX to DD node, there is no inter-node interference between the UD node and DD node, 3) when AP initiates the PTX to DD node, all UD nodes can proceed to decrease their backoff counters, and there is no inter-node interference either. In our simulation, we set the packet size to 1500 bytes, and fix the number of the DD node (n) and the UD node (m) to n=m=5. The parameter settings are listed in Table 1. The parameters in Table 1 are set by IEEE 802.11b. For each simulation, we only change the contention window size of each node, CW_i, and set $CW_{\max} = CW_{\min} = CW_i$ for node i. Each simulation run lasts for 200 seconds. Below, we detail the simulation results.

Table 1. Parameter settings in simulation.

σ	20 us	Data Rate	11 Mbps
SIFS	10 us	PHY Header	26 byte
DIFS	50 us	MAC Header	28 byte
CN	100 us	Data	1500 byte
Basic Rate	1 Mbps	ACK	38 byte

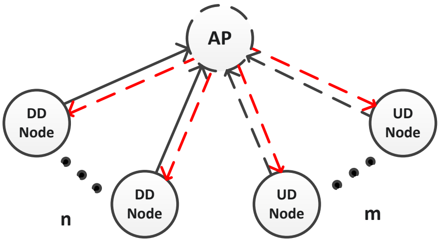


Fig. 6. The topology used in our simulation.

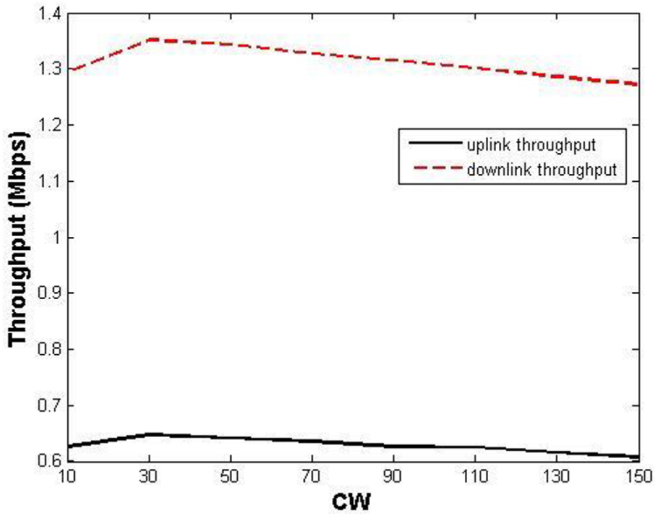


Fig. 7. The average uplink and downlink throughput for the link between the DD node and AP, where n = m = 5, and CW = 10, 20, ..., 150.

Fig. 7 plots the average uplink throughput and downlink throughput for the link between DD node and AP, when CW=10,30,...,150. In this figure, the dotted line with red color represents the average downlink throughput that from AP to DD node, and the solid line with black color represents the uplink throughput that from DD node to AP. From this figure, we can see that 1) as CW increases from 10 to 150, the uplink throughput and the downlink throughput first increase and then decrease. 2) No matter how CW varies, the downlink throughput is always higher than the uplink throughput. And 3) we can infer that there exists the optimal CW value which can maximize the uplink throughput and downlink throughput. This manifests that our AT-MAC design can well achieve the asymmetric transmission for the link between AP and DD node, namely, the downlink throughput is larger than the uplink throughput.

Fig. 8 plots the average uplink throughput and downlink throughput for the link between UD node and AP, when CW=10,30,...,150. From this figure, we can clear see that the result is similar to Fig. 7, except that the uplink throughput is larger than the downlink throughput. This manifests that our AT-MAC design can well achieve the asymmetric transmission for the link between AP and UD node, namely, the uplink throughput is larger than the downlink throughput.

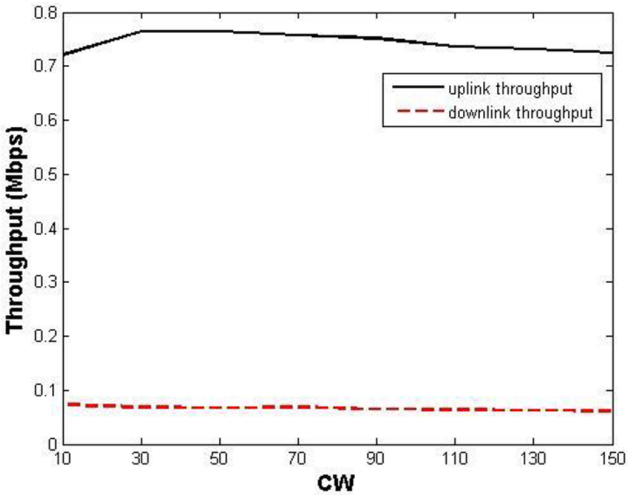


Fig. 8. The average uplink and downlink throughput for the link between the UD node and AP, where n = m = 5, and CW = 10, 20, …, 150.

5 Conclusion

In this paper, we propose a novel full duplex MAC design, AT-MAC. Through our design, each node can obtain the desired uplink and downlink throughput according to its requirement. We validate our design through extensive simulation, and the simulation results verify that our design is feasible and efficient.

Acknowledgment

This work is supported by the Macao Science and Technology Development Fund under Grants 081/2012/A3, 104/2014/A3, 013/2014/A1, and 096/2013/A3.

References

1. J. L. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti, "Achieving single channel, full duplex wireless communication," in Proceedings of the sixteenth annual international conference on Mobile computing and networking, pp. 1-12, ACM, 2010.
2. D. Bharadia, E. McMilin, and S. Katti, "Full duplex radios," in Proceedings of the ACM SIGCOMM, pp. 375-386, ACM, 2013.
3. A. Sahai, G. Patel, and A. Sabharwal, "Pushing the limits of full-duplex: Design and real-time implementation," arXiv preprint arXiv:1107.0607, 2011.
4. S. Goyal, P. Liu, O. Gurbuz, E. Erkip, and S. Panwar, "A distributed MAC protocol for full duplex radio," in Signals, Systems and Computers, 2013. Asilomar Conference on, pp. 788–792, IEEE, 2013.
5. M. Duarte, A. Sabharwal, V. Aggarwal, R. Jana, K. Ramakrishnan, C. W. Rice, and N. Shankaranarayanan, "Design and characterization of a full duplex multiantenna system for WiFi networks," IEEE Transactions on Vehicular Technology, vol. 63, no. 3, pp. 1160–1177, 2014.
6. R. Dunaytsev, D. Moltchanov, Y. Koucheryavy, O. Strandberg, and H. Flinck, "A survey of P2P traffic management approaches: best practices and future directions," Journal of Internet Engineering, vol. 5, no. 1, pp. 318-330, 2012.
7. IEEE 802.11 - wireless lan medium access control (MAC) and physical layer (PHY) specifications, IEEE Std. 02.11-2007, June 2007.
8. S. Sen, R. R. Choudhury, and S. Nelakuditi, "CSMA/CN: carrier sense multiple access with collision notification," IEEE/ACM Transactions on Networking (TON), vol. 20, no. 2, pp. 544–556, 2012.

Effective Noise Reduction Methods for Rear-View Monitoring Devices Based on Microprocessors

Huy Toan Nguyen[1], YongSu Choi[2], Gwon Seok Sun[3], Seung You Na[4], Jin Young Kim[5]

Chonnam National University, Gwangju, Rep. of Korea
[1]nguyenhuytoantn@gmail.com
[4]syna12@chonnam.ac.kr
[5]beyondi@jnu.ac.kr
Novasez Co., Gwangju, Rep. of Korea
[2]yschoi@novasez.com
Korea Polytechnics College, Rep. of Korea
[3]mtrsks@hanmail.net

Abstract. This paper suggests effective noise reduction methods for the neckband devices which are able to monitor the rear-view images. This device helps the user to keep an eye on the rear areas which he/she cannot see in normal ways without turning back. In order to be used as a wearable device, the neckband device should have some particular properties such as small sizes, light weight, and low power consumption. When users walk or move, the neckband device also moves, which causes many noise effects to the system such as illumination changes and especially severe camera motion. We propose effective noise reduction methods to deal with these noise effects. The experiments demonstrate that the proposed methods have good performance in rear-view monitoring under the practical difficulties that arise due to users' movement such as illumination changes and out of focus problems.

1 Introduction

Video monitoring systems have been in use to monitor dangerous and interested areas. The video monitoring system needs to be fast, reliable and easy to use. A vast amount of research work has been presented to tackle these problems. O. Jafari *et al* [1] and D. Mitzel *et al* [2] proposed algorithms to detect pedestrian in close-range using head-mounted camera based on Kinect RGB-D input data, and this work was run on single CPU core and achieved 18 fps. In [3] R. Rashmi *et al* built an application using IP camera to send and receive data via a network and the Internet; the image is transferred from mobile to PC. Lefang *et al* [4] designed a remote video monitoring system based on ARM processing chip using Linux operating system combined with GPRS technologies. X. Jiangsheng [5] proposed a video monitoring system based on TMS320DM642 DSP, local network and GPRS for large railway maintenance machinery. However, most of these mobile monitoring systems are based on

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_6

computers or super computers for processing acquisition videos. The high cost and inconvenience for a single user limit the popularity of these systems in a real life. The design and implementation of a wearable device, which is small and light-weight as SenseCam device [6] with real-time processing of input videos, is a hot issue now.

In this paper, we propose and implement a neckband device for the rear-view monitoring task. The main purpose of this small device is monitoring blind spot areas behind the user without turning back. The system constraints such as memory resources, microprocessor computation ability are challenges for researchers. The proposed algorithms to deal with the illumination changes, random noises, and strong camera motion when the user walks, are presented.

The paper is organized as follows. In section 2, we present our proposed system including hardware and software architectures in detail. After that, in section 3, we show the experimental results. Finally, the conclusion is presented in section 4.

2 The Proposed system

The implementation of neckband device can be divided into two main parts: 1) Hardware Architecture and 2) Software architecture.

2.1 Hardware Architecture

The neckband device hardware can be divided into several parts. The system schematic diagram is presented in Fig. 1.

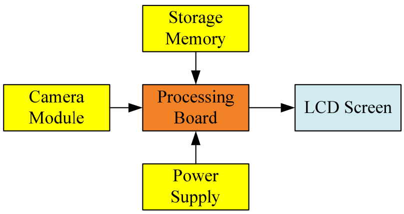


Fig. 1. System diagram.

The camera module, which is plugged on neckband, is responsible for image acquisition and transmission via cable to Processing Board, which is the main computation part of the system. The operating system and software algorithms are stored in storage memory. The software algorithm will be run on Processing Board and send the results to the LCD Screen. All parts of the system get power from the Power Supply.

2.2 Software Architecture

In this section, we propose an algorithm for monitoring moving objects in a rear-view range from the camera mounted on the neckband when the user walks. The object which we choose for monitoring is a human in a close-range of from 1 to 10 meters. The proposed method has three main steps: pre-processing, detection of moving foreground objects and objects classification. The main purpose of the first step is reducing noise and correcting the illumination changes in the image. The second step is proposed to extract the moving foreground object regions. Finally, we extract features in moving object regions and classify them to get the output results. The flowchart of the software architecture is presented in Fig. 2.

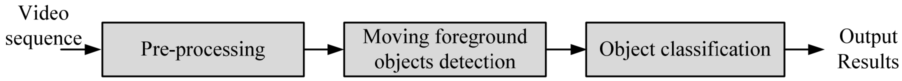


Fig. 2. The software architecture flowchart**.**

Due to hardware limitations, we choose simple algorithms for the pre-processing step. When the camera module is activated, the input image $I(x, y, t)$ immediately is resized to 320x240 pixels and converted to gray scale $I^{gray}(x, y, t)$. In order to reduce noise from the environment, we apply Gaussian Filter on the converted image.

In our case, the camera is mounted on the neckband. The camera would be moved when the user walks, which causes the background to be unstable. There are two kinds of motion that we have to consider here: the ego-motion of the camera and the motion of moving objects. In case of non-static background, to know what parts of the picture are changing because of the camera motion and what parts are changing independently, the camera motion has to be estimated [7]. Let $I^{gray}(x, y, t-1)$ and $I^{gray}(x, y, t)$ be two consecutive frames. Ideally, the relationship between $I^{gray}(x, y, t-1)$ and $I^{gray}(x, y, t)$ can be presented as Eq. (1):

$$I^{gray}(x, y, t) = HI^{gray}(x, y, t-1) \tag{1}$$

or

$$\begin{pmatrix} x_t \\ y_t \\ 1 \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \\ 1 \end{pmatrix} \tag{2}$$

where H is a homography matrix.

In order to extract the moving foreground objects, we extract feature points from the frame $I^{gray}(x, y, t-1)$ as explained by J. Shi *et al* [8]. Next, we use Kanade-Lucas-Tomashi (KLT) for tracking the features points in the frame $I^{gray}(x, y, t)$. After that, we find a projective transformation matrix between the two frames which is called the homography matrix *H* by RANSAC method [9]. From matrix *H*, we compute the reference motion-compensated image $I_{\mathrm{Ref}}^{gray}(x,y,t-1)$:

$$I_{\mathrm{Ref}}^{gray}(x,y,t-1) = HI^{gray}(x,y,t-1) \tag{3}$$

We subtract $I^{gray}(x,y,t)$ from $I_{\mathrm{Ref}}^{gray}(x,y,t-1)$ to get the error image $E^{gray}(x,y,t)$. The error image $E^{gray}(x,y,t)$ presents the moving foreground objects independently from the camera motion:

$$E^{gray}(x,y,t) = \left| I_{\mathrm{Ref}}^{gray}(x,y,t-1) - I^{gray}(x,y,t) \right| \tag{4}$$

In an ideal case, the error image $E^{gray}(x,y,t)$ only contains the moving foreground objects. However, in our real experiments, the camera motion, when the user walks, is changing, even stronger than the object movements. As a result, the error image also contains the false positive pixels. Using normal methods which are applied on a gray scale image to get the moving foreground objects is impossible in this case. Hence, we propose to use color information from the original image $I(x,y,t)$. The proposed error image is presented as the following equation:

$$E(x,y,t) = \begin{cases} 0\ if\ E^{gray}(x,y,t) = I^{gray}(x,y,t)\ or\ E^{gray}(x,y,t) < th_E \\ \quad I(x,y,t) \qquad else \end{cases} \tag{5}$$

Finally, we apply color threshold on error color image $E(x,y,t)$ and morphology to get the moving foreground objects.

After the detection step of moving foreground objects, we get the region which contains moving objects for monitoring. When a person is in the rear-view range, the camera cannot capture the full-body images. The other full-object detectors [10, 11] are not suitable for our case. Therefore, in this study, we only focus on face detection in a close-range. We choose Local Binary Pattern descriptor and AdaBoost classifier for classification tasks. The original LBP [12] forms labels for the image pixels by using the threshold of the 3x3 neighborhood of each pixel with the center value and then by considering the result as a binary number. The basic methodology for LBP based face description was proposed by Ahonen *et al* [13]. The facial image is divided into local regions, and LBP texture descriptors are extracted from each region independently. Due to the necessity of the fast response, we select the Adaboot algorithm[14] for classification tasks.

3 Experiments and results

In order to make a small wearable device with high processing ability, low-power consumption, and light weight, we use Raspberry Pi 2 model B (RPI) as the main processing board. Raspberry Pi 2 model B is a credit card sized single-board computer. We utilize Raspberry Pi camera module, as the system input, which has 5-

Megapixel photography, a high frame rate of 720p/60fps high-definition (HD) video capture, high sensitive, low noise and low crosstalk image capture with an extremely small size (25x20x9mm) and light weight (3 grams). The neckband device hardware is shown in Fig. 3.

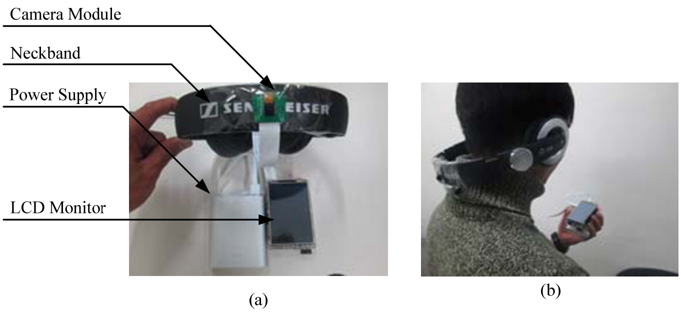


Fig. 3. The proposed neckband device: (a) Hardware structure, (b) Setup device.

To evaluate the proposed method on the neckband device, we record videos in real conditions where illumination changes and the user keeps walking when recording the videos. The software algorithms are implemented by using OpenCV library in Python programming language. The processing time of each step is shown in Table 1.

Table 1. The time consumption of each step for one frame.

Step	Time consumption(second)
Pre-processing	0.0109
Moving foreground object detection	0.5634
Face classification	0.1969
Total Time	**0.7712**

We test our methods for cancelling the noise due to user walking movements on several video sequences with different walking speeds when recording data both indoors and outdoors. The results are presented in Table 2.

Table 2. The noise cancellation results.

Video sequence	Environments	Walking speed	Accuracy
Video 1	Indoor	Fast	69.71%
Video 2	Indoor	Normal	73.04%
Video 3	Indoor	Slow	88.70%
Video 4	Outdoor	Slow	81.48%

We compare time consumption when using LBP descriptors and traditional Haar-features which were proposed by other researchers in [15-16]. The results are presented in Table 3.

Table 3. Comparison of face classification using difference features

Features	Time consumption(second)
Haar-features	0.556
LBP descriptors	**0.1969**

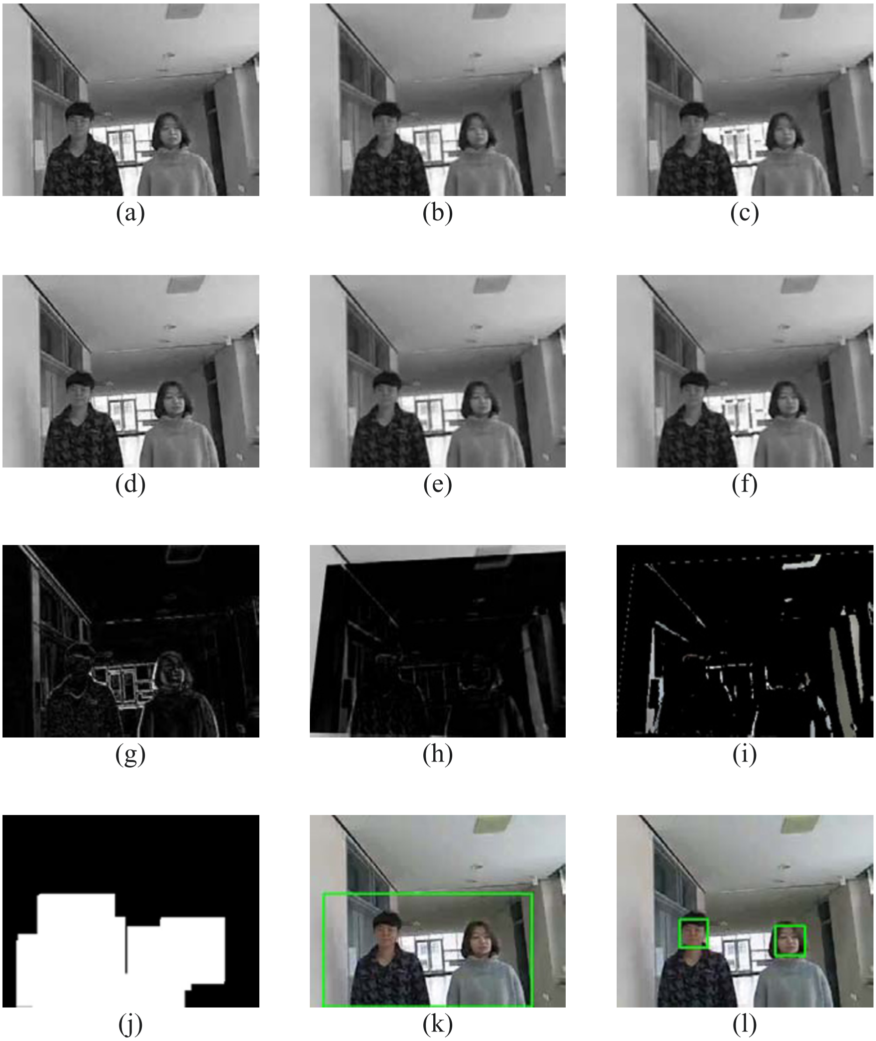


Fig. 4. (a) (d) Two consecutive frames in gray scale, (b) (e) Two consecutive frames after correct illumination, (c) Features from Good Feature to Track, (f) KLT tracker result, (g) Normal subtraction of two frames, (h) Error image, (i) Modified error image, (j) Region of moving background objects, (k) Foreground moving objects detection results, (l) Detection results.

4 Conclusion

In this paper, we implement a neckband device for rear-view monitoring. To reduce noise components in the recorded video data, we propose a method of combining color feature and motion to reduce noise and camera motion effects on the system. When compared with other mobile monitoring systems, our system is even cheaper and much more suitable for an individual user. The device works with a high speed and high accuracy for human detection in rear-view and different environmental conditions. The users are able to monitor the blind rear-side areas without turning back. This mobile low-cost device can be applied to other monitoring applications in the future such as monitoring blind areas around cars, or surveillance purposes using drones and mobile navigation applications.

Acknowledgments. This research was supported by the Ministry of Education, Science Technology (MEST) and National Research Foundation of Korea (NRF) through the Human Resource Training Project for Regional Innovation.

References

1. Jafari, O., Mitzel, D., Leibe, B.: Real-time RGB-D based people detection and tracking for mobile robots and head-worn cameras. Proceedings of IEEE International Conference on Robotics and Automation (ICRA ’14), (2014)
2. Mitzel, D., Leibe, B.: Close-range human detection for head-mounted camera. British Machine Vision Conference (BMCV ’12), (2012)
3. Rashmi, R., Latha, Dr. B.: Video surveillance system and facility to access Pc from remote areas using smart phone. International Conference on Information Communication and Embedded System (ICICES), (2013) 491-495
4. Lefang, Z., Jian-xin. W., Kai. Z.: Design of embedded video monitoring system based on S3C2440. International Conference on Digital Manufacturing & Automation (ICDMA), (2013) 461-465
5. Jiangsheng, X.: Video monitoring system for large maintenance machinery. International Conference on Electronic Measurement & Instrument (ICEMI), Vol. 3, (2011) 60-63
6. Hodges, S., Williams, L., Berry, E., Izadi. S., Srinivasan, J., Butler, A., Smyth, G., Kapur, N., Wood, K.: SenseCam: A retrospective memory aid. In Proceedings of the 8th International Conference of Ubiquitous Computing, (2006) 177-193
7. Szcolgay, D., Benois-Pineau, J., Mégret, R., Gaëstel, Y., Dartigues, J. F.: Detection of moving foreground objects in videos with strong camera motion. Pattern analysis and applications, Springer Verlag, (2011) 331-328
8. Shi, J., Tomasi, C.: Good Features to Track. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (1994) 593-600
9. Fischler, M. A., Bolles, R. C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Magazine Communications of the ACM, Vol. 24, Issue 6, (1981) 381-395
10. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, (2005) 886-893
11. Felzenszwalb, P., Girshick, B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 32, No. 9, (2010) 1627-1645

12. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. Pattern Recognition, Vol. 29, Issue 1, (1996) 51-59
13. Ahonen, T., Hadid, A., Pietikäinen, M.: Face description with Local Binary Patterns: Application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28, No. 12, (2006) 2037-2041
14. Freund, Y., Schapire, R. E.: A decision-theoretic generalization of on-line learning and an application to boosting. Journal of computer and system sciences, Vol. 55, Issue 1, (1997) 119-139
15. Moreno, R. J.: Robotic explorer to search people through face detection. International Congress of Engineering Mechatronics and Automation (CIIMA), (2014) 1-4
16. Abaya, W. F., Basa, J., Sy, Abad, A. C.: Low cost smart security camera with night vision capability using Raspberry Pi and OpenCV. IEEE International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control Environment and Management (HNICEM), (2014) 1-6

A Dual-mode Beacon Profile for Normal and Disaster Environments

Sunggyun Jang, Guangqiu Jiang, Inwhee Joe

Department of Electronics Computer Engineering, Hanyang University
Seoul, 133-791 Republic of Korea
Corresponding Author: Inwhee Joe
{mrjang28, jgqandsh, iwjoe}@hanyang.ac.kr

Abstract. In these days, the beacon has been widely used in all aspects of life. It can be used in marketing and loyalty, based on the location of advertisements or messages. But it has not been applied to the disaster environment. Recently, the disasters such as fire disaster and earthquake occur frequently around the world. They cause great threats to life and property. However, Modern architectures have high-rise and complex features that bring danger and greater challenges to the search and rescue work for firefighters. This paper proposes the design and implementation for a portable, power saving dual-mode beacon profile. The dual-mode beacon not only has the function of sending an advertisement, but also can sense the fire and earthquake to send out alarms. This system supports both the original beacon system and the disaster system. The beacon system gives you advanced information using Indoor Positioning System, while the disaster system helps you to exit and service escape route.

Keywords. Embedded Software, Beacon, IPS, Advertising packet, dual-mode beacon

1 Introduction

The dual-mode beacon profile is realized by the beacon profile based on BLE (Bluet ooth Low Energy). In the normal mode, a dual-mode beacon like an ordinary beacon s ends some discount coupons or shop event messages to nearby smart devices. In the d isaster environment, a dual-mode beacon is divided into four states according to how l ong the disaster happened. In this paper, we first introduce the importance of the devel opment of a dual-mode beacon profile. In Section 2, we introduce Beacon. Then we a nalyze the characteristics of beacon and structure of broadcast data packet. We design the dual-mode beacon and define the dual-mode beacon advertising packet in section 3. The performance results of a dual mode beacon are reported in section 4. Finally se ction 5 gives the conclusion and further works.

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_7

2 Related work

2.1 Beacon

Beacon enables consumers to pay at stores completely hands-free [3]. Beacon is a device that can transmit a small amount of information to central devices through radio signals [5]. Android4.3 or IOS 7 and above version of smartphone or tablet as a central device can scan the signal from the beacons. According to the signal strength can calculate how far away the central device is from the beacon. The results are divided into three states: Immediate, Near, Far [5].

Table 1. Beacon three states

Types	Description
Immediate	approx. 10 centimeters away
Near	approx. 2–3 meters away
Far	approx. 5–70 meters away

Currently the best approach to determine the position indoor is using signal strengt h values. There is an equation that works well for distance estimation:

$$\text{RSSI (dBm)} = -(10\log_{10}(d) + A) \tag{1}$$

The ‘d’ is the distance we want to know, ‘A’ is received signal strength in dBm at one meter, ‘n’ is the path-loss exponent. The path loss exponent has to be determined experimentally. To cover this equation to an algorithm first we need to normalize the RSSI and txPower values. Second we need to measure a bunch of RSSI measurements at known distances, then do a best fit curve to match the data points. Finally we can convert the best fit curve into an algorithm.

2.2 Non-contact Infrared Thermometers

The DTPM11, Non-contact Infrared Thermometers, is a kind of temperature sensor module which can measure the temperature of the object surface in 500 milliseconds without touching the object. The DTMP11 embeds a microprocessor for Temperature calculating function, which can output the correct temperature measurements. One advantage of this is that Master Controller does not need a temperature calculation algorithm. The DTPM11 outputs the temperature measurement by the way of Serial Peripheral Interface and the maximum value of the SPI clock is 1 MHz. It can simultaneously measure the environment temperature and the target temperature. The range of temperature measurement of the DTMP11 is -20℃ to 100℃ and its temperature resolution is 0.1℃. It operates from an input voltage of 2.4V to 3.6V with as much as 2 percent accuracy.

2.3 Three axis linear accelerator

The LIS3DSH is a three axis linear accelerator and support voltage range from 3 to 5 volts. It measures the acceleration at scales of ±2g/±4g/±6g/±8g/±16g to detector the earthquake but not enough. User stores data to limit intervention by using LIS3DSH's integrated FIFO buffer. Land Grid Array ensures that LIS3DSH works between -40 °C and 85 °C. LIS3DSH has two communication interfaces, one is serial peripheral interface, and the other is inter IC control interface. Each state machine can be applied to such fields as gesture recognition, 4D/6D orientation, and pulse counter and so on. Write 1 on the ST bit to enable the self-test function which can test the performance of the sensor. The actuation force acting on the sensor can simulate the input acceleration, and if the DC level change of output signals is within the amplitude, then the sensor is working properly. Choose a working mode of FIFO from Bypass mode, FIFO mode, Stream mode and Stream-to-FIFO mode by setting the FIFO_MODE bits.

2.4 An advertising data packet

An advertising data packet is transmitted once every 350 milliseconds. An advertisi ng data packet consists of four parts: 1 byte preamble, 4 bytes access address, 39 byte s advertising channel PDU, and 3 bytes CRC.

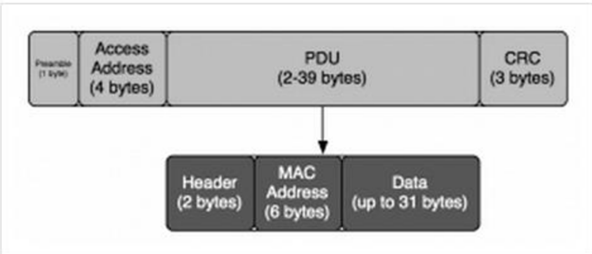


Fig. 1. Advertising data packet

The access address is a fixed value for 0x8E89BED6. Two bytes of header and 6 by tes of the MAC address and 31 bytes of payload data together form the PDU. The pay load data is from the peripheral device in order to be found by the central device. CR

C is used to confirm whether the central device accurately receives the advertising dat a packet. The first 9 bytes of payload data are a fixed data called beacon prefix. The pr oximity UUID, 16 bytes data on the back of beacon prefix, is used to distinguish betw een your beacon and someone else's beacon. The major number plays the role to grou p all your beacon and it is 2 bytes in size. The minor number, the size of two bytes, is used to distinguish each beacon in the group. The TX power is the signal strength that measured at 1 meter far from beacon.

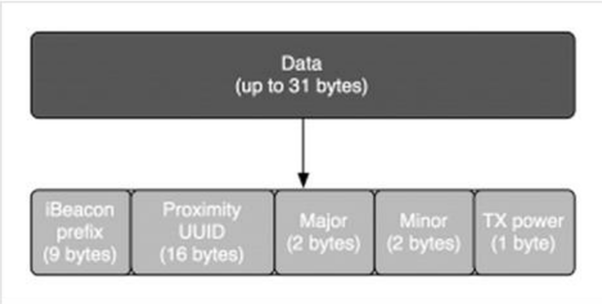


Fig. 2. Payload data unit of beacon

3 Design and Implementation

3.1 Two environments

There are two environments in dual-mode beacon profile;

1. Beacon mode: Normal Execution Environment (NEE)
2. Disaster mode: Disaster Execution Environment (DEE)

If the dual-mode beacon profile is in NEE, there is a smart device close to it, the beacon will send some information about the product introduction to the smart device. When the items are on sale the beacon can also send some discount coupons or shop event messages to smart devices. The defined byte is set to 00. If the dual-mode beacon profile is in DEE, the defined byte is set to 01 to represent the beacon is in Pre-disaster state1 of fire. At this time DTPM11 temperature sensor perceives suspected combustible material with temperature increase and shows that the possibility of fire is very high. Pre-disaster state2 is represented by 03, indicating the target temperature has exceeded 40 degrees. It can be sure that the fire has already happened. The beacon will be in disaster state3 after ten minutes in the first state and indicates the escape route to the application's users. Disaster state4 will be activated in disaster state3 after 60 minutes and be used to help firefighters determine the location of trapped persons.

The defined byte is set to 09 to represent the beacon is in Pre-disaster state1 of earthquake. At this time acceleration sensor senses a slight vibration on the ground. Pre-disaster state2 is represented by 11, indicating the acceleration has exceeded 40 degrees. It can be sure that the earthquake has already happened. The disaster state3 and state4 of earthquake are similar to fire.

Table 2. Normal and Disaster mode

Services	Normal mode	Disaster mode			
		Pre-disaster		Disaster	
		State1	State2	State3	State4
Shopping	Buy the items				
Advertisement	Shop events				
Exits Broadcast				Disaster level 1 (10 min)	Disaster level 1 (10 min)
Confirm Survivors					Disaster level 2 (30 min)
Rescue Survivors					Disaster level 3 (60 min)

3.2 Dual-mode Advertising packet

We define the last byte of the advertising packet to dual mode beacon. Its first bit is called M bit which is used to set the beacon mode: when the value of M bit is 0, it means that the beacon is in the normal mode; when it is 1, it means that it is in a disaster mode. The second bit and third bit are used to describe the current status of the disaster (a disaster is going on, or it happened a few minutes ago).

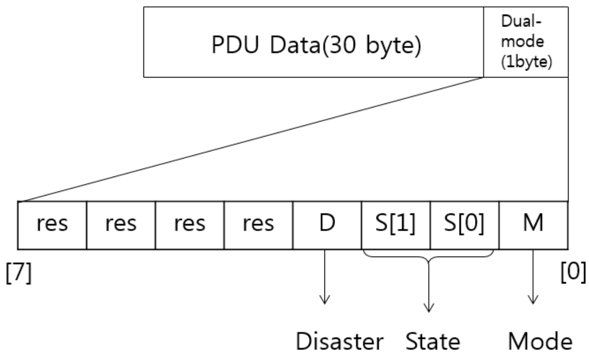


Fig. 3. Dual-mode advertising packet

We use the fourth bit to indicate that whether it is a fire disaster or an earthquake. T he next four bits are reserved bits.

Table 3. Dual-mode (1 byte)

Bits	Description
Mode[0]	0: Normal, 1: Disaster
State[2:1]	00: State1, 01: State2, 10: State3, 11: State4
Mode[3]	0: Fire, 1: Earthquake
Reserved[4:7]	Reserved

If the modified beacon is in NEE (Normal Execution Environment) state and there is a smart device close to it, the beacon will send some information about the product introduction to the smart device. When the items are on sale the beacon can also send some discount coupons or shop event messages to smart devices. The defined byte is set to 00. The defined byte is set to 01 to represent the beacon is in Pre-disaster state1 of fire. At this time DTPM11 temperature sensor perceives suspected combustible material with temperature increase and shows that the possibility of fire is very high. Pre-disaster state2 is represented by 03, indicating the target temperature has exceeded 40 degrees. It can be sure that the fire has already happened. The beacon will be in disaster state3 after ten minutes in the first state and indicates the escape route to the application's users. Disaster state4 will be activated in disaster state3 after 60 minutes and be used to help firefighters determine the location of trapped persons. The defined byte is set to 09 to represent the beacon is in Pre-disaster state1 of earthquake. At this time acceleration sensor senses a slight vibration on the ground. Pre-disaster state2 is represented by 11, indicating the acceleration has exceeded 40 degrees. It can be sure that the earthquake has already happened. The disaster state3 and state4 of earthquake are similar to fire.

Advertising Data for Normal mode	Advertising Data for Fire state4
AdvData 02 01 04 1B FF 01 00 02 15 01 12 23 34 45 56 67 78 89 9A AB BC CD DE EF F0 01 02 03 04 C3 00	AdvData 02 01 04 1B FF 01 00 02 15 01 12 23 34 45 56 67 78 89 9A AB BC CD DE EF F0 01 02 03 04 C3 07

Fig. 4. Each mode packets

3.3 Disaster Detection Algorithm

The dual-mode beacon not only has the function of sending an advertisement, but also can sense the fire and earthquake to send out alarms. When a dual mode beacon does not detect any abnormality, it will enter the normal state. However, if it detects fires or earthquakes, it will switch to a disaster mode. When beacon is in disaster mode, the pre-disaster state is divided into four types: perceived state, prepare for evacuation state， evacuation state, immediately evacuate state. The fourth states are used to detect whether there are people in disaster area when the disaster occurred 30 minutes later. Dual mode beacon can identify the type of disaster that it is a fire or an earthquake by several sensors.

Table 4. Calculating function for target temperature

	Calculating
Temperature	high byte + low byte = high byte (0x6D) + low byte (0x1) = 0x16D = 365(decimal) = 36.5 ℃
Temperature of the object surface	high byte + low byte = high byte (0x00) + low byte (0xFA) = 0x00FA = 250(decimal) = 25.5 ℃

4 Performance

For analyzing the possibility of the IoT-based a dual-mode beacon with indoor positi oning system, several tests have been done. To verify our suggestion, the experiment e nvironment is as follows:

- Ubuntu: 12.04.1
- MySQL for the database server: 5.5.40-0
- R version 3.2.0 (2015-04-16) for the data mining server
- IoT Device: BLE nrf51822
- Smart Device: LG G Pad 7.0
- AndroidOS: 4.4.2 and IoT beacon: 7 ea

We designed two hypothetical scenarios to test the system.

The first hypothetical scenario is Exits Broadcast when the situation is Disaster level 1. All the survivors escaped from the disaster areas by the exit in C1 area. The beacon called F7C1 will broadcast the escape message to guide survivors. Figure 6 shows the result of the data mining for this situation.

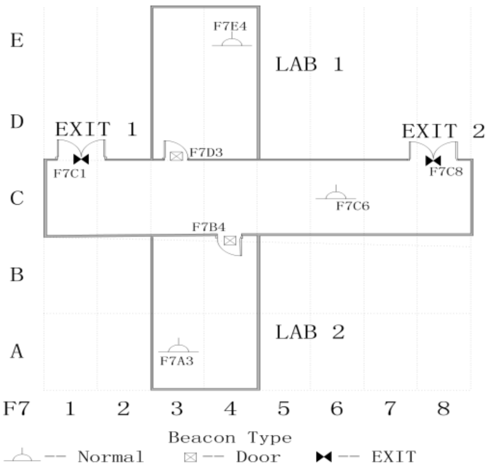


Fig. 5. Sketch map of the experiment scene. The experiment has been done in our lab oratory. There are two lab rooms and two exits at the both end of the corridor. 7 beaco ns were deployed to cover the experiment space. These beacons are divided into three types: normal, door and exit.

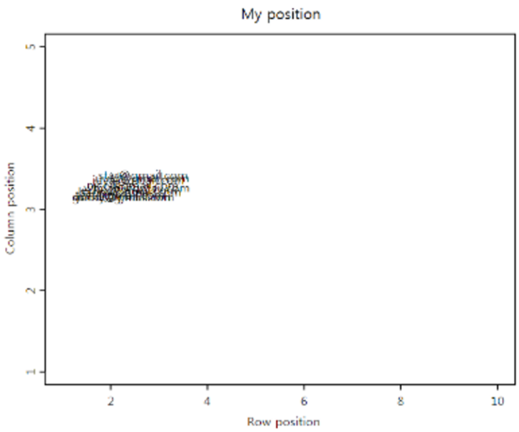


Fig. 6. The result of data mining for the first hypothetical scenario

The second hypothetical scenario is walking through every beacon shown in Figure 7: F7A3→F7B4→F7C6→F7C8→F7C6→F7D3→F7E4→F7D3→F7C1

Users must register with their own e-mail address so that we can know who are ther e. Figure 8 shows the result of the data mining. The walked location data of the experi menter whose email is “good@gmail.com” is shown on the figure. So that we can get the people's location and also confirm whether there is any people in the disaster area s.

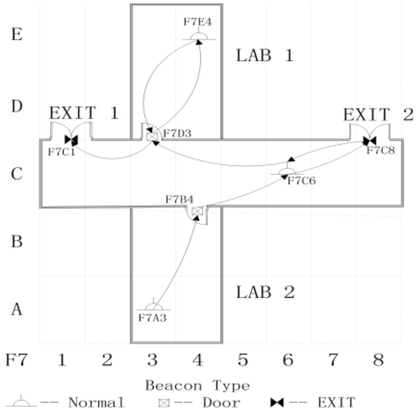


Fig. 7. Sketch map of the experiment scene. Tracing user's movement at F7A3→F7B 4→F7C6→F7C8→F7C6→F7D3→F7E4→F7D3→F7C1

Data mining server can calculate the value which can be consider as a person location inside the building[3].

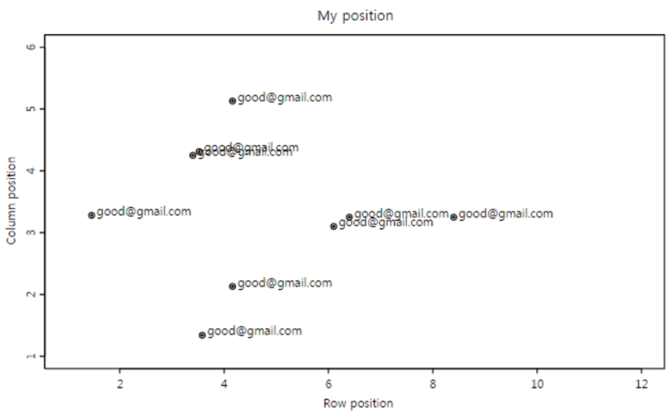


Fig. 8. The result of data mining for the second hypothetical scenario

5 Conclusion

In this paper we described an emergency evacuation system with indoor positioning when disaster suddenly occurred in the buildings. A dual-mode beacon profile has been designed to fit the normal or the disaster environment. By detecting the beacons' BLE signal strength, mobile device send the information of the nearest beacon to the database server. The data mining server use these data to broadcast the exits and report the position or the walking path. The major facts of this system are: Strong stability, Low power consumption, accurate positioning, the disaster levels situation, and normal and disaster environments detection. For the future our work shall be focused on development of the more nice escape algorithms for the DEE and development of the advanced disaster detection algorithm when cause DEE.

6 Reference

1. Hun Choi, Hyeon-Deok Bae: Design and Implementation of Seismic Data Acquisition System using MEMS Accelerometer, The Transactions of the Korean Institute of Electrical Engineers Vol. 61, No. 6, pp. 851 ~ 858, 2012

2. Sugata Sanyal, Ajith Abraham, Dhaval Gada: Security Scheme for Distributed DoS in Mobile Ad Hoc Networks, School of Technology and Computer Science, Tata Institute of Fundamental Research, India.

3. Sriharsha Kuchimanchi : Bluetooth Low Energy Based Ticketing Systems, AALT O UNIVERSITY School of Electrical Engineering, February 2015.

4. Apple's iBeacon: Location via Bluetooth 4.0, http://www.dailywireless.org/2013/12/04/apples-ibeacon-bluetooth-4-0-for-location/, Sam Churchill, December 4th, 2013.

5. Richter magnitude scale, https://en.wikipedia.org/wiki/Richter_magnitude_scale, WIKIPEDIA

6. Ki-Yun Kim, Min-Woo Lee, Jae-Sang Cha: Analysis of Indoor Localization Technology for Exhibition Services and Application Method of VLC/RF Converged Localization, The Journal of Korean Institute of Communications and Information Sciences '15-01 Vol.40 No.01

7. Myungin Ji, Jooyoung Kim, Juil Jeon, Youngsu Cho: Analysis of Positioning Accuracy corresponding to the number of BLE beacons in Indoor Positioning System, July 1-3, 2-15 ICACT2015

8. Mi-Young Bae, Dae-Jea Cho: Design and Implementation of Automatic Attendance Check System Using BLE Beacon, International Journal of Multimedia and Ubiquitous Engineering. Vol.10, No.10(2015), pp.177-186

Energy-Efficient Activity Monitoring System Using a Wearable Acceleration Sensor

Ik-Hyun Youn[1], Sangil Choi[2], Jong-Hoon Youn[1]

[1] Department of Computer Science, University of Nebraska at Omaha, Omaha, USA
{iyoun, jyoun}@unomaha.edu
[2] Department of Computer Science, Swaziland Christian University, Mbabane, Swaziland
sgilchoi@gmail.com

Abstract. As people pay more attention to their health issues, different types of human activity monitoring systems are emerging in the market. Many researchers have proposed various accelerometer sensor- based mobility monitoring systems. However, the energy efficiency of wearable activity monitoring systems has not been well studied. In this paper, we develop and test an application-level solution for achieving energy savings in a human daily activity monitoring system using a wearable wireless sensor. All functionalities including data processing, activity classification, wireless communication, and storage of classified activities are implemented in a single sensor without degrading the classification accuracy of the activities. Based on the experimental protocol with five major physical activities, the system achieves an average of 98 percent accuracy in classifying these daily activities with significant energy savings.

Keywords: Wearable sensor, energy efficiency, accelerometer, physical activity classification

1 Introduction

Wireless sensors are widely used for collecting and monitoring environmental data in areas of interest. These devices usually use wireless communications protocols such as Bluetooth for data exchanges. One of the critical research topics in developing a wearable activity monitoring system is energy efficiency. . In general, wearable activity monitoring systems use small sensors to collect human activity data and then analyze the data to produce mobility patterns of the human subjects.

Although the development of activity monitoring systems using wearable sensors has made significant progress over the past decade, some technical problems still need to be solved in order to realize the full potential of these technologies. First, most wearable sensor-based monitoring systems rely on a small battery and have a short lifetime. Second, many sensor-based activity monitoring systems focus on the data collection process without integrating data analysis into the sensor platform. As a result, limited energy resources are further wasted by exchanging data wirelessly between the sensor and data sink.

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_8

Few studies have focused on energy conservation in designing wearable activity classification systems. Battery conservation is crucial in developing a daily activity monitoring system because the system must be able to operate for several days to characterize and profile meaningful activity patterns of subjects. An ideal activity monitoring system should maintain classification accuracy for patients with mobility disorders and also address the energy consumption issues with wearable systems.

To enhance the energy efficiency of a wearable activity monitoring system, it is vital to consider the functionality of onboard data processing and efficient wireless data communication. Thus, we design a real-time, onboard data processing application to classify human activity in an energy-efficient manner. Our system collects 3-axis acceleration data, extracts three key parameters, and produces one byte of activity data every half-second. By performing the algorithm, the system dramatically reduces the amount of data to be transmitted wirelessly. We use a wearable wireless sensor called Shimmer [1] to implement the proposed system. Shimmer is a small and lightweight accelerometer sensor platform principally built for wireless body sensor networks in the health sector. We consider this gadget to be a miniature personal computer because it contains computing, sensing, and communication capabilities.

This paper is structured as follows. Section 2 surveys related research on physical activity classification and monitoring systems. Section 3 details how our system works. In Section 4, we describe the proposed activity classification algorithm. Next, we numerically analyze the energy consumption of a wearable sensor in Section 5. The results of our experimental study to measure the accuracy of the activity monitoring system are given in Section 6. Finally, conclusions are presented in Section 7.

2 System Environment

Due to limited battery resources in wearable activity monitoring devices, energy efficiency directly affects the lifetime of these monitoring systems. In this study, we focus on the energy consumption of wireless communications and the computations for activity classification.

Wireless communications typically come at the cost of high energy consumption. In addition, complex computations with high volume data streams can consume a significant amount of energy. It is widely known that packet transmission and reception are most energy-intensive processes in wireless sensors. Therefore, we focus on minimizing the demand for energy consumption for wireless data exchanges. Instead of sending raw acceleration data in the x-, y-, and z-directions, the proposed system processes raw data collected from the accelerator and produces just one byte to represent the classified activity. To minimize the energy consumption caused by computations, we focus on developing a lightweight real-time classification method. The proposed physical activity classification system uses a triaxial wearable accelerometer sensor. By attaching this sensor to the waist of a human body, the system 1) collects acceleration data from a built-in accelerometer inside the sensor; 2) classifies major human activities including walking, standing, sitting, lying, and running by analyzing the collected raw data; and 3) transmits the classified activity information to the base station through IEEE

802.15.1 wireless communication. If the base station is not within the range of the sensor's communication, the system stores activity data on a MicroSD card.

2.1 Shimmer

Shimmer is a wireless sensor platform for various types of wearable sensor applications [1]. It was originally developed for wearable healthcare sensing applications. The Shimmer device consists of integrated and extended sensors, a central processing unit, wireless communication modules, and storage devices. It can be considered as a small portable computer equipped with wireless communication and sensing capabilities. In this study, we implement the proposed system on a Shimmer2r device, which has a low-power 8MHz MSP430 CPU, 10 KB RAM, and 48 KB flash memory. The device also has a MicroSD card for secondary storage of up to 2 GB. Shimmer2r can collect and capture physiological or environmental data by interacting with different kinds of integrated or extended sensing gadgets. Typically, an MMA7361 accelerometer is integrated into Shimmer2r. A Gyroscope or magnetic sensors can be added to the Shimmer system through an extension board.

The size of raw acceleration data is 4 bytes long on each axis. After including other data such as the timestamp and sequence number, the size of one set of raw data is typically over 20 bytes. This data report is transmitted at a high rate of speed of about once every tens of milliseconds. Therefore, to minimize the energy consumption of wireless sensor nodes, we implement on-board data collection and activity classification in the Shimmer platform. The proposed approach produces only a single byte of classification data once every half-second. Preventing the wireless exchange of the raw acceleration data reduces energy consumption considerably and dramatically prolongs the lifetime of the system.

3 Algorithm Development

3.1 Data Collection

Before we process raw acceleration data from the accelerometer sensor in Shimmer, we initialize some variables related to accelerometer readings. The Freescale MMA7361 accelerometer integrated into Shimmer2r provides us a selectable sensitivity range from 1.5 g to 6 g. For the activity monitoring, we choose 4 g as a sensitivity range.

Since the sensor orientation varies from person to person, the calibration phase is important in order to improve the accuracy of the system. During the calibration phase, we measure the upper body angle (ANG_{VA}) while standing and the standard deviation of acceleration (SD_{LOW}) while walking. We also measure a standing average vertical acceleration (AVG_V) for static activities such as lying, sitting, standing and transition, and a standard deviation value of vertical acceleration (SD_V) for dynamic activities such as walking and running. These thresholds are fed into the system for accurate classification of the five target activities. This calibration phase ensures that significant errors from acceleration drift are minimized [3].

Raw acceleration data are collected every 50 milliseconds (*ms*). In Shimmer2r, the x-axis represents an anterior-posterior direction and the y-axis represents a vertical direction. We use these two axes to calculate the inclination angle. The equation for computing the inclination angle (Φ) is shown below in Equation (1), where Ax and Ay represent accelerations of the x- and y-axis, respectively.

$$\Phi = \arctan \frac{A_x}{A_y} \tag{1}$$

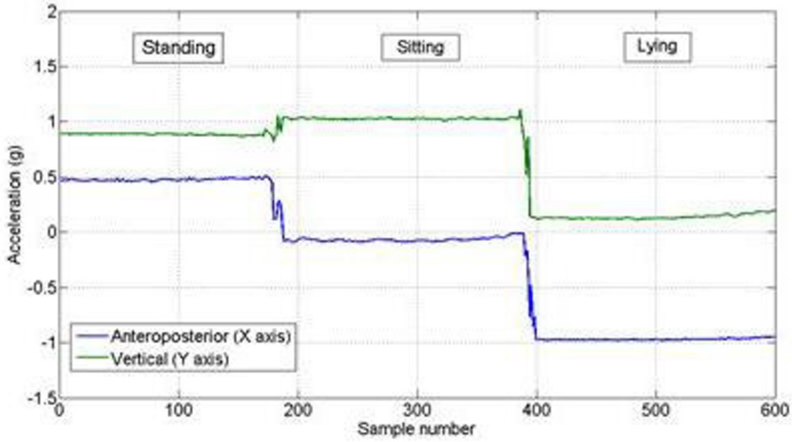


Fig. 1. Acceleration of two directions in static activities.

After attaching the Shimmer to the subjects' waists, we collect the raw acceleration data. As shown in Figure 1, there are clear differences in the ranges of the x- and y-axis acceleration values for standing and sitting. In Figure 1, the measured ANG_{VA} values for three static postures are shown. Because standing has an almost upright upper body angle, whereas sitting has relatively tilted upper body angles, standing can be distinguished from sitting and lying by measuring the upright upper body angle. The inclination angle is also useful to identify the lying status, since lying apparently has a smaller amount of vertical acceleration than other activities. AVG_V is also a useful discriminator between lying and other static activities.

To determine dynamic postures, we choose the standard deviation of the vertical acceleration (SD_V). Standard deviation is a well-known measure to distinguish between static and dynamic activities [2], [4], [9]. Most dynamic activities, including walking and running, show significant changes in vertical acceleration. The dispersion of the vertical acceleration readings is statistically calculated by the standard deviation of the y-axis. We measure the dispersion of the y-axis acceleration using Equation (2), where n is the number of samples, y_i is the acceleration of the y-axis at time i, and $\bar{y}$ represents the average of the vertical acceleration (AVG_V) over n samples, respectively.

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{2}$$

We set two standard deviation values, SD_{LOW} and SD_{HIGH}, as a threshold to identify three different levels of dynamicity – static, walking and running activities. If the measured standard deviation value is below the lower threshold (SD_{LOW}), the algorithm classifies the activity as a static activity. If the standard deviation value is between the thresholds of SD_{LOW} and SD_{HIGH}, the activity is classified as a walking activity. Otherwise, the activity is considered as running.

Figure 2 shows the measured SD_V of these activities. The first activity represents standing for 100 seconds, the next activity is walking, and the last activity is running. Note: your Y axis title in Figure 2 below has a misspelling (vertical).

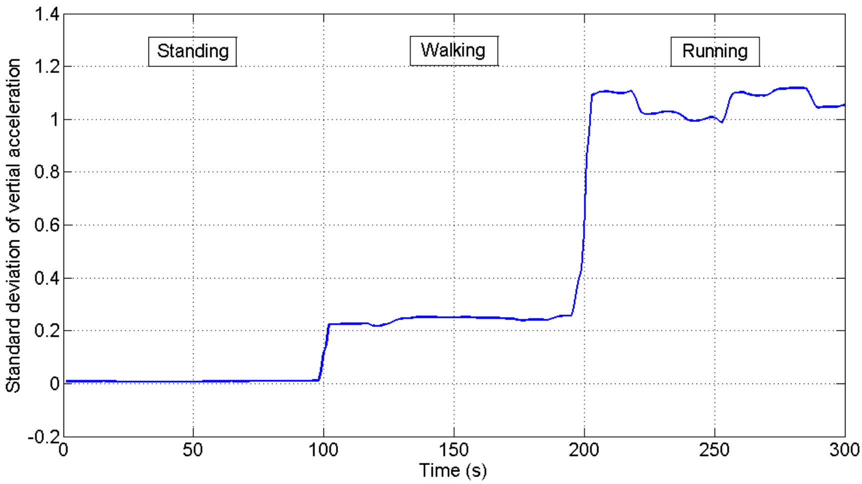


Fig. 2. Standard deviations of vertical acceleration (SD_V) while standing, walking, and running.

3.2 Algorithm

The proposed algorithm is developed for a resource-constrained and battery-operated hardware platform for real-time activity classification. An activity monitoring system is required to operate for a relatively long period, ranging from several days to a few weeks without recharging its battery. Battery conservation is always crucial to wearable sensor systems, so we focus on minimizing the use of computational power to reduce energy consumption. Note that the proposed algorithm utilizes only three parameters, and these parameters are computed sequentially. Figure 3 shows a flow chart of the proposed classification algorithm. The proposed system collects 3-axis acceleration data from the accelerometer sensor every 50 milliseconds. Each half-second, three key parameters, AVG_V, ANG_{VA} and SD_V, are computed for a total of 20 acceleration data samples.

Once the algorithm gets the parameters from 20 raw acceleration data, it starts with relatively easy activities to identify. First, if the calculated SD_V value is higher than the higher SD_V threshold (SD_{HIGH}) value, the activity is classified as running. If not, the obtained AVG_V value is then compared to the AVG threshold. Due to the Earth's gravity, the vertical acceleration of many activities except lying is near 1G ≈ 9.8 meter/sec^2. Thus, the AVG_V value for lying clearly differentiates it from other activities. If the activity is neither running nor lying, the lower SD_V threshold (SD_{LOW}) is compared to the current SD_V value. If SD_V is greater than SD_{LOW}, the activity is categorized as walking. If none of the conditions for the above three activities are met, ANG_{VA} is tested to determine whether the activity is standing or sitting. The result represents an activity for the collected 20 acceleration samples for a half second.

When classifying activities, we do not consider parameters that require heavy computation, such as skewness or SMA. Further our system mostly relies on vertical acceleration measurements; anterior/posterior accelerations are only used as needed.

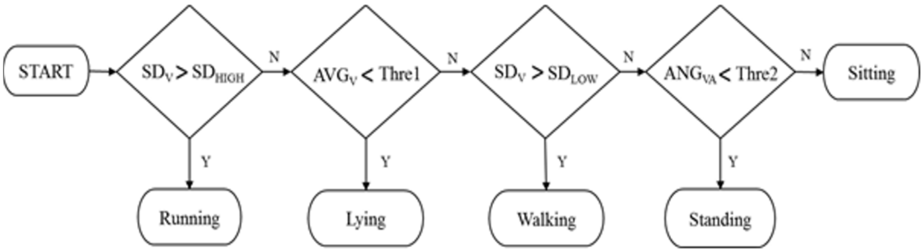


Fig. 3. Flowchart of the activity classification algorithm.

4 Numerical Analysis of Energy Consumption

In this section, we analyze the energy cost for transmitting a packet wirelessly. The proposed system collects x-, y-, and z-axis acceleration data from the accelerometer sensor, and produces one byte of human activity data every half second using the inclination angle and standard deviation. Finally, it sends the classified data to the base station. We implement all of these steps in a Shimmer2r sensor. In this paper, we call our approach on-board processing. Current wearable activity monitoring systems typically gather the acceleration data and send these data to the base station, and the activity classification algorithm runs in the base station.

Let us assume that we develop our system without adapting the on-board processing approach using Shimmer. The Shimmer device includes a tri-axial accelerometer, and the size of raw acceleration data is 4 bytes per axis. Therefore, the system needs to send at least 12 bytes (4 bytes × 3 axes) once every sampling period, and the base station performs post-processing for activity collection after receiving raw data. Assume 50 milliseconds is the sampling rate of the system. Without on-board processing, 240 bytes (20 × 12 bytes) of acceleration data have to be transmitted per second, versus 2 bytes of activity data per second with on-board processing. The post-processing data size is 120 times greater than that of on-board processing. The large difference in terms of the size of data being transmitted wirelessly could significantly affect the energy consumption of wireless sensors.

Based on the energy model in [5], we perform a numerical analysis of the post-processing and on-board processing approaches discussed above. The transmission energy cost for a packet (E_{tx}) is

$$E_{tx} = V \cdot I_{tx} \cdot S/C \tag{3}$$

In Equation (3), V, I_{tx}, S, and C represent the supply voltage, current drawn, packet size, and channel capacity, respectively. On-board processing, in particular, spends additional CPU energy (E_{cpu}) for calculating the inclination angle and standard deviation. The modified energy model (E_{board}) for on-board processing is as follows:

$$\mathrm{E_{cpu}} = \mathrm{V} \cdot \mathrm{I_{tx}}$$

$$\mathrm{E_{board}} = \mathrm{E_{tx}} + \mathrm{E_{cpu}} = \mathrm{V} \cdot \mathrm{I_{tx}} \cdot \mathrm{S/C} + \mathrm{V} \cdot \mathrm{I_{cpu}} \tag{4}$$

By reducing the size of transmitted data, we can enhance the energy efficiency of the system. For example, as the transmission time increases, the gaps between the energy consumption of post-processing and that of on-board processing increase linearly. We assume that the energy cost of wireless communication is higher than that of sensing or CPU usage.

5 Experimental Study

In this section, we evaluate the classification accuracy of the proposed system. Five protocols were selected for this evaluation study. This experiment was performed indoors with twenty healthy subjects. The subjects followed a given activity schedule as shown in Table 1. Each subject performed five different activities for 3 minutes each. Before shifting to the next activity, an observer alerted the subjects in order to achieve a smooth and quick activity change. Moreover, in the protocols, each activity included a few detailed directions while maintaining that particular activity. For example, when a subject was sitting on a chair, the subject was required to change the natural posture from forward to backward within 180 seconds. Other activities also had detailed activity descriptions for subjects.

Table 1. Experimental protocol.

No.	Activity	Description	Time (seconds)
1	Lying	Lying on back, lying on stomach	180
2	Sitting	Sitting back, forward, and upright	180
3	Standing	Turning trunk while standing	180
4	Walking	Walking various speeds	180
5	Running	Running on treadmill	180
Total duration			900

Table 2 and Figure 5 present the results of the experiments. The average accuracy of the system was above 97% for all five activities across the twenty subjects. Based on the accuracy of each activity in Figures 5, lying and running activities were precisely identified for every subject, whereas about 10 percent of the standing activity was misidentified. The main reason for the error in standing activity is that some ambiguous upper body postures were classified as sitting. This is because both sitting and standing have similar acceleration patterns in terms of the standard deviation and body inclination angle.

Table 2. Experimental results.

Activity	Statistics		
	Min.	Max.	Avg.
Lying	98 %	99 %	98 %
Sitting	92 %	100 %	97 %
Standing	89 %	100 %	97 %
Walking	93 %	100 %	98 %
Running	98 %	100 %	99 %

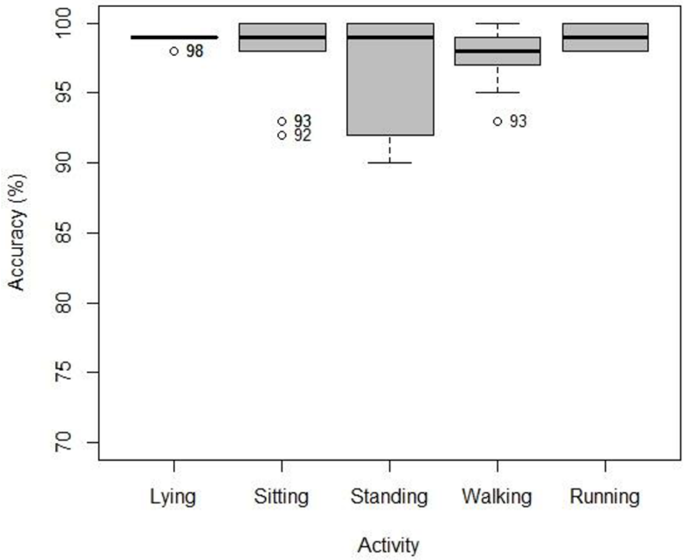


Fig. 4. The accuracy of activity classification.

6 Conclusions

We propose an on-board processing algorithm for classifying human daily activities using a single accelerometer. The proposed system is implemented on a resource-constrained wearable sensor for better portability and usability for day-to-day activities. The system transmits classified activity information instead of raw acceleration data by enabling classification processing on the wearable sensor system. The compact algorithm and its implementation on the sensor platform improve power efficiency by reducing the size of the data packet to be transmitted. We also evaluated the energy efficiency of the proposed system by analyzing the energy consumption of the system based on the energy consumption model in the literature and the sensor specifications. The results of our theoretical evaluation show a great deal of improvement in energy efficiency. The results of the experimental study also demonstrate a high degree of classification accuracy (about 98 percent on average) for five major daily activities.

7 References

1. Shimersensing, `http://www.shimmer-research.com.`
2. K. M. Culhane, G. M. Lyons, D. Hilton, P. A. Grace, and D. Lyons, "Long term mobility monitoring of older adults using accelerometers in a clinical environment," Clinical Rehabilitation, vol. 18, no. 3, pp. 335-343, Mar. 2004.
3. G. Hache, E. Lemaire, and N. Baddour, "Development of a Wearable Mobility Monitoring System," in Proceedings of Canadian Medical and Biological Engineering Conference, May 2009.
4. G. M. Lyons, K. M. Culhane, D. Hilton, P. A. Grace, and D. Lyons, "A description of an accelerometer-based mobility monitoring technique," Medical Engineering and Physics, vol. 27, no. 6, pp. 497-504, Jul. 2005.
5. M. Quwaider, J. Rao, and S. Biswas, "Body-Posture-Based Dynamic Link Power Control in Wearable Sensor Networks," IEEE Communications Magazine, July 2010.
6. P. H. Veltink, H. B. J. Bussmann, W. De Vries, W. L. J. Martens, and R. C. Van Lummel, "Detection of static and dynamic activities using uniaxial accelerometers," IEEE Trans. Rehabilitation Engineering, vol. 4, no. 4, pp. 375-385, Dec. 1996.

Multiweighted Petri Net Based Control in Indoor Navigation

Gairatzhan Mavlankulov[1], Laula Zhumabayeva[2], Mohamed Othman [1], Tamara Zhukabayeva[2], Zhanat Abdugulova[2]

[1] Department of Communication Technology and Network, Universiti Putra Malaysia 43400 UPM Serdang, Selangor D.E., Malaysia,
gairatjon@gmail.com, mothman@upm.edu.my

[2] Department of Information Technologies, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan,
Lau_la@mail.ru, tamara.kokenovna@gmail.com , janat_6767@mail.ru

Abstract. In this paper we use Multiweighted Petri Net, which is an extension of Petri Net, to model Indoor Navigation System. The core of the problem is to find the path from one object to another satisfying several requirements including those defined by the user.

Keywords: **P**etri Nets, Multi-weighted Transitions, Parallel Computing, Indoor Navigation

1 Introduction

Petri nets were introduced by Carl Adam Petri are a powerful modeling formalism in computer science, system engineering and many other disciplines. Petri nets combine a well defined mathematical theory with a graphical representation of the dynamic behavior of systems. The theoretic aspect of Petri nets allow precise modeling and analysis of system behavior, while the graphical representation of Petri nets enable visualization of the modeled system state changes. This combination is the main reason for the great success of Petri nets.[1]

Moreover, Petri nets allow to describe and analyze the flow of information and control in concurrent systems, which have widely been used in the study of formal languages[2]. In [3, 4, 5] have introduced different variants of a Petri net controlled grammar, which is a context-free grammar equipped with a Petri net, whose transitions are labeled with rules of the grammar or the empty string, and the associated language consists of all terminal strings which can be derived in the grammar and the sequence of rules in every terminal derivation corresponds to some occurrence sequence of transitions of the Petri net which is enabled at the initial marking and finished at a final marking of the net, as mathematical models for the study of concurrent systems appearing in systems biology and automated manufacturing systems. The distinguished feature of all of these variants is that the transitions of a Petri net fire sequentially. In [6,7,8,9] were introduced and extended context free concurrent grammars, grammars controlled by context free

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_9

Petri nets under parallel firing strategies, i.e., the transitions of a Petri net fire simultaneously in different modes. An important requirement of navigation systems is the ability to find optimal path for users. Optimal paths can be the least time-consuming paths, the shortest path, or paths according to user requirements. In outdoor space, they are often calculated on the basis of a path network. However, constructing a network that can support shortest path search in indoor space is more challenging. Although there are some comparable concepts in indoor space and outdoor space like corridors and roads, in indoor space, there are also concepts like rooms and lobbies for which we do not find counterparts in outdoor space. For example, inside a room, there are many implicit paths from one location to another. This makes it difficult to construct path networks for indoor space. The problem of finding the path with some requirements between objects is reduced to finding a multiweighted string of fired transitions of a Petri net. The places of a Petri net represent the objects/points, its transitions represent the direct connections between objects. Since in Petri nets, edges are replaced with transitions, different edges from one object to many objects can be reduced into one transition. The arcs and tokens of a Petri net are labeled with a vector of weights corresponding to distances, cost, time, probability, etc.

1.1 Grammars and languages

Let $\boldsymbol{N}$ be the set of all non-negative integers and $\boldsymbol{N^k}$ be the set of all vectors of $\boldsymbol{k}$ non-negative integers. The cardinality of a set $\boldsymbol{X}$ is denoted by $|\boldsymbol{X}|$**.** Let $\boldsymbol{\Sigma}$ be an alphabet which is a finite nonempty set of symbols. A string over the alphabet $\boldsymbol{\Sigma}$ is a finite sequence of symbols from $\boldsymbol{\Sigma}$. The empty string is denoted by λ . The set of all strings over the alphabet $\boldsymbol{\Sigma}$ is denoted by $\boldsymbol{\Sigma}^*$. A subset of $\boldsymbol{\Sigma}^*$ is called a language. The length of a string w, denoted by $|\mathbf{w}|$, is the number of occurrences of symbols in w. The number of occurrences of a symbol a in a string w is denoted by $|w|_a$. A context-free grammar is a quadruple $\mathbf{G=(V,\Sigma,S,R)}$ where $\mathbf{V}$ and $\boldsymbol{\Sigma}$ are the disjoint finite sets of nonterminal and terminal symbols, respectively, $S \in V$ is the start symbol and $R \subseteq V\times(V\cup \Sigma)^*$ is a finite set of (production) rules. Usually, a rule **(A, x)** is written as $\mathbf{A\rightarrow x}$. A rule of the form $\mathbf{A\rightarrow \lambda}$ is called an erasing rule. $x \in (V\cup \Sigma)+$ directly derives $y \in (V \cup \Sigma)^*$, written as $x \Rightarrow y$, iff there is a rule $r = A \rightarrow \alpha \in R$ such that $x = x_1Ax_2$ and $y = x_1\alpha x_2$ The rule $r : A\rightarrow \alpha \in R$ is said to be applicable in sentential form x, if $x = x_1 A x_2$, where $x_1,x_2\in(V \cup \Sigma)^*$. The reflexive and transitive closure of $\Rightarrow$ is denoted by $\Rightarrow^*$. A derivation using the sequence of rules $\pi = r_1 r_2 \cdots r_n$ is denoted by $\Rightarrow\pi$. The language generated by G is defined by

$$L(G) = \{w \in \Sigma^* \mid S \Rightarrow^* w\}.$$

1.2 Petri Nets Controlled Grammars

A Petri net is a triple (P,T,δ) where P and T are finite disjoint sets of places and transitions, respectively, a mapping $\delta : T \rightarrow P^{\oplus}\times P^{\oplus}$ is a mapping which assigns to each transition $t \in T$. Graphically, a Petri net is represented by a bipartite directed graph with the node set $P \cup T$ where places are drawn as circles, transitions as boxes. For each transition $t \in T$ with $\delta=(\alpha,\beta)$, the multiplicities $\alpha(p)$, $\beta(p)$ of a place $p \in P$, give the number of arcs from p to t and from t to p, respectively. A multiset $\mu \in P^{\oplus}$ is called a marking . For each $p \in P$, μ (p) gives the number of tokens in p. Graphically, tokens are drawn as small solid dots inside circles. A place/transition net (p/t net for short) is a quadruple $N =(P,T,\delta,\mu_0)$, where (P,T,δ) is a Petri net, $\iota \in P^{\oplus}$ is the initial marking . A transition $t \in T$ with $\delta(t) = (\alpha,\beta)$ is enabled at a marking $\mu \in P^{\oplus}$ if and only if $\alpha \sqsubseteq \mu$. In this case t can occur (fire). A labeled Petri net is a tuple $K=(\Delta,N,\gamma,F)$, where Δ is an alphabet, $N=(P,T,\delta,\iota)$ is a p/t net, $\gamma : T \to \Delta \cup \{ \lambda \}$ is a transition labeling function and $F \subseteq R(N,\iota)$. The labeling function γ is extended to occurrence sequences in natural way, i.e., if $\nu t \in T^*$ is an occurrence sequence then $\gamma(\nu t) = \gamma(\nu)\gamma(t)$ and $\gamma(\lambda)=\lambda$. For an occurrence sequence $\nu \in T^*$, $\gamma(\nu)$ is called a label sequence.

A timed Petri net is a six-tuple $N = (P, T, A, w, M_0, f)$, where (P, T, A, w, M_0) is a marked Petri net $f : T \to R+$ is a firing time function that assigns a positive real number to each transition.

2 Constructing and Navigating with Multiweighted Petri Net.

Navigation is a process which successfully leads object from one place to another. In general, it is needed to find the optimal paths to destinations, which can be the shortest paths with respect to time, the shortest path with respect to distance, the path without paying fees, or any other paths according to users' requirements. The optimal path with respect to distance can be obtained by applying the shortest path algorithm on a path network that reflects the global path information of the entire indoor space.

2.1 Constructing Multiweighted Petri Net

One of the important matter of the path finding process is the aspect of accessibility of architectural cells in the indoor space. For instance, while an employee in a building may have access to certain office rooms, in the same time these rooms are may be inaccessible for guests (customers). Another example is a construction site that prevents people from walking through a corridor and forces them to bypass it. We include accessibility attributes to both access points and path segments in our model. The accessibility in an access point is controlled by a time attribute in the weight of Petri net. Assume the accessibility of the doors of the building in Figure 1. are given in Table1

Table 1. Accessibility of Doors of the indoor building

Door #	Accessibility		
	Time	Status	Access card
D1	10:00-16:00	Employee	Yes
D2	9:00-17:00	Employee	No
D3	10:00-16:00	Employee	Yes
D4	9:00-17:00	Employee	No
D51	9:00-17:00	Employee	Yes
D52	9:00-15:00	Employee	No
D6	10:00-16:00	Customer	No
D71	9:00-17:00	Employee	No
D72	11:00-15:00	Customer	No
D8	10:00-16:00	Employee	No
D9	10:00-16:0	Employee	Yes
D10	9:00-17:00	Customer	No

Taking an example from Figure1 and Table 1, we can assume the accessibility of the doors “D10” and “D6” are “9:00-17:00” and “10:00-16:00” correspondingly. Therefore, customers can enter directly to room6 using D10 and D6 only during this period of time. Beside these doors employees can enter to room6 by using D51 and D52 during the time period “9:00-17:00”. The relationship between the accessibility of access points and the accessibility of path segments is stated in the following statements.

Statement 1: If an access point is inaccessible, all its emanating path segments are inaccessible, and vice versa.

Statement 2: If an access point is accessible, there is at least one emanating path segment that is accessible, and vice versa.

Statement 3: If two end points of a path segment are accessible, the path segment can be inaccessible.

Statement 1 and Statement 2 are obvious. If an access point is inaccessible, then it does not make sense that any of its incident path segments is accessible since the access point can never be reached. An accessible access point must have at least one path segment that leads to it. Statement 3 indicates that a path segment can be blocked although its two end points are accessible due to other paths traversing them.

2.2 Navigating with Multiweighted Petri Net

A *Multiweighted Petri net* is a tuple $K=(\Delta, N, \gamma, \theta)$, where Δ is a vector (in our case a vector created by set of attributes like distance, accessibility, privilege and so on) , $N=(P,T,\delta,\iota)$ is a p/t net, $\gamma : \iota \rightarrow \Delta \cup \{ \lambda \}$ is a marking labeling function and $\theta : \alpha(p,t) \rightarrow \Delta \cup \{ \lambda \}$ is a weight of the arc comes from p to t. It is clear, multiweighted Petri net with $\Delta \in N$ is a classical marked Petri net. A transition $t \in T$ with $\delta(t) = (\alpha, \beta)$ is enabled at a marking μ if and only if $\alpha \sqsubseteq \mu$. In this case t can occur (fire). As α, μ are vectors $\alpha=(\alpha_1, \alpha_2, \ldots \alpha_n)$ and $\mu=(\mu_1, \mu_2, \ldots \mu_n)$ $\alpha \sqsubseteq \mu$ means that α_i

$\sqsubseteq \mu_i$ ($0 \leq i \leq n$), where n is number of attributes. If t fire then this occurrence transforms the marking μ into the marking μ'. μ' is also vector $\mu'=(\mu'_1, \mu'_2, ...\mu'_n)$ defined by $\mu'_i=\omega'_i(\alpha, \mu)$, where ω'_i is a function of α and μ.

In indoor space, rooms, corridors and lobbies, are considered as the basic units which are called cells. In fact, in order to find out detailed path information, it is insufficient to only consider the sequence of the visited cells. Moreover, we can notice that there are some implicit path in indoor space which are commonly used by people.

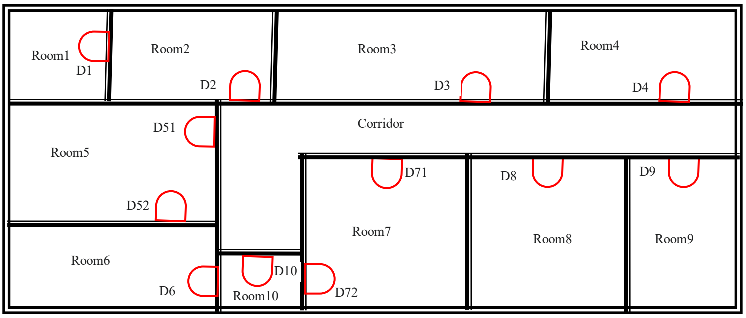


Figure 1. Sample Map of the indoor building

We note, there are a number of different kinds of architectural cells in indoor space. Some of them have similar shapes but may serve different purposes, and some of them are totally different in shapes but may purposed the same role during the path. For example, rooms with multiple doors can be a part of a passage to a certain destination, while rooms with only one door cannot. Thus, we classify cells into different categories according to their geometric and architectural features from the path perspective.

The notion of distance could be understood in a variety of different ways depending on the attributes. Here some examples :

– Doors can be locked or, more general, access requires authorization (key, card, biometric scan etc.).

– Admission of entry or exit can be limited in time, e.g. office hours of an office.

– Certain sections of a public building may be restricted in access (“staff only”, high-security wings, laboratories).

– Special exits and base level windows can be used for emergencies.

Moreover, there are person-related properties like roles, privileges, or preferences also have a significant impact on navigation.

The geometric data structure consists of polygons. For the purpose of navigation, we must define connectivity between the separated polygons (rooms, floors, etc.). Even though two adjoining separated polygons are touching, they can still be physically separated by walls or other divisions. Enter and exit are only possible through specific points on their shared boundary (i.e. doors and other openings), called boundary nodes. Therefore, an optimal path of the object depend on its attributes(has key or nor, staff or not, etc). We model such requirements in multiweighted nets as weights in arcs.

3 Conclusion

In this paper we defined a notion a *Multiweighted Petri net* which allows to simplify of model of the Navigation in complex spaces(in our case in indoor space). The concurrency feature of a Petri net allows to search for multiweighted strings parallelly, which reduces the time complexity of an algorithm significantly. In this work we used deductive and constructive methods to solve stated problems. We showed that proposed our model of Indoor Navigation System by special Petri Nets simplify and optimize control and finding the path from one place to another satisfying requirements of the system.

References

1. Peterson, J.L., Petri net theory and modeling of systems. Prentice-Hall, Englewood Cliffs, NJ, (1981)
2. Hack, M., Petri net languages. Computation Structures Group Memo, Project MAC 124, MIT, Cambridge Mass., (1975)
3. Dassow, J. and Turaev, S., k-Petri net controlled grammars. In C. Mart´ın-Vide, F. Otto, and H. Fernau, editors, Language and Automata Theory and Applications. Second International Conference,LATA 2008. Revised Papers, volume 5196 of LNCS, Springer, (2008) 209–220
4. Dassow, J and Turaev, S., Grammars controlled by special Petri nets. In A.H. Dediu, A.-M. Ionescu, and C. Mart´ın-Vide, editors, Language and Automata Theory and Applications, Third International Conference, LATA 2009, volume 5457 of LNCS, Springer, (2009) 326–337
5. Dassow, J and Turaev, S., Petri net controlled grammars: the power of labeling and final markings.Romanian Jour. of Information Science and Technology, 12(2): (2009) 191–207.
6. Mavlankulov, G., Othman, M., Selamat, M.H. and Turaev, S., Concurrent context-free grammars. In First International Conference on Advanced Data and Information Engineering, LNEE, Springer, (2014) 521–528.
7. Mavlankulov, G., Othman, M., Selamat, M.H. and Turaev, S., Some properties of the concurrent grammar. In International Conference on Mathematical Sciences and Statistics, Springer, (2014) 223–231
8. Zhumabayeva, L., Zhukabayeva, T., Mavlankulov, G. and Othman, M.,. Arbitrary Petri Nets Controlled Grammars Under Parallel Firing Strategy, 5th International Conference on IT Convergence and Security (ICITCS), IEEE, (2015) 1-3
9. Mavlankulov, G., Turaev, S., Zhumabayeva, L. and Zhukabayeva, T., Parallel firing strategy on Petri nets: A review, International Conference On Mathematics, Engineering And Industrial,AIP Publishing, (2015)

Using GSPNs for Performance Analysis of a New Admission Control Strategy with Retrials and Guard Channels

Nawel Gharbi

Department of Computer Science,
University of Sciences and Technology USTHB,
Algiers, Algeria
ngharbi@usthb.dz

Abstract. The Small Cell Networks (SCNs) based on the idea of employing much smaller size cells cause the mobile subscribers to cross several cells during an ongoing conversation resulting in frequent handovers. Most of works dealing with Call Admission Control (CAC) problems based on guard channels scheme in cellular mobile networks, consider models without retrials (repeated calls). On the other hand, almost all existing works which take into account retrial phenomenon, consider a multi-servers retrial queueing model for which the analytical solution is not available. In this paper, we propose to use the Generalized Stochastic Petri Nets (GSPNs) formalism as a novel alternative for modeling and evaluating performances of a new CAC strategy with retrial policy of fresh and handover calls generated by a finite source of subscribers, based on prioritization of retrial calls and adopting a guard channels borrowing concept, operating in SCNs. A performance study shows that our approach is more efficient compared to basic guard channel scheme with better blocking probability, especially for low to medium handover calls rates.

Keywords: Small Cell Networks, Call Admission Control, Retrials, Guard Channels, Generalized Stochastic Petri nets, Performance Measures.

1 Introduction

The ever-increasing number of subscribers and the need for higher data rates and multimedia services require the deployment of *Small Cell Networks (SCNs)*, which represent a novel networking paradigm based on the idea of deploying short-range, low-power, and low-cost base stations in cells where the size gets smaller and thus the number of subscribers served in a cell will be relatively smaller, such that traffic models with a finite source of subscribers should be considered. The concept of small cell is available in almost all cellular network technologies such as GSM, UMTS, LTE and WIMAX. These small cells operate in spectrum that have a range of 10 to 200 meters, compared to a mobile macrocell which might have a range of a few kilometers. As a consequence, the

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_10

mobile subscribers cross several cells during an ongoing conversation resulting in frequent handovers. As dropping a call in progress is more annoying than blocking a fresh call request for subscribers, handovers calls are typically given higher priority than fresh calls in access to the wireless resources by implementing admission control policies. *Call admission control (CAC)* restricts the access to some network resources in order to prevent network congestion and service degradation for already accepted calls [1]. The most popular CAC approach to prioritize handover calls is based on reserving a portion of available channels (called *guard channels*) in each cell to be used exclusively by handover calls. The remaining channels (called *ordinary* or *common channels*) are shared both by fresh and handover calls [2]. Furthermore, the consideration of subscribers behavior in the performance analysis of cellular mobile networks and in particular, the repeated attempts (called *repeated calls* or *retrials*) of subscribers whose call was refused due to the lack of available resources is crucial to determine the network performances because they can have a significant negative impact on the *quality of service (QoS)* offered to subscribers and should therefore not be neglected in network design and planning.

Most of works dealing with CAC problems based on guard channels scheme in cellular mobile networks consider models without retrials (see [3–5]). This is essentially due to the difficulty to obtain explicit formulae for models with retrial phenomenon and multiple resources [6]. Since the study of Tran-Gia and Mandjes [7], provided a CAC model with only one guard channel and demonstrated that in the context of cellular networks, the retrial phenomenon is no longer negligible because of the significant negative influence on the network performances, the consideration of that phenomenon is more and more being undertaken. Recently, CAC models with guard channels and retrials have been extensively studied, see for example [7–11]. Moreover, almost all works on cellular networks consider an infinite population cell and generally use a multi-servers retrial queue to represent the cell behavior. In [8], CAC models with finite population of fresh calls and infinite population of handover calls were considered. Further, *Molina et al.* in [12] have proposed a novel approach using *Generalized Stochastic Petri nets (GSPNs)* as a new performance evaluation tool to study CAC policies in cellular mobile networks. GSPNs [13] offer a flexible graphical model and it was proved to be an excellent and very effective mathematical tool for modeling and analyzing the dynamics of cellular networks compared to retrial queues. More recently, some papers of *Gharbi* (see [14, 15]) study the impact of repeated calls on the performances of SCNs by means of GSPNs formalism. However, no CAC policies models with retrials and finite population cells have been provided in the literature for prioritizing handover calls in cellular networks, in particular, for guard channels schemes with retrials using GSPNs formalism.

On the other hand, the existing guard channels schemes often reduce the handover calls blocking probability at the expense of increasing the fresh calls blocking probability and reducing the total carried traffic in cell. Moreover, in certain situations where subscribers mobility is relatively limited and handovers are less frequent, for example in certain periods of day and non-peak hours, it is possible that fresh calls be blocked because of lack of available ordinary channels, while guard channels are unoccupied. In this case, cell resources are wasted by serving neither the handover requests nor the fresh calls requests. Here we are face with a bad exploitation of the available base station resources. With the motivation of overcoming these two problems, we propose in this paper a novel CAC strategy for SCNs, where subscribers population in each cell is considered finite, for both of fresh and handover calls, based on the concept of *guard channels borrowing* and the priority for the blocked fresh calls retrial, called *Priority for Fresh calls Retrial based on Guard Channels Borrowing Scheme (PFR-GCBS)*, using the GSPNs formalism. The priority allocated to blocked fresh calls retrial can be justified by the fact that this kind of calls came the first into the system and are always applicant and waiting for service.

The rest of the paper is organized as follows: First, we give an overview of syntax and semantics of GSPNs formalism. In Section 3, the proposed PFR-GCBS description is detailed. In Section 4, we present the mathematical model describing the subscribers behavior with PFR-GCBS in SCNs. Next, the GSPNs model of PFR-GCBS is developed in section 5. Next, the computational formulas for evaluating exact performance indices are derived in Section 6. Finally, based on some experimental examples, we illustrate the effect of new guard channels borrowing concept rate on the network performances.

2 Description of the proposed Call Admission Control mechanism: PFR-GCBS

The proposed CAC called Priority for Fresh call Retrial based on Guard Channel Borrowing Scheme (PFR-GCBS) considers a mobile network with finite subscribers population cells, single-class of service (voice call), fixed channel allocation and the use of guard channels scheme to prioritize handover calls with taking into consideration the repeated calls phenomenon (retrial) of both blocked fresh and handover calls by using two finite size imaginary waiting spaces called *orbits*. The principal novelty of PFR-GCBS lies in the priority allocated to retrials of blocked fresh calls compared to primary fresh calls arrivals via the incorporation of the guard channels borrowing concept. This concept consists in a light modification of the traditional guard channels semantic for which the whole set of guard channels is divided into two parts using a threshold t called threshold of borrowing: a part of t ($t < g$, g represents guard channels number) guard channels known as borrowable or shareable, are accessible by all handover calls with the possibility of being borrowed and temporary allocated to the blocked fresh calls retrials if they are unoccupied; and the other part (of g-t guard channels) called dedicated, rest exclusively reserved for handover calls.

For the mathematical model, we consider a particular cell with a finite subscribers population of size $\mathbf{L}$ ($L < +\infty$), here there are C ($C \geq 2$) channels in each base station (BS) to serve incoming subscribers calls. On the other hand, we consider two different and independent arrival traffic streams: The arrival of fresh calls (calls initiated in the cell coverage area) and the arrival of handover calls (the incoming handovers flow entering the cell). The inter-arrival times of fresh and handover calls are independent and identically-distributed random variables with rate λ_f and λ_h, respectively. Thus, the total call arrival rate is $\lambda = \lambda_f + \lambda_h$. Call duration follows an exponential distribution with rate μ.

In the proposed scheme, the priority is given to handover requests by reserving g *($g < C$)* channels (known as *guard channels*) exclusively for handover calls among the C available channels in a cell. The remaining $n = C - g$ channels (known as *common* or *ordinary channels*) are shared by fresh and handover calls. To this end, we assume that:

- An arriving fresh call which finds at least $g + 1$ idle channels (at most $C - g - 1$ busy channels) will be immediately served. Otherwise, if there are at most g idle channels (at least $C - g$ busy channels), the fresh call will be blocked and will join the orbit of fresh calls then starts generation of an exponentially distributed flow of repeated calls with probability θ_0 and rate α or gives up and leaves the system with probability (*1-* θ_0) (to model impatient subscribers). According to the CAC policy of PFR-GCBS, more priority is assigned to the retrial of blocked fresh calls that find all the ordinary channels busy, by allowing them to not only occupy the $n = C - g$ ordinary channels, but they have also the possibility to borrow and temporarily occupy an idle guard channel from the set of *'t' borrowable guard channels*. Hence, the retrial of a blocked fresh call has access to *(C-(g-t))* channels. The retrial of a blocked fresh call will be immediately served if there are at least $(g - t) + 1$ idle channels (at most *(C-(g-t))-1* busy channels) and it will be blocked again if there are at least $(C - (g - t))$ busy channels. A subscriber waiting in the orbit of fresh calls may become impatient and decide to give up and leaves the system definitively with probability (*1-* θ_0) after a finite number of unsuccessful reattempts.
- An arriving handover call which finds at least *one* idle channel (at most *C-1* busy channels) will be immediately served. Otherwise, the call will be blocked and will join the orbit of handover calls then starts generation of an exponentially distributed flow of repeated calls with rate β until it finds one idle channel with probability θ_1 or leaves the system (which represents the handover failure) with probability (*1-*θ_1), when the mobile station (*MS*) moves outside the handover zone or the signal quality degrades below the handover threshold. The retrial of blocked handover call is treated as a new arriving handover call request. We assume that the waiting time of a handover call in the handover orbit is short.

Note that θ_0 (*resp.* θ_1) is used to represent the degree of *impatience* of subscribers. We assume also that both fresh and handover subscribers retry at least once.

The CAC policy PFR-GCBS assigns the highest priority to handover calls, the blocked fresh calls retrials come after with a reasonable priority and finally the smallest priority is assigned to fresh calls.

3 GSPN model of the PFR-GCBS strategy

A GSPN consists of two kinds of nodes, called *places* (drawn as circles) and *transitions* that are partitioned into two different classes: *timed transitions* (represented by means of rectangles), which describe the execution of time consuming activities and *immediate transitions* (represented by thin bars), which have priority over timed transitions and fire in zero time once they are enabled. Immediate transitions model logic activities as synchronization. Places and transitions are connected by oriented and labeled arcs (represented by arrows). An arc connects either a place to a transition or a transition to a place. Each arc is labeled with a value (or a weight), which is a positive integer. An arc which has no label is an arc whose weight is equal by default to 1. An *inhibitor arc* is represented by a line terminating with a rounded rather than an arrow-pointed head. The presence of a token in the inhibitor place inhibits the firing of the transition.

The system state is described by means of markings which specify the number of tokens in each place of the net. A transition is said to be enabled in a given marking, if and only if each of its normal input places contains at least as many tokens as the multiplicity of the connecting arc, and each of its inhibitor input places contains fewer tokens than the multiplicity of the corresponding inhibitor arc. The firing of an enabled transition creates a new marking (state) of the net. Hence, the evolution of the Petri net is represented by a reachability graph, which is isomorphic to a continuous time Markov chain (CTMC). The solution of this CTMC at steady-state is the stationary probability vector π which can be expressed as the solution of the linear system of equilibrium equations $\pi.Q = 0$ with the normalization condition $\sum_i \pi_i = 1$, where π_i denotes the steady-state probability that the process is in state M_i and Q is the infinitesimal generator. Having the probabilities vector π, we can compute several stationary performance indices of the modeled system.

In Fig.1, we present the GSPN model describing the subscriber behavior and the channels allocation policy of the PFR-GCBS strategy. In this model, the place Sub_Free contains the free subscribers, the place CAC_Fr (*resp.* CAC_Ho) represents the arrival and the admission condition of fresh calls (*resp.* handover or a repeated handover calls) for service and the place Cha_Idle represents the idle channels. Initially, it contains C tokens because all channels are available. Place $Call\_In\_Proc$ contains the calls in process. Place $Orbit\_Fr$ (*resp.* $Orbit\_Ho$) represents the orbit of fresh call (*resp.* orbit of handover call). Place $Choice\_Fr$ (*resp.* $Choice\_Ho$) represents the choice made by a blocked fresh call or a blocked repeated fresh call (*resp.* a blocked handover call or a blocked repeated handover call), to enter in the corresponding orbit or to leave the system and the place $Cond\_Borw$ represents the condition for a repeated blocked fresh call (retrial) demand to borrow an unoccupied guard channel.

Initially, the **L** subscribers are free, the two orbits are empty, the **C** channels are idle and no call is in progress.

In the GSPN model, there are two streams of arrivals : the arrivals of fresh calls with rate λ_f, and the arrivals of incoming handover calls with rate λ_h.

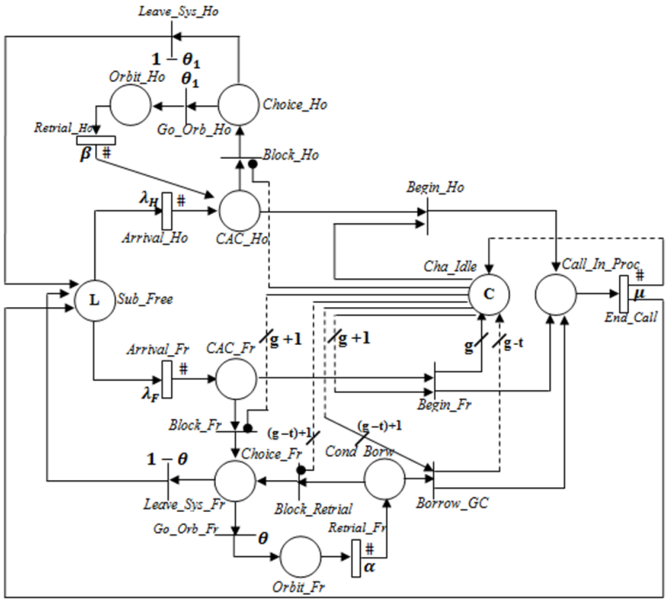


Fig. 1. GSPN model describing the PFR-GCBS strategy.

The firing of transition *Arrival_Fr* (*resp.* *Arrival_Ho*) indicates the arrival of a fresh call (*resp.* a handover call). The service semantics of these transitions is (represented by symbol #) because free subscribers can independently generate fresh calls or handover calls. Hence, the firing rate depends on the marking of place *Sub_Free* and is equal to $\lambda.M$ *(Sub_Free)*, such that $\lambda = \lambda_f + \lambda_h$.

At the arrival of a fresh call (*resp.* a handover or a repeated handover call) to place *CAC_Fr* (*resp.* *CAC_Ho*), if place *Cha_Idle* contains at least *g+1* tokens (*resp.* one token), *i.e.* if there are at least *g+1* idles channels (*resp.* one idle channel), immediate transition *Begin_Fr* (*resp.* *Begin_Ho*) fires. Hence, the subscriber starts being served (the call is in process) and the channel moves into busy state. Otherwise, if place *Cha_Idle* contains at most g tokens (*resp.* is empty), *i.e.* if there is no idle ordinary channel (*resp.* no idle channel), the immediate transition *Block_Fr* (*resp.* *Block_Ho*) fires, the call will be blocked and the subscriber immediately joins the place *Choice_Fr* (*resp.* *Choice_Ho*).

The marking of place *Choice_Fr* (*resp.* *Choice_Ho*) enables two concurrent transitions *Go_Orb_Fr* and *Leave_Sys_Fr* (*resp.* *Go_Orb_Ho* and *Leave_Sys_Ho*).

– The firing of immediate transition *Go_Orb_Fr* (*resp.* *Go_Orb_Ho*) with probability θ_0 (*resp.* θ_1), means that the blocked subscriber joins immediately place *Orbit_Fr* (*resp.* *Orbit_Ho*) and starts generating a flow of repeated calls with rate α (*resp.* β), until it finds an idle channel. In fact, subscribers in orbit of either fresh or handover calls, behave independently of each other. The firing of transition *Retrial_Fr* (*resp.* *Retrial_Ho*) represents the arrival of a repeated call of fresh blocked call (*resp.* a repeated call of handover blocked call) and produces a token in place *Choice_Fr* (*resp. Cond_Borw*). As subscribers independently generate repeated calls, this transition has an ∞-servers semantics.

- The firing of immediate transition *Leave_Sys_Fr* (*resp. Leave_Sys_Ho*) with probability $1\text{-}\theta_0$ (*resp.* $1\text{-}\theta_1$), means that the subscriber became impatient and it decided to leave definitively the system (*resp.* the failure of the handover procedure). After firing of this transition subscriber becomes inactive (free) and joins the set of free subscribers present in the cell coverage area.

At the arrival of a repeated fresh call to place *Cond_Borw*:

- If there is at least one idle channel among *C-g* ordinary channels (place *Cha_idle* contains at most *C-g-1* busy channels), the subscriber starts being served immediately. Otherwise, the retrial call will check for a borrowable guard channel. If it is unoccupied, immediate transition *Borrow_GC* fires and the repeated fresh call will borrow and occupy this guard channel. Hence, the subscriber starts being served and the channel moves into busy state.
- Otherwise, if all ordinary channels are busy (occupied) and there is no unoccupied borrowable guard channel (*i.e.* there are at most *g-t* idle channels), immediate transition *Block_Retrial* fires and the fresh call retrial will be again blocked and the subscriber immediately joins place *Choice_Fr*.

At the end of the call duration, timed transition *End_Call* fires with rate μ. The subscriber under service returns to free state (to place *Sub_Free*) and the channel becomes idle and ready to serve another subscriber. As services take place in parallel, transition *End_Call* has an ∞-servers semantics.

4 Performance measures

As the proposed model is bounded and the initial marking is a home state, the underlying process is ergodic. Hence, the steady-state probability distribution vector π exists and is unique. Once this probability vector is computed, several performance measures of SCNs with the PFR-GCBS strategy can be derived as follows. In these formulas, $M_i\ (p)$ denotes the number of tokens in place p in marking M_i, A the set of reachable tangible markings, and $A\ (t)$ is the set of tangible markings reachable by transition t and $E\ (t)$ is the set of markings where the transition t is enabled.

- **Mean number of busy channels ($\boldsymbol{n_s}$)**:

$$n_s = \sum_{i:M_i \in A} M_i(\mathit{Call\_In\_Proc}).\pi_i$$

– **Mean number of subscribers in orbit of fresh calls (n_{of}):**

$$n_{of} = \sum_{i:M_i \in A} M_i(Orbit\_Fr).\pi_i$$

– **Mean number of subscribers in orbit of handover calls ($\boldsymbol{n_{oh}}$):**

$$n_{oh} = \sum_{i:M_i \in A} M_i(Orbit\_Ho).\pi_i$$

– **Mean generation rate of fresh calls ($\overline{\lambda}_f$):**

$$\overline{\lambda}_f = \sum_{i:M_i \in A(Arrival\_Fr)} \lambda_f.M_i(Sub\_Free).\pi_i$$

– **Mean generation rate of handover calls ($\overline{\lambda}_h$):**

$$\overline{\lambda}_h = \sum_{i:M_i \in A(Arrival\_Ho)} \lambda_H.M_i(Sub\_Free).\pi_i$$

– **Mean generation rate of repeated fresh calls ($\overline{\alpha}$):**

$$\overline{\alpha} = \sum_{i:M_i \in A(Retrial\_Fr)} \alpha.M_i(Orbit\_Fr).\pi_i$$

– **Mean generation rate of repeated handover calls ($\overline{\beta}$):**

$$\overline{\beta} = \sum_{i:M_i \in A(Retrial\_Ho)} \beta.M_i(Orbit\_Ho).\pi_i$$

– **Blocking probability of primary fresh calls ($\boldsymbol{P_{Bf}}$):**

$$P_{Bf} = \Big(\sum_{j:M_j \in A} \sum_{i=1}^{L-(C-g)} i.\lambda_f.Prob[M_j(Sub\_Free) = i \ \&\ M_j(Cha\_Idle) \leq g] \Big) \ / \ \overline{\lambda}_f$$

– **Blocking probability of repeated fresh calls ($\boldsymbol{P_{Brf}}$):**

$$P_{Brh} = \Big(\sum_{j:M_j \in A} \sum_{i=1}^{L-(C-(g-t))} i.\alpha.Prob[M_j(Orbit\_Fr) = i \ \&\ M_j(Cha\_Idle) \leq (g-t)] \Big) \ / \ \overline{\alpha}$$

– **Global blocking probability of fresh calls ($\boldsymbol{P_F}$):**

$$P_F = \frac{\overline{\lambda}_f}{\overline{\lambda}_f + \overline{\alpha}}.P_{Bf} + \frac{\overline{\alpha}}{\overline{\lambda}_f + \overline{\alpha}}.P_{Brf}$$

– **Blocking probability of handover calls (P_{Bh}):**

$$P_{Bh} = (\sum_{j:M_j \in A} \sum_{i=1}^{L-C} i.\lambda_h.Prob[M_j(Sub\_Free) = i \ \& \ M_j(Cha\_Idle) = g] \ ) \ / \ \overline{\lambda}_h$$

–**Blocking probability of repeated handover calls (P_{Brh}) :**

$$P_{Brh} = (\sum_{j:M_j \in A} \sum_{i=1}^{L-C} i.\beta.Prob[M_j(Orbit\_Ho) = i \ \& \ M_j(Cha\_Idle) = 0] \ ) \ / \ \overline{\beta}$$

– **Global blocking probability of handover calls (P_H):**

$$P_H = \frac{\overline{\lambda}_h}{\overline{\lambda}_h + \overline{\beta}}.P_{Bh} + \frac{\overline{\beta}}{\overline{\lambda}_h + \overline{\beta}}.P_{Brh}$$

– **Mean response time ($\overline{R}$)**: This corresponds to the time spent by a call in system.

$$\overline{R} = \frac{n}{\overline{\lambda_f} + \overline{\lambda_h}}$$

5 Numerical results

In this section, we illustrate In Fig. 2, the effect of the borrowable guard channels number variation (t) on the blocking probability of fresh calls (BPF) and that of handover calls (BPH), that for different experiences varying the guard channels number (g) between 1 and 5. TOo this end, we consider a SCN with the following basic configuration parameters: L=40, C=15, g=4, t=2, λ_f^{-1}=10s, λ_h^{-1}=40s, μ =60s, α^{-1} =5s, β^{-1} =2s, θ_0=0.6, θ_1=0.8. As we can clearly see, the calls blocking probability is very sensitive to changes in the borrowable guard channels number (t). Also, the fresh calls blocking probability decreases significantly with increasing of borrowable guard channels number (t), however that of handover calls weakly increases. This can be justified by the effect of the guard channels borrowing concept.

References

1. Ahmed, M.H.: Call admission control in wireless networks: a comprehensive survey. IEEE Communications Surveys and Tutorials 7, 49-68 (2005)
2. Kolate, V.S., Patil, G.I., Bhide, A.S.: Call admission control schemes and handoff prioritization in 3G wireless mobile networks. Inter. Journal of Engineering and Innovative Technology (IJEIT) 1 (2012)
3. Yang, S.F., Wu, J.S., Hwang, B.J.: Performance evaluation of priority based adaptive multiguard channel call admission control for multiclass services in mobile networks. Inter. Journal of Communication Systems 26, 597-609 (2013)

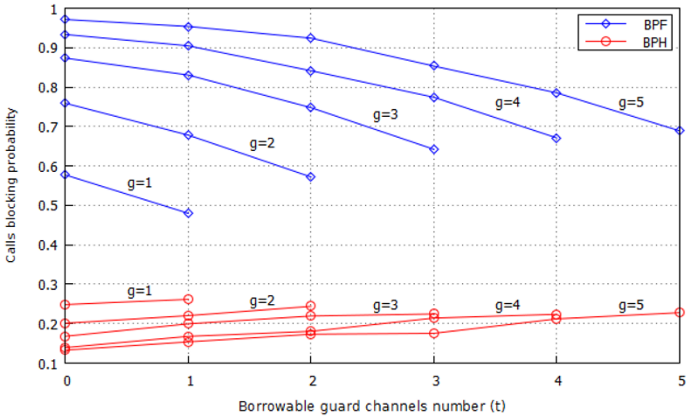


Fig. 2. Calls blocking probability versus the borrowable guard channels number (t).

4. Chowdhury, T.A., Bhattacharjee, R., Chowdhury, M.Z.: Handover priority based on adaptive channel reservation in wireless networks. Inter. Conference on Electrical Information and Communication Technology (EICT), 1-5 (2014)
5. Akpan, U.I., Kalu, C., Inyang, A.K.: Performance Analysis of Prioritized Handoff Schemes in Wireless Systems, Performance Analysis of Prioritized Handoff Schemes in Wireless Systems. Communications. Vol. 2, No. 1, 1-6 (2014)
6. Kim, J., Kim, B.: A survey of retrial queueing systems. Annals of Operations Research, 1-34 (2015)
7. Tran-Gia, P., Mandjes, M.: Modeling of customer retrial phenomenon in cellular mobile networks. IEEE Journal on selected areas in communi.15, 1406-1414 (1997)
8. Artalejo, J.R., Lopez-Herrero, M.J.: Cellular mobile networks with repeated calls operating in random environment. Comput. and Oper. Res. 37, 1158-1166 (2010)
9. Do, T.V.: Solution for a retrial queueing problem in cellular networks with the fractional guard channel policy. Mathematical and Computer Modelling 53, 2059-2066 (2011)
10. Zhou, Z., Zhu, Y.: Optimization of the (MAP1, MAP2)/ (PH1, PH2)/N retrial queue model of wireless cellular networks with channel allocation. Systems Engineering and Electronics 22, 739-748 (2011)
11. Alagu, S., Meyyappan, T.: An efficient call admission control scheme for handling handoffs in wireless mobile networks. Inter. Journal on AdHoc Networking Systems (IJANS) 2 (2012)
12. Jaimes-Romero, F.J., Munoz-Rodriguez, D., Molina, C., Tawfik, H.: Modeling resource management in cellular systems using Petri nets. IEEE Transactions on vehicular technology 26, 298-312 (1997)
13. Ajmone Marsan, M., Balbo, G., Conte, G., Donatelli, S., Franceschinis, G.: Modelling with Generalized Stochastic Petri Nets. John Wiley & Sons, New York (1995)
14. Gharbi, N., Charabi, L.: Wireless Networks with Retrials and Heterogeneous Servers: Comparing Random Server and Fastest Free Server Disciplines. Inter. Journal on Advances in Networks and Services 5, 102-115 (2012)
15. Gharbi, N., Nemmouchi, B., Mokdad, L., Ben-Othmane, J.: The Impact of Breakdowns Disciplines and Repeated Attempts on Performances of Small Cell Networks. Journal of Computational Sciences 5, 633-644 (2014)

Part II
Security in Mobile and Wireless

A Proposed Model for User Acceptance of Mobile Security Measures – Business Context

Basel Hasan[1], Eduard Rajski[1], Jorge Marx Gómez[1], Joachim Kurzhöfer[2]

[1]Department of Computing Science, Oldenburg University, Oldenburg, Germany

{basel.hasan,eduard.rajski,jorge.marx.gomez}@uni-oldenburg.de

[2]Lufthansa Industry Solutions, Norderstedt, Germany

joachim.kurzhoefer@lhind.dlh.de

Abstract. Companies demand Mobile Enterprise Applications (MEAs) due to many advantages concerning their employees' flexibility and productivity. However, adopting MEAs can be risky without applying suitable security measures and policies on mobile devices, namely smartphones and tablets. On the other hand, applying these on mobile devices can restrict their usage. In this paper, we argue that potential restrictions affect the user acceptance of using mobile devices for working purposes. Hence, mobile security measures were researched along with their possible consequences on mobile users. Based on that consequences, a questionnaire was designed and conducted to collect data needed to measure user acceptance. Furthermore, an extension of Technology Acceptance Model (TAM) was proposed taking the "perceived restrictions" as an important construct that affects the user acceptance of applying security measures on their mobile devices. The extended model will help companies by giving an indicator of the user acceptance keeping a balance between security and usability when adopting MEAs.

Keywords: mobile security measures, user acceptance, mobile devices, mobile business.

1 Introduction

The new properties and advances made in mobile technologies have influenced the popularity and usability of mobile devices[1]. Certain features such as working capacity, storage and user friendliness have become comparable to the properties of laptops and computers [1]. As a result of these characteristics, user interactions with mobile devices have increased considerably [2]. Hence, these features have had an impact on the number of enterprises that use mobile technologies. Furthermore, mobile devices are steadily proliferating in both personal and business life [3] [4].

[1] Within this paper, mobile devices refer to smartphones and tablets.

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_11

In the age of globalization, the need of mobility and flexibility can be seen as one of the driving forces for a company to integrate mobile devices into their business processes. Advantages in this regard are cost-efficiency, location-independent business processes, read/write access to corporate data and content at any point of need, allowing data capturing in real time, and an overall improved user experience [5] [2]. For instance, sale personnel can access their mobile Customer Relationship Management System (mobile CRM) and update their customer details while they are away from their offices. However, beside the aforementioned advantages, many business risks arise from adopting mobile solutions in business sectors [6] [7] [8] [9]. This is due to new challenges concerning the large number of interfaces (like SD-cards, USB, Bluetooth ...etc.), types of communications (Wi-Fi, UMTS …etc.). Furthermore, mobile devices are small and portable and therefore, they can easily be stolen or lost.

Therefore, companies have to apply proper mobile security measures to mitigate the risks while still retaining the advantages of mobility. Applying security measures for mobile device has consequences on the mobile users' flexibility and productivity. In case of high restriction levels across all implemented measures, regardless of the actual data sensitivity or department-specific requirements, employees may be less acceptant of such measures, as they may feel overly restricted in using the mobile devices. Due to that, employees may cease to use mobile devices or they may circumvent the implemented measures for easier use. Moreover, password complexity, device encryption, network restrictions and other techniques that restrict the access to information on the mobile device encourage users to find other less secure alternatives that will eventually compromise enterprise data and access [10]. The research behind this paper tries to answer the following questions. a) Which security measures exist in regards to mobile devices and applications? b) Which potential consequences are caused by applying these measures on mobile devices? c) What affects the user acceptance of using mobile devices for working proposes?

First, a literature review has been conducted to determine the existing mobile security measures, along with their possible consequences on the mobile user. Based on that, a questionnaire has been designed and distributed in companies to collect the data needed to investigate user acceptance. Consequently, user acceptance was investigated side by side with extending TAM for MEAs concerning security. In addition, a statistical analysis has been conducted using SPSS (Statistical Package for the Social Sciences) to assess the quality of the results and possible correlations in regards to the extended model.

2 Overview

Within scope of this paper, mobile applications refer primarily to applications that run on mobile devices and that are optimized for that purpose. In business sectors, mobile applications are to be considered from different perspectives: a) mobile business applications, which run business processes to produce a specific service or product for customers, e.g. CeWe Fotobuch App[2], and b) Mobile Enterprise Applications (MEAs), which are used by employees for working purposes, e.g. Good for Enterprise[3]. However, in this paper, user acceptance was investigated under the second perspective, MEAs.

The recent growing trend is Bring Your Own Device (BYOD) to workplace for official use, which was developed due to consumerization in IT [11]. Thus, employees can have the possibility to work using their private mobile devices. This work considers both corporate devices, which are owned and offered by companies, and private devices, which are owned by employee.

This research uses TAM to address key factors that affect the adoption of MEAs by companies, taking security as its main focus. The original TAM is depicted in Fig. 1. It presents two main constructs, perceived usefulness and perceived ease of use, that influence the attitude towards using the technology, which in turn influences the actual use of the system.

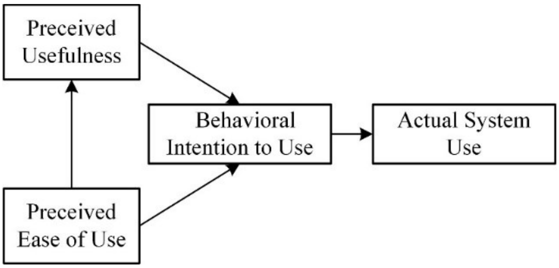


Fig. 1. The original Technology Acceptance Model

This model introduces two main constructs. Perceived usefulness is defined as "the degree to which a person believes that using a particular system would enhance his or her job performance." [12]. In this paper, this factor reflects the degree to which an employee believes that the applied security measures are important for secure usage of mobile devices in business context. Perceived ease of use refers to "the degree to which a person believes that using a particular system would be free of effort." [12], this reflects the degree to which an employee believes that using the applied security measures would be effort free.

3 Related Work

Mobile security measures and restrictions are well studied in security literature [2] [13]. However, those studies did not discuss the potential consequences to mobile users, which might be caused by applying security measures on mobile devices. Some of these consequences were mentioned in [14]. The user acceptance rate of such consequences

[2] http://www.cewe.de/cewe-apps.html

[3] https://www1.good.com/applications/collaboration-suite/good-for-enterprise/

has to be considered when companies want to offer mobile technologies to their employee. When it comes to the user acceptance studies, TAM was often applied. TAM was originally presented by Davis [15] and was extended in TAM2 by incorporating additional constructs of social influence process and cognitive instrumental processes [16]. In this study, the original TAM was considered, since the additional constructs for the TAM2 are mostly irrelevant to the studied domain.

User acceptance studies for security are rare. However, two works were found which corresponded with our topic. Osman [17] provided an extended model of the acceptance of mobile government systems. In that model, trust was included as a construct that directly affects the intention to use of mobile government systems. Security was not included in that model. Arpaci et al. [18] presented a study aimed at investigating the impact of perceived security on organizational adoption of mobile communication technologies, specifically smartphones. That work investigated how the perception of security levels, when using mobile technologies, can affect the organizational adoption of smartphones, since the perception of low security levels can increase the technological risks of adopting these technologies in organizations [18]. However, to achieve a certain security level, proper security measures have to be applied. Applying security measures on mobile devices has consequences for mobile users. The perception of such consequences is the focus of our work. Some further works studied user acceptance, but they focus on different technologies. Benenson et al. [19] studied the users' understanding, usage and acceptance of attribute-based credentials. Kashi and Zheng [20] investigated how the impression of recruitment websites can affect job applicants' intentions to apply for a job. However, TAM extension for using mobile devices in business sectors concerning security and its consequences on the mobile users has not been studied so far.

4 Mobile Security Measures and Restrictions

Compared to Personal Computers (PCs), mobile devices have very different security principles. According to [7], three major factors distinguish mobile security from traditional computer security. First, mobile devices are carried by their owners and have high mobility. Therefore, they are on high risk of being stolen or lost. Second, mobile devices are strongly personalized, and the device owner is normally its unique user. Third, they have strong connectivity accessing various Internet services. This makes them more vulnerable by malware through a variety of channels. Furthermore, mobile devices have limited resources (CPU power, memory, battery life) compared to PCs [7]. Due to these limited resources, the security measures cannot be simply ported from traditional PCs and laptops into mobile devices, e.g. complex intrusion detection algorithms are not suitable for mobile devices because they require excessive CPU power [7]. By integrating mobile devices into business sectors, important and critical business data can easily find their way to the mobile devices. In order to protect such data against threats, companies apply security measures and policies on mobile devises. In the following, existing mobile security measures along with their consequences for mobile users is presented based on the literature review and best practices within companies.

Some of security measures seem similar to those applied on PCs, but their implementation and usage is accompanied by potential consequences for mobile users. The questionnaire has been designed based on the following measures and their consequences.

Authentication generally refers to methods that are used for user identification and verification. Possible implementations for authentication on mobile devices are:

- *Knowledge Based Methods* (like passwords, PIN – Personal Identification Number, pattern locks and graphical passwords) are based on exclusive user knowledge [21].
- *Biometric Authentication* uses biological features (usually voice, face or fingerprint) for the identification and verification of a user [22] [23].
- *Continuous Authentication* is basically a behavioral monitoring that continuously compares the current user activities or specific actions with the usual behavior of the authorized user, rejecting users that deviate to a higher degree than is allowed [24]. For instance, Finger-gestures authentication using Touchscreen [25].

Potential consequence of authentication can be a high chance of error, when entering a complex alphanumeric password. But usability can be supported by using multi-level authentication, which allows the implementation of passwords with varying complexities depending on the required security level. Biometric authentication is problematic because the recognition may produce false negatives and stops the user from rightfully accessing content. The reverse can be true as well and an unauthorized user may gain access due to adequate similarity. Pattern authentication is usually easy to handle for users but it is also easily predictable, which decreases security. Finally, continuous authentication simplifies the authentication process through automatization, but may increase the usage of device resources due to the need for continuous monitoring.

Encryption is the main key to ensure data confidentiality and to ensures that only the right users can read the information.

- *Data Encryption for on-device data*, for instance, FDE (Full Disk Encryption). FDE encrypts the entire hard drive and it is generally classified into software-based and hardware-based solutions [26]. FDE is especially important for mobile devices that can be physically lost or stolen. However, there are software-based encryptions, which do not encrypt the entire hard drive, but encrypt app-specific file or folders.
- *Data Encryption for transmitted data*, such as VPN (Virtual Private Network), which allows the user to use unsecure access points to transfer the data through a secured tunnel. A VPN uses the encryption to keep the transmitted data secure and it also verifies the identity of any one using the network [27].

By encrypting the data on mobile device while they are transmitted, security is enhanced and employees can work safely everywhere. However, users may experience longer loading and saving times when in case of data encryption on mobile devices due to the limited resources concerning CPU power and memory.

Containerization and Virtualization are especially important when using BYOD devices.

- Containerization technology provides each managed mobile application, and its data, with its own secure runtime container [28]. Containerization causes a mobile application to transform in multiple ways: the mobile application data is encrypted and segregated from all other apps; native OS runtime system calls are replaced with equivalent secure versions; and secure back end connectivity and app-to-app secure workflows are enabled [28].
- Virtualization technology can be as an alternative to Containerization. A number of mobile virtualization technologies have been presented in [10].

Both technologies, Containerization and Virtualization, allow a clear separation between private and business data. However, they can cause higher consumption of device resources, that may slow down the device, as well as higher battery drain. In case of server/client solutions the required high bandwidth as well as low latency may also be of consequence. This can cause extra costs depending on mobile tariff. Moreover, Virtualization requires compromises that fracture the mobile user experience, for example, they require an always-on connection to be effective [28].

Protection Software like Antivirus Software are used to detect and protect the mobile device from malware and to protect users whilst surfing the internet, primarily against fishing attacks, by defining blacklists or whitelists of specific websites [29].

Due to the need of continues monitoring, the device has a higher consumption of device resources, like RAM and battery. Another consequence of this is that the security of the system is directly dependent on the up-to-dateness of the virus signature as well as the firewall rules.

Other Security Measures

- Security techniques applied through MDM, like Monitoring the device for any unauthorized changes or mobile remote wiping to remotely wipe data from lost or stolen mobile device [30]. Monitoring of mobile devices can violate user privacy in the case of everything are monitored (like internet activities, email ... etc.). The monitoring can be applied only to detect policy violations. Remotely wiping lost or stolen mobile device increases the security of both corporate as well as private data.
- Restrictions applied on mobile devices. The following restrictions list was defined by [2]:
  - Restrict user and mobile application access to hardware, like digital camera, GPS, and removable storage, etc.
  - Restrict user and mobile application access to native OS services, such as built-in web browsers, email client, contacts, etc.
  - Restrict which mobile application stores may be used.
  - Restrict which mobile applications may be installed through whitelisting or blacklisting.
  - Restrict updating mobile applications.
  - Restrict the permissions (e.g., camera access, location access) assigned to each mobile application.
  - Restrict the use of mobile operating system and synchronization services.
  - Restrict the connection of mobile devices.

Those restrictions form clear guidelines for mobile users, but on the other hand, some mobile functionalities and features might be lost due to restricted application rights. The following section presents methodology of research conducted and the results concerning user acceptance.

5 Research Methodology

The main research objective is to investigate the factors that drive user acceptance of using mobile devices in business sectors. For this study, a questionnaire was used to gather the required information. The questions in this questionnaire are primarily derived from the security measures and their possible consequences on user (see Section 4). Applying security measures on mobile devices has consequences, which can restrict the use of mobile devices. We mainly hypothesize that perceived restrictions would have a significant impact on perceived usefulness and perceived ease of use of the security measures.

5.1 The Design of the Questionnaire

The basic design of this questionnaire consists of three primary sections. In first section, the employees were asked questions concerning their ages, business and private usage of mobile devices and whether they use private or company-owned mobile device. A second Section included specific questions on security measures, the possible restrictions, perceived ease of use and perceived usefulness of security measures. In the final section, the participant is confronted with questions regarding personalized policies and intention to use the security measures introduced. Apart from the general questions in the first section, which use question-specific options, respondents were asked to rate their acceptance using 7-Point Likert Scale ranging from 0 = no acceptance to 6 = very high acceptance. The questionnaire has some limitations. First, the majority of questions ask for the subjective perception of the participant regarding the consequences of certain security measures. Although subjectivity is usually discouraged, the topic at hand requires the participant's perception to assess future user acceptance and it is therefore allowed. A second limitation can be found in the different levels of understanding by the participants concerning the security measures introduced, the accompanying restrictions and their consequences. This may skew the participant's perception on the subject matter.

5.2 Sample and Data Collection

As the main topic of this research is the security of mobile devices and mobile applications in a business context, the target group has to reflect these aspects as well. Thus, the target group consists of employees from three German companies that make it possible for their employees to use mobile devices for working purposes. Hence, those employees would be the most affected by implemented security measures on mobile devices. Furthermore, it is important to consider, whether the device is predominantly used for business or for private purposes, as it may influence the impact of the implemented measures on the overall user acceptance as well. The questionnaire was designed online and distributed in three German companies. 88 participants have taken part in this questionnaire with ages ranging from 24 to 63 years.

5.3 Descriptive Results

After collecting the data, the internal consistency measurement was assessed using of the Cronbach's alpha coefficient [31]. Cronbach's alpha was calculated for each of the multi-item factors of the proposed TAM model, which is depicted in Fig. 4. Table 1 presents the Cronbach's alpha values for the investigated variables. According to [32], reliabilities less than 0.6 are considered to be poor, those in the 0.7 range, acceptable, and those over 0.8 good.

Table 1. Internal Consistency for the investigated variables

Variables	Cronbach α	Acceptable if in 0.7 range [32]
Perceived Ease of Use - PEOU	0.701	Yes
Perceived Usefulness - PU	0.782	Yes
Intention to Use - ITU	0.749	Yes
Perceived Restriction - PR	0.713	Yes

First, we present some general information we obtained from the collected data about the use of mobile devices for working purposes. As depicted in Fig. 2, about 86% of the participants use mobile devices (private device 28% or corporate devices 58%) in a business setting, with the majority of 55% using smartphones and 27% using both smartphones and tablets together. 71% of participants, who use their corporate devices for business, use these corporate devices privately as well.

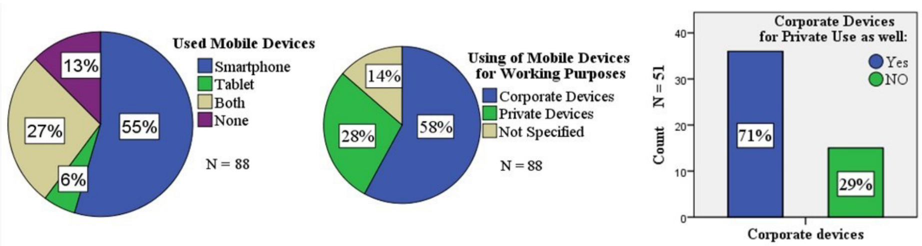


Fig. 2. Mobile devices' usage for work

In regard to the user acceptance of security measures, we found that the user acceptance depends significantly on the restriction level of the security measures Fig. 3 exemplifies this through showcasing several measures with varying restriction levels. On the same figure, the difference between corporate and private devices is also presented. For instance, applying whitelisting is less acceptable than allowing standard AppStores, like Google Play and Apple AppStore. However, it seems users, who use corporate devices, are more flexible about accepting high restrictions than users, who use their private mobile devices.

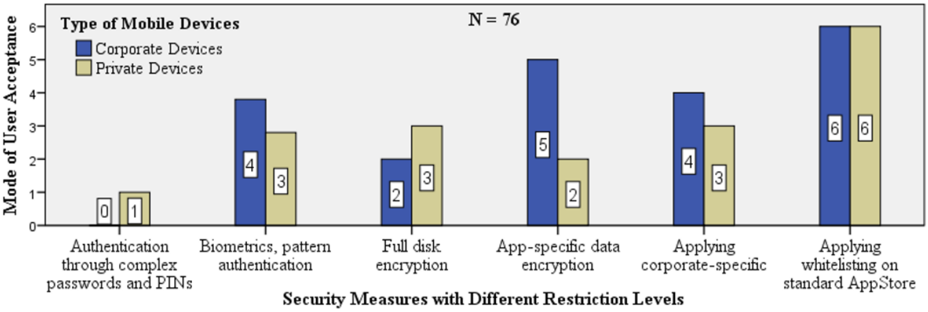


Fig. 3. Showcase for user acceptance of security measures with different restriction levels

5.4 The Proposed Extension of TAM

The original Technology Acceptance Model (TAM), which is presented in Fig. 1, does not cover specific requirements concerning mobile security, such as the perceived restrictions when applying security measures on mobile devices when they are used in a business context. This in turn leads to distortion of the actual user acceptance of the studied technology. Additionally, the correlation between the standard factors may vary as well, due to different priorities in different application areas. The measured results in this work play a major role in the construction of the model proposed in this paper. Fig. 4 shows the proposed extension of TAM along with the correlations between its constructs. The construct "perceived restriction" was added to the original model. A statistical analysis using SPSS was conducted on the collected data in order to measure the Pearson Correlation between the constructs of the proposed model.

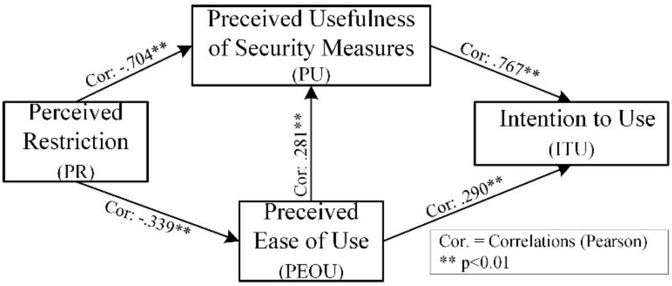


Fig. 4. The proposed user acceptance model

Looking at these correlations displayed in Fig. 4, it is evident that the perceived restriction has a significant impact on the perceived usefulness of security measures. As this correlation is negative, the higher restriction level the lower usefulness of mobile security measures. The perceived restrictions show also a negative correlation to ease of use. The ease of use shows a positive correlation to both, usefulness and intention to use the security measures. These two correlations are relatively low. Furthermore, perceived usefulness is significantly correlated with the intention to use of mobile security measures. The complete correlations between the constructs of the proposed model are displayed in Table 2.

Table 2. Correlation between the constructs

	PEOU	PU	ITU	PR
PEOU	1			
PU	.281**	1		
ITU	.290**	.767**	1	
PR	-.339**	-.704**	-.769**	1

**Correlation is significant at the 0.01 level (2-tailed).

5.5 Discussion

Achieving a balance between security and usability is very important to encourage productive use of mobile devices. Applying high restrictive security measures lowers the overall user acceptance and may trigger the employee to circumvent those measures. Derived from the proposed model, if security measures with a high restriction level are applied, the usefulness of the applied security measures will decrease, which in turn minimize the intention towards using that measures, or in other words, minimize the intention to use mobile devices for working purposes. Consequently, the companies will lose the advantages in employee flexibility and productivity when adopting mobile enterprise applications. Data analysis has also revealed that the users' acceptance of the perceived restrictions is relatively dependent on whether they use corporate or private devices, where the users who use corporate devices seem more accepting a higher level of restrictions than those who want to use their private mobile devices for working purposes. However, it is difficult to generalize the results presented due to some limitations. Beside the limitations mentioned in section 5.1, further limitations can be seen concerning cultural aspects, since different countries and cultures may emphasize different aspects. Moreover, the work relevancy and the application area can limit the results of this work. The employees who deal with sensitive data can show a higher acceptance rate compared to those who deal with less sensitive data. With other user groups, especially high-tech or health sectors, the results on perceived usefulness and perceived restrictions could be also different.

6 Conclusion

In this work, a literature review has been conducted in mobile security domain taking the security of mobile devices, namely smartphones and tablets, as the main focus. Based on that, mobile security measures have been determined along with their potential consequences on the mobile users. Based on the determined mobile security measures and their consequences, a questionnaire has been designed and the questions have been validated through discussions with experts within companies. After that, the questionnaire has been distributed inside three companies in Germany, two of them IT companies. Based on the collected data, some descriptive results have been presented concerning the usage of mobile devices for working purposes and the user acceptance of mobile security measures with different restriction levels. Finally, we proposed an extension of the technology acceptance model by including the perceived restrictions. The data analysis using SPSS showed that the perceived restrictions are negatively correlated with perceived ease of use and perceived usefulness as well as the intention to use of security measures on mobile devices. The study done in this paper help the companies by providing an argument that the user acceptance has to be given a special attention when adopting mobile enterprise applications. But due to the limitations discussed, it is difficult to generalize the results.

However, during the research, several interesting aspects presented themselves for further investigation, including the sensitivity of data and mobile security awareness as additional constructs that affect the user acceptance.

References

1. Giessmann, A., Stanoevska Slabeva, K., Visser, B. de: Mobile Enterprise Applications--Current State and Future Directions. 45th Hawaii International Conference on System Science (HICSS), (2012) 1363–1372
2. Souppaya, M. and Scarfone, K.: Guidelines for Managing the Security of Mobile Devices in the Enterprise, DOI: 10.6028/NIST.SP.800-124r1, (2013)
3. Disterer, G., Kleiner, C.: Mobile Endgeraete Im Unternehmen. Technische Ansaetze, Compliance-anforderungen, Management. Vieweg + Teubner Verlag (2014)
4. Mylonas, A., Kastania, A., Gritzalis, D.: Delegate the smartphone user? Security awareness in smartphone platforms. Computers & Security 34, (2013) 47–66
5. Basole, R.: Strategic Planning for Enterprise Mobility: A Readiness-Centric Approach. AMCIS 2007 Proceedings. (2007)
6. Beckett, P.: The business risks of using smartphones. Network Security 2015, (2015) 16–17
7. Daojing He, Chan, S., Guizani, M.: Mobile application security: malware threats and defenses. Wireless Communications, IEEE 22 (1), (2015) 138–144
8. Copeland, W., Chiang, C.C.: Securing Enterprise Mobile Information. Consumer and Control (IS3C) Computer, (2012) 80–83
9. Qing, L., Clark, G.: Mobile Security: A Look Ahead. Security & Privacy, IEEE 11 (1), (2013) 78–81
10. Jaramillo, D., Furht, B., Agarwal, A.: Mobile Virtualization Technologies. In: Virtualization Techniques for Mobile Systems. Springer International Publishing, Cham (2014) 5–20
11. Vignesh, U., Asha, S.: Modifying Security Policies Towards BYOD. Procedia Computer Science 50, (2015) 511–516
12. Davis, F.D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 13 (3), DOI: 10.2307/249008, (1989)
13. Park, J.H., Yi, K.J., Jeong, Y.-S.: An enhanced smartphone security model based on information security management system (ISMS). Electron Commer Res 14, (2014) 321–348
14. Hasan, B., Mahmoud, T., Marx Gómez, J., Pramod, R., Kurzhöfer, J.: User Acceptance Identification of Restrictions Caused by Mobile Security Countermeasures. In: MOBILITY 2015, The Fifth International Conference on Mobile Services, Resources, and Users, (2015) 31–37
15. Davis, F.D.: A technology acceptance model for empirically testing new end-user information systems : theory and results. Diss, http://hdl.handle.net/1721.1/15192, (1986)
16. Venkatesh, V., Davis, F.D.: A Theoretical Extension of the Technology Acceptance Model. Four Longitudinal Field Studies. Management Science 46, (2000) 186–204
17. Osman, N.B.: Extending the Technology Acceptance Model for Mobile Government Systems. development 5, (2013)

18. Arpaci, I., Yardimci Cetin, Y., Turetken, O.: Impact of Perceived Security on Organizational Adoption of Smartphones. Cyberpsychology, behavior and social networking 18, (2015) 602–608
19. Benenson, Z., Girard, A., Krontiris, I., Liagkou, V., Rannenberg, K., Stamatiou, Y.: User Acceptance of Privacy-ABCs: An Exploratory Study. In: Human Aspects of Information Security, Privacy, and Trust, 8533. Springer International Publishing (2014) 375–386
20. Kashi, K., Zheng, C.: Extending Technology Acceptance Model to the E-recruitment Context in Iran. Int J Select Assess 21, (2013) 121–129
21. Rogowski, M., Saeed, K., Rybnik, M., Tabedzki, M., Adamski, M.: User Authentication for Mobile Devices. In: Computer Information Systems and Industrial Management, 8104. Springer Berlin Heidelberg, Berlin, Heidelberg (2013) 47–58
22. Knopper, K.: Mobile Security. In: App4U. Springer Fachmedien Wiesbaden, Wiesbaden (2014) 217–249
23. Grudzien, W.: Biometrie im Banking — Ein Plädoyer gegen Vorurteile. Datenschutz Datensich (2015) 39, 7–11
24. Kambourakis, G., Damopoulos, D., Papamartzivanos, D., Pavlidakis, E.: Introducing touchstroke. Keystroke-based authentication system for smartphones. Security Comm. Networks, DOI: 10.1002/sec.1061, (2014)
25. Feng, T., Liu, Z., Kwon, K.-A., Shi, W., Carbunar, B., Jiang, Y., Nguyen, N.: Continuous mobile authentication using touchscreen gestures. In: 2012 IEEE International Conference on Technologies for Homeland Security (HST), (2012) 451–456
26. Muller, T., Freiling, F.C.: A Systematic Assessment of the Security of Full Disk Encryption. IEEE Trans. Dependable and Secure Comput. 12, (2015) 491–503
27. Turban, E., King, D., Lee, J.K., Liang, T.-P., Turban, D.C.: E-Commerce Security and Fraud Issues and Protections. In: Electronic Commerce. Springer International Publishing, Cham (2015) 457–518
28. Good Technology Corporation: Evaluating Approaches to Mobile Security. [White Paper], https://media.good.com/documents/wp-evaluting-approaches-to-mobile-security.pdf (2015)
29. v Do, T., Lyche, F.B., Lytskjold, J.H., van Thuan, D.: Threat assessment model for mobile malware. In: Kim, K.J. (ed.) Information Science and Applications, 339. Springer Berlin Heidelberg, Berlin, Heidelberg (2015) 467–474
30. Kizza, J.M.: Mobile Systems and Corresponding Intractable Security Issues. In: Kizza, J.M. (ed.) Guide to Computer Network Security. Springer London, London (2015) 491–507
31. Cronbach, L.J.: Coefficient alpha and the internal structure of tests. Psychometrika 16, (1951) 297–334
32. Sekaran, U.: Research methods for business. A skill-building approach. New York, Great Britain: Wiley (2003)

Speaker Verification Using A Modified Adaptive GMM Approach Based On Low Rank Matrix Recovery

Xinjie Ma[1], Tan Dat Trinh[2], Jin Young Kim[3*], and Hyoung Gook Kim[4]

[1,2,3]Chonnam National University, Gwangju, Rep. of Korea
[4]Kwangwoon University, Seoul, Rep. of Korea
[1]84515045@qq.com
[2]trinhtandat1230@yahoo.com
[3]beyondi@jnu.ac.kr
[4]hkim@kw.ac.kr
*corresponding author

Abstract. In this paper, we propose a new method to calculate observation confidence values that are applied in a modified adaptive GMM training for speaker verification. First, we use low rank matrix recovery (LRR) to find enhanced speeches and estimate frame SNR values. Then, a simple sigmoid function is applied to convert the frame SNR values into the observation confidence values. We also combine the frame SNR values estimated by the MMSE log-STSA and LRR methods in order to enhance performance of speaker verification system. To verify the accuracy of the system, we use utterances from a Korean movie "You came from the stars". The experimental results show that our proposed approach achieves better accuracy than the baseline GMM-UBM under both clean and noisy environments.

1 Introduction

Speaker verification is an interesting task and widely used in many applications such as biometric systems, access control systems to facilities, secured transactions, over a network, and so on [1]. Text-independent speaker verification under uncontrolled noise conditions, inter-speaker and channel/ session variabilities is one of the most important challenges. Text-independent speaker verification has been actively studied since early. Many various approaches have been introduced to overcome this problem. An adaptive Gaussian mixture model (GMM) based Universal background model (UBM) has been applied to model target speakers [2]. In the GMM-UBM framework, GMM-supervector approach has been explored by stacking means of adapted mixture components [3]. The GMM-supervector based Support vector machine (SVM) has achieved the most promising performance [3].

To deal with the inter-speaker and channel/session variabilities problems, Joint factor analysis (JFA) [4] and identity vector (called as i-vector) [5] approaches have been investigated to derive robust models and features for text-independent speaker verification system. These approaches have become state-of-the-art performance in speaker verification applications. Besides, sparse representation (SR) has recently

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_12

been used to improve the performance of the speaker verification system. Li et al. [6] has introduced the sparse representation on the total variability i-vector based method for the robust speaker verification.

Recently, a promising method based observation confidence values under an adaptive GMM framework has been successfully applied to the robust speaker verification [7, 8]. Authors considered observation vectors affected or corrupted by noise and channel variabilities as having low confidence values and decreasing the performance of the speaker verification system. A modified adaptive GMM training based the observation confidence values has been proposed to overcome these problems. They calculated the observation confidence values based on frame SNR values. The frame SNR values were computed by using an input speech and an enhanced speech estimated by the optimal minimum mean square error logarithm short-time spectral amplitude (MMSE log-STSA) [9].

In this paper, we apply the modified adaptive GMM based the observation confidence framework [7] for the speaker verification. We propose a new approach to calculate the observation confidence values based low rank matrix recovery (LRR). First, the LRR is applied to estimate the enhanced speech from the input speech. Then, the frame SNR values are calculated by using the input speech and the enhanced speech. The observation confidence values are derived from the frame SNR values and a simple sigmoid function. In addition, we also propose to fuse the frame SNR values computed by the MMSE log-STSA and LRR approaches in order to calculate the observation confidence values for improving the speaker verification performance.

2 Proposed Method for Robust Speaker Verification

2.1 Modified Adaptive GMM based Low Rank Matrix Recovery

The overall structure of the speaker verification system using the modified adaptive GMM training based on the observation confidence mentioned in [7,8] is shown in Fig. 1.

a) Feature extraction: First, RASTA_PLP [11] is applied to extract feature vectors from audio signals.
b) Observation confidence calculation: The confidence values are measured based on the frame SNR values and the sigmoid function. The frame SNR values were computed by using the input speech and the enhanced speech estimated by the optimal MMSE log-STSA.
c) UBM model: All samples in the training database are used to train the UBM model through the modified EM algorithm.
d) Target speaker model: The GMM models for target speakers are adapted from the UBM model via the modified MAP estimation.

In this system, the observation confidence values of the observation vectors have been introduced to counter noise problems such as music effect, street sound, cell phone ringing sound, sob sound, and so on that have low confidence values and can decrease the performance of the speaker verification system. In [7], authors have used the MMSE log-STSA approach to estimate the enhanced speech from the input speech and calculate the frame SNR values. Then, the frame SNR values are transformed into the observation confidence values ranging from 0 and 1 using the simple sigmoid function.

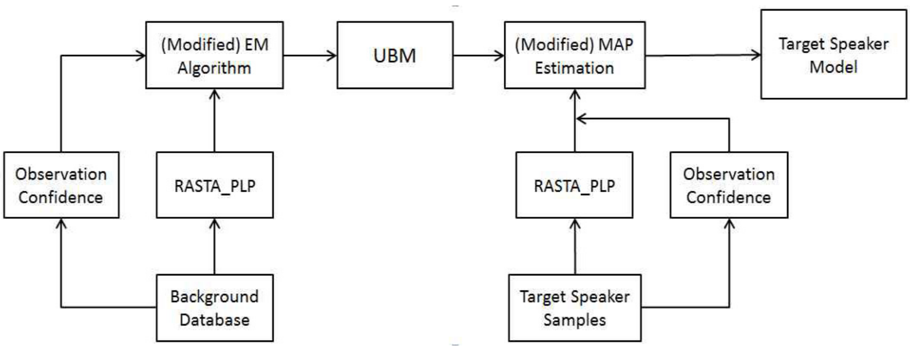
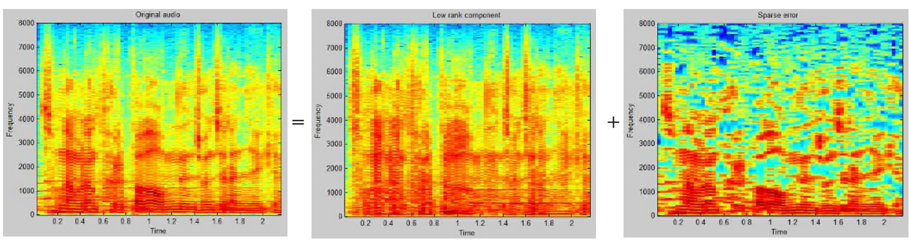


Fig. 1. Block diagram of the modified adaptive GMM training algorithm based the observation confidence value.

In this paper, we propose to use the low rank matrix recovery (LRR) in order to calculate the observation confidence values. In our database, input speeches are affected and corrupted by different types of noise such as background music, street noise, camera/phone sound, crowd sound, etc. We consider noise components that lie in a low subspace as low rank matrix because of its repetition characteristics. The speech has more variation and is relatively sparse error matrix. We apply the LRR based time-frequency masking method as mentioned in [10] to separate the speech and noise components. Fig. 2 shows an example of spectrograms of the original speech corrupted by music, the low rank component and the sparse error component estimated via LRR based time-frequency masking method.



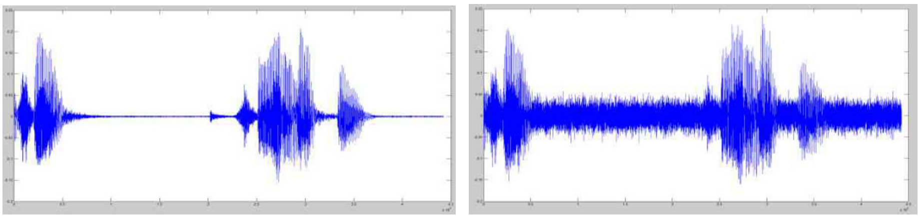
a) b) c)

Fig. 2. The spectrogram of a) the original audio (speech + music) b) the low rank component (music) c) the sparse error (speech)

The LRR based time-frequency masking method can be considered as an estimator to estimate the enhanced speech from the noisy input signal. Then, the frame SNR values are computed by using the input speech and the enhanced speech. However, ranges of amplitude of the input speech and the enhanced speech are different. This can make incorrectly SNR estimation. We also investigate effects of amplitude normalization of the enhanced speech to calculate SNR values. In Fig. 3, we add additive white Gaussian noise to a clean speech to make a noisy speech in desired SNR in 5dB. Then, we use the LRR based time-frequency masking to find the enhanced speech and the frame SNR values. Fig. 4a shows the enhanced speech and the frame SNR values based the LRR without the amplitude normalization. In this figure, the global SNR is equal to -0.2 dB, which means incorrect SNR value. To overcome this problem, we propose to normalize the range of amplitude of the enhanced speech into the range of amplitude of the input speech by using Min-Max mapping

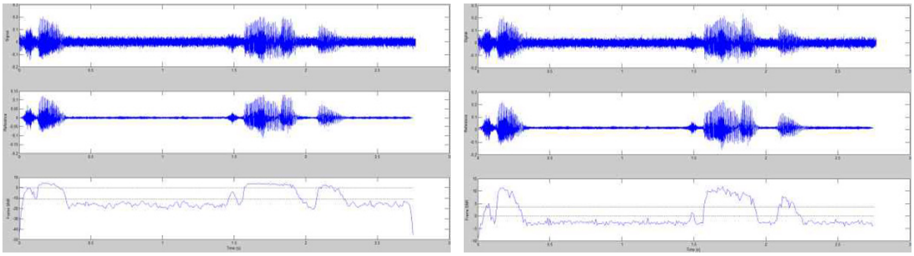
$$X\_norm = \frac{(X - X\_\min) * (\text{Input\_max} - \text{Input\_min})}{X\_\max - X\_\min} + Input\_\min \cdot \quad (1)$$

where X is the enhanced speech, X_max and X_min are maximum and minimum amplitude values of the enhanced speech, respectively. $Input\_max$ and $Input\_min$ are maximum and minimum amplitude values of the input speech, respectively. Fig. 4b shows the enhanced speech and the frame SNR values based the LRR with the amplitude normalization. The global SNR is 5.6 dB; it is very close to the true SNR value. After computing the frame SNR values, the simple sigmoid function is used to transform them into the observation confidence values [7].



a) b)

Fig. 3. a) Clean speech. b) Noisy speech (corrupted by additive white Gaussian noise).



a) b)

Fig. 4. The estimation of the enhance speech and the frame SNR values based the LRR a) without the amplitude normalization, b) with the amplitude normalization.

2.2 Observation Confidence Computation based Fusion of Frame SNR Values

In our study, we have tried to combine the frame SNR values calculated by using the MMSE log-STSA and the LRR approaches in order to improve the performance of the speaker verification system. The two frame SNR values are fused using a simple linear combination rule.

$$FSNR_{SL} = a_S FSNR_S + a_L FSNR_L \,. \tag{2}$$

where a_S+a_L=1, a_S and a_L are weights assigned to each frame SNR values. $FSNR_S$ and $FSNR_L$ are the frame SNR values computed by using the MMSE log-STSA and the LRR approaches, respectively. The weight factors affect the performance of verification system and can be randomly chosen. In this study, we run thirty times with randomly initialized weight factors to evaluate the performance of the fusion approach for the speaker verification system.

3 Experimental Results

We use the database from the first twelve episodes of the Korean drama "You came from the stars" as mentioned in [7, 8] to evaluate the performance of proposed method. We manually segmented and grouped audio samples corresponding to target speakers and other events. The audio signals can take different length. The minimum length of a sample is 0.5s, and the maximum length of 10s. Seven target speakers are chosen to evaluate the accuracy of the verification system. We use the clean speeches of each target speaker from the first episode to the sixth episode to form the training groups. Test group is separated into two parts: Clean Test and Noisy Test. The Clean Test consisted of clean samples of the target speakers from the last six episodes. The Noisy Test is constructed using noisy target speaker samples from the seventh episode to the twelfth episode. In addition to, 1200 random samples from other events are also added to the two test groups to verify the performance of the system. Table 1 shows the detail information of the dataset.

Table 1. Information of training and test dataset.

Speakers	Number of Samples		
	Training	Testing	
		Clean Test	Noisy Test
Speaker 1	**170**	**208**	**611**
Speaker 2	**348**	**513**	**1167**
Speaker 3	**44**	**114**	**179**
Speaker 4	**162**	**137**	**326**
Speaker 5	**53**	**22**	**216**
Speaker 6	**57**	**61**	**33**
Speaker 7	**11**	**87**	**111**

RASTA_PLP [11] is applied to extract features from audio samples. We combine thirteen PLP coefficients with delta and double delta coefficients to obtain 42-dimensional cepstral feature vector. We use 16 ms Hanning window that overlap 50% to calculate the feature vector. The numbers of windows depend on the length of audio samples. Feature warping is then implemented to reduce channel effect [12]. We use 3000 random samples to train GMM-UBM model via the modified adaptive GMM training with a relevance factor of 16. We choose 128 mixture components in the GMM-UBM model. To evaluate the performance, we use equal error rate (EER) rate where false rejection error (miss detection) is equal to false acceptance error (false alarm). During the testing phase, we first identify the test utterances, then the scores are calculated and compared with a threshold to decide to accept or reject.

The experimental results of the speaker verification systems are shown in Table 2. In this table, BGMM-UBM denotes the baseline GMM-UBM system without applied observation confidence value. MGMM-UBM-log-STSA is described as the modified adaptive GMM-UBM based the observation confidence values computed using the MMSE log-STSA method. Our proposed method, the modified adaptive GMM-UBM based observation confidence value computed by using LRR method, is denoted as MGMM-UBM-LRR.

Table 2. Performance comparison of various systems on our dataset.

System	EER(%)	
	Clean Test	Noisy Test
BGMM-UBM	10.97	14.89
MGMM-UBM-log-STSA	**10.29**	**13.39**
MGMM-UBM-LRR	**10.36**	**13.13**

The experimental results show that the proposed method significantly improves the performance of the speaker verification system and is better than the baseline GMM-UBM system in both clean and noisy conditions. The proposed method is also compared to the MMSE log-STSA method. It can achieve a little better accuracy than the MMSE log-STSA in noisy cases.

Fig. 5 shows the performance comparison of the modified adaptive GMM-UBM system based the observation confidence value that are computed by using the fusion of the frame SNR values using the MMSE log-STSA and the LRR approaches (fusion approach) with three approaches shown in Table 2. In this figure, the fusion approach using the linear combination rule with the weight factors, $a_S = 0.2609$ and $a_L = 0.7391$, gives the highest performance compared with other approaches. Therefore, finding good observation confidence value based the frame SNR value is important to increase the accuracy of the speaker verification system.

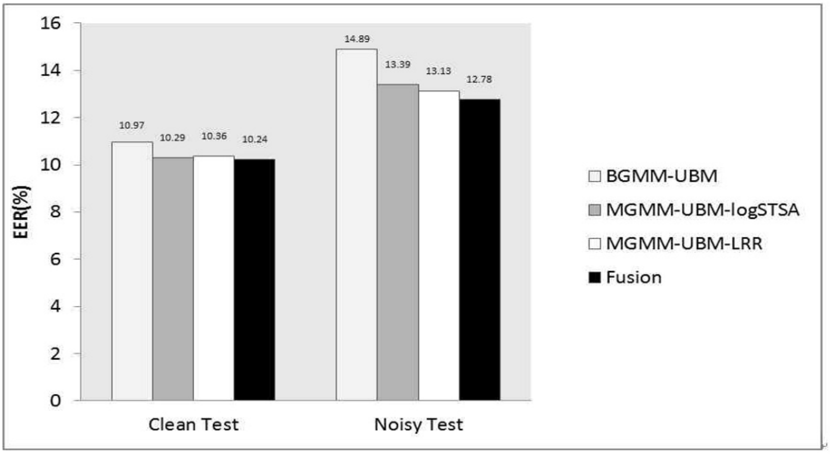


Fig. 5. Performance comparison of the fusion of frames SNR values using the MMSE log-STSA and the LRR with other methods.

4 Conclusion

We have proposed to use the LRR to compute the observation confidence value applied in the modified adaptive GMM training for the speaker verification system. To improve the performance of the system, we introduce to apply a combination of frame SNR values calculated by the MMSE log-STSA and the LRR. The proposed approaches are more robust and effective in both clean and noisy environments. The results have proven that our proposed approach provides better performance than the baseline GMM-UBM approach. In the future, we will consider different measurements to calculate observation confidence value and inter-speaker and channel compensation methods.

Acknowledgments. This research was supported by the Ministry of Education, Science Technology (MEST) and National Research Foundation of Korea (NRF) through the Human Resource Training Project for Regional Innovation.

References

1. Bimbot, F., et al.: A Tutorial on Text-Independent Speaker Verification. EURASIP Journal on Applied Signal Processing, Vol. 4, (2004) 430–451
2. Reynolds, D.A., Quatieri, T.F., Dunn, R.B.: Speaker Verification using Adapted Gaussian Mixture Models. Digital Signal Processing, No. 10, (2000) 19–41
3. Campbell, W.M., Sturim, D.E., Reynolds, D.A., Solomonoff, A.: SVM based Speaker Verification using a GMM Supervector Kernel and NAP Variability Compensation. In Proc. IEEE ICASSP, Vol. 1, (2006) 97–100
4. Dehak, N., Kenny, P., Dehak, R., Glembek, O., Dumouchel, P., Burget, L., Hubeika, V., Castaldo, F.: Support Vector Machines and Joint Factor Analysis for Speaker Verification. In Proc. IEEE ICASSP, (2009) 4237-4240.
5. Dehak, N., Kenny, P. J., Dehak, R., Dumouchel, P., Ouellet, P.: Front-End Factor Analysis for Speaker Verification. IEEE Trans. Audio, Speech, Lang. Process., Vol. 19, No. 4, (2011) 788-798
6. Li, M., Zhang, X., Yan, Y., Narayanan, S.: Speaker Verification Using Sparse Representations on Total Variability I-Vectors. Proc. Interspeech, (2011) 4548-4551
7. Trinh, T.D., Park, M.K., Kim, J.Y., Lee, K.R, Choi, S.H., Cho, K.S.: A Modified Adaptive GMM Approach Based GMM Supervector and I-Vector Using NMF Decomposition for Robust Speaker Verification. Journal of KIIT, Vol. 13, No. 7, (2015) 117-125
8. Trinh, T.D., Park, M.K., Kim, J.Y., Lee, K.R, Cho, K.S.: Enhanced Speaker Verification Using GMM-Supervector Based Modified Adaptive GMM Training. Lecture Notes in Electrical Engineering, Vol. 310, (2015) 147-153
9. Ephraim, Y., Malah, D.: Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator. IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 33, No. 2, (1985) 443-445
10. Huang, P.S., Chen, S.D., Smaragdis, P., Johnson, H.M.: Singing-Voice Separation from Monaural Recordings Using Robust Principal Component Analysis. Proc. IEEE ICASSP, (2012) 57–60
11. Hermansky, H., Morgan, N.: RASTA Processing of Speech. IEEE Trans. on Speech and Audio Proc., Vol. 2, No. 4, (1994) 578-589
12. Pelecanos, J., Sridharan, S.: Feature Warping for Robust Speaker Verification. In Proc. Speaker Odyssey, Crete, Greece, (2001) 213–218

A SDN-based Network Intrusion Detection System to Overcome UPnP Security Drawbacks

Duc-sun Lim, Jin-yeong Kang, and In-whee Joe

Department of Computer and Software, Hanyang University,
Seoul, 133-791 Republic of Korea
Corresponding Author: In-whee Joe
{imcoms, achieve365, iwjoe}@hanyang.ac.kr

Abstract. Internet of Things (IoT) can be defined as an intelligent infrastructure service that provides the means by which information can be exchanged through the interconnection of devices and people. While IoT services continue to be developed, the need to overcome its drawbacks is great. One weakness of IoT is that its technical characteristics expose it to diverse security threats. Because technical equipment has several limitations including restrictive hardware options and low power, it is vulnerable to physical attack or other security threats. In particular, universal plug and play (UPnP) plays a crucial role in connecting heterogeneous devices for IoT services to enable information exchange between them. However, UPnP has inherent security-related drawbacks because it is operated using simple service discovery protocol and user datagram protocol (UDP). Solving the issues of network security and access by unauthorized users is now critical. In this paper, we analyze security vulnerabilities of UPnP, define the rules for distributed Denial-of-service attack situations, attempt to overcome the drawbacks of Software-defined-networking (SDN)-based network intrusion detection systems. Also, we propose to establish an internal IP address of a network device in the virtual environment and present rules to prevent attack scenarios.

Keywords: IoT, UPnP, SDN, Stack Overflow, DDoS.

1 Introduction

With the recent dissemination of various smart devices and new developments in network technology, we have entered the era of what is known as Internet of Things (IoT), in which all devices are connected to the Internet and share information. IoT technology, which is based on conventional network technology, is an upgrade to the Internet whereby devices connected to the Internet exchange data on their own. While IoT is similar to conventional machine to machine (M2M) protocol in that it exchanges information independent of humans, it can be seen as a more developed concept that enables expanded communication among not only humans but also between objects connected to the Internet. The economic value creation effect of IoT by the year 2022 is estimated to surpass $14 trillion[1]. The forecast for 2020 is that approximately 50

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_13

billion devices will be connected to the Internet[2]. Currently, a high level of technological competition is occurring to satisfy the technical needs of devices and network equipment. Among these, universal plug and play (UPnP) technology is a software architecture designed to identify the networks of all home appliances and connect them to other devices.

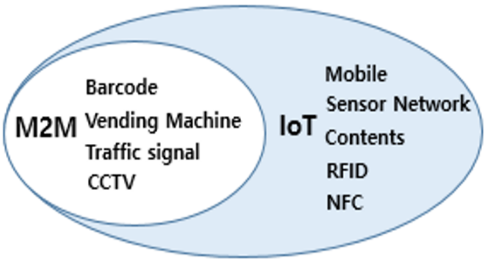


Fig. 1. M2M and IoT Concept Map

UPnP uses the simple service discovery protocol (SSDP) and user datagram protocol (UDP), which are based on TCP/IP technology, and is independent of other networking equipment [3]. One advantage of using UPnP is that it enables automatic identification through its connection to the Internet and use of web protocol. In addition, neither installation of other software nor additional device settings on PCs and surrounding devices is required. To satisfy the technical needs of present network devices, UPnP can connect each device without authorization and enables free composition. IoT combines various technologies. With the linkage between applications and heterogeneous devices, various security-related threats can occur. Therefore, a need exists to address not only these IoT-related threats, but simultaneously those that occur in infrastructure service environments [4]. Because UPnP operates using SSDP and UDP protocols, it is vulnerable to these many security threats. This study proposes a software defined networking (SDN)-based network intrusion detection system the objective of which is to overcome the drawbacks of UPnP. In addition, we propose rules relate to intrusive attacks.

2 Related Works

2.1 SDN(Software Defined Networking)

In conventional network devices, the control and data processing functions are not separated. Because network devices are controlled individually, they must be managed independently of one another. The network must be built to match the functions of the devices it services. When adding devices to the network, settings for all networks must be changed individually. Therefore, network administrators face difficulties in terms of management. In addition, because networks are hardware-centered, they incur high costs and have unique characteristics that make them difficult to configure. However, unlike in a conventional network, in an SDN, the control plane and data plane are separate.

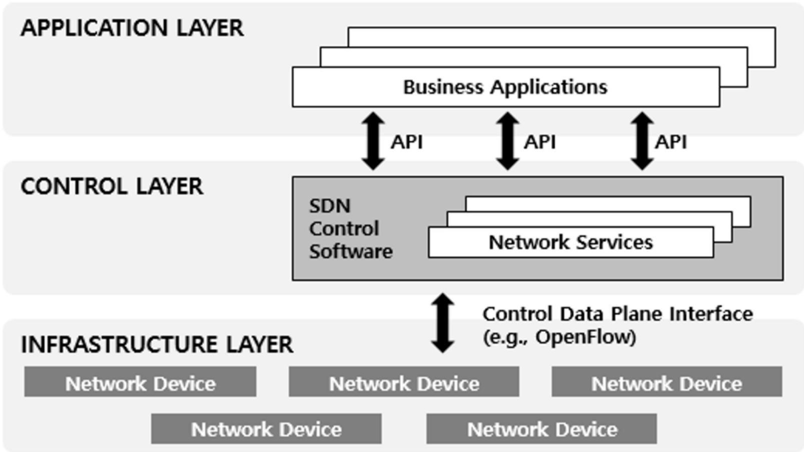


Fig. 2. SDN Structure Map

As shown in Fig. 2, SDN has a three-level architecture[5]. Network devices such as switches or routers are included at the infrastructure level. At the control level, where control of the overall network occurs, the essential function of the network is located using the SDN controller that is based on software programming. The controller can be managed by programming network functions. At the application level, an application is required to control the network. Adding the SDN controller simplifies the network device, thereby reducing the network administrator's burden of manually changing settings. It also enables resources to be optimized in terms of network settings and security. Regarding security, SDN assists the administrator in establishing high-level policies through OpenFlow. The OpenFlow-based SDN technology reduces the burden of individually setting network devices when a change in policy occurs. By managing the networks from a central location, SDN has been evaluated as a superior system in terms of security and stability compared to conventional networks[6].

2.2 Overview of Upnp and Security Vulnerability

Many policy security domains exist within the home network of UPnP. These can be applied to a virtual home network without restrictions on physical space. Encryption or authorization by device units is conducted, and standards are defined to support security service between the control point that transmits security orders and those devices that process the specific orders they receive[7][8][9]. In short, UPnP is a protocol in which users can be automatically allocated an address **without network settings** inserted malicious code **having to be accessed**. It operates using the following procedures. The UPnP security standard not only provides security services to customers who simply use devices, but also provides services for each device manufacture for the authorization conducted within the home network. It also provides an access control function based on an access control list (ACL) or authorization certificate for device units. Moreover, for simple object access protocol message units, it provides integrity and confidentiality services based on the selected methods of customers.

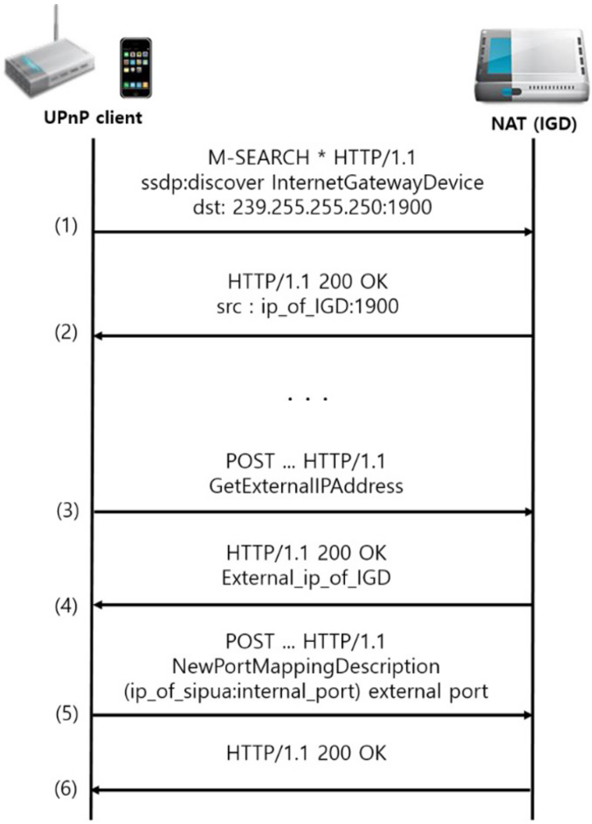


Fig. 3. Potential threats of UPnP service exposure to the internet

Access control for devices is conducted using an ACL, a group membership certificate, and authorization certificate. The ACL of devices consists of agent, certificate information, proxy, and expiration date, and is managed by each device. The security service for a device unit is regulated by the control device that has control rights for that control device unit, and consists of a section key exchange flow, as well as ACL edit-related flow, ownership-related flow, and authorization certificate-related flow of each device[10][11][12].

UPnP security uses the technology and standards for communication between devices. Users can monitor and control devices that are connected to the home network control at all times. In the case of a major event, the device can notify a user[13]. Although security regulations exist for devices in the UPnP-based home network, unauthorized users can obtain easy access because no separate security regulations exist for users. The vulnerabilities of wireless LANs are also cited as a major drawback.

2.2.1 Remote control

The most cited drawback in UPnP concerns the seizing of the highest access permissions by external users with bad intentions. In other words, even when one has access to the device based on general user access rights, he or she can upgrade authority to possess greater control. Filling a manipulated packet in the UPnP with meaningless data (NOP) and inserted malicious code lead to a violation of access restrictions to the target system. As the following example shows, in the unsupervised relation identification field of port, protocol, and URL, NOP increases the buffer, and transmitting the session with a cycle of 10,000 ms causes a violation of access restrictions in the system.

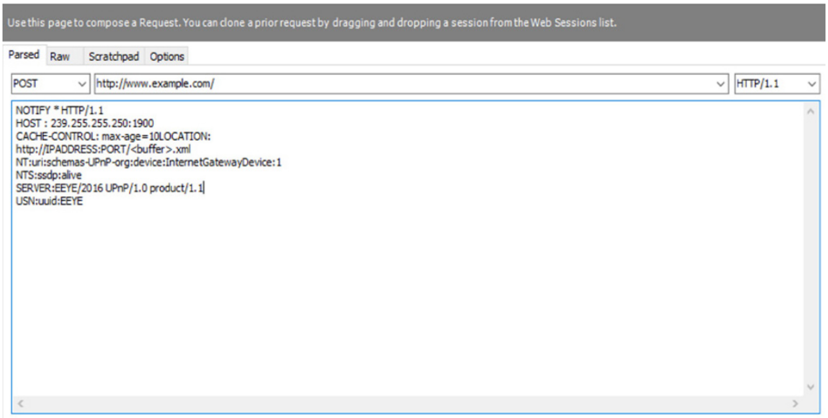


Fig. 4. Protocol parse results in an UPnP environment

When this is considered from an IoT equipment perspective instead of that of the manager's through packet parse, IoT can be defined as a buffer overflow attack. Because buffer overflow attacks have been occurring since the beginning days of the Internet, many preventative and defense methods has been introduced. However, buffer attacks even occur in IoT devices connected to a home network is a major issue that prevention can be applied even to the IoT devices that are connected to the home network. When an attack has been successful moving beyond being possible, the hacking range cannot be known. Since the problem resides within the OS itself, more time and research is needed to treat it.

3 Proposed Method

This study proposes a method of automatically detecting illegal traffic using an SDN-based switch. Detection of malicious traffic in a conventional network environment is difficult. However, in an OpenFlow network, network security and stability can be ensured based on the policies established by the manager. This can enable the identification and processing of all traffic for DDoS attacks on the network. When an SDN control plane is used to control the data plane, real-time calculation of the increased value within the network is possible and will help in detecting attacks. Because the network route data are then saved in a flow table, prevention is also possible. In addition, setting the threshold by collecting statistical data about the network, which is based on OpenFlow, can enable migration about networking programming and thus assist in reacting to attacks automatically. Using a centralized SDN-based network SDN grant effective methods for access control or security policy rights for the overall network.

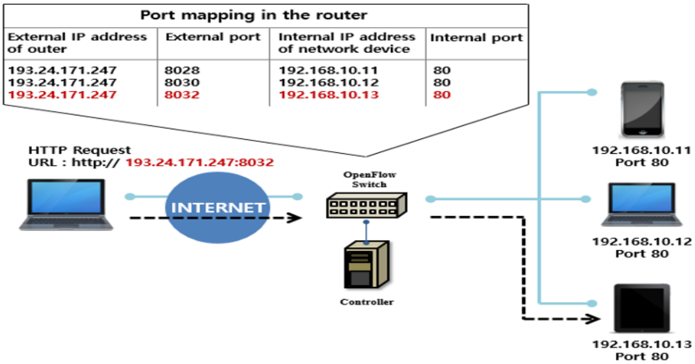


Fig. 5. Malicious packet detection and control structure using an OvS

As the structure in Fig. 5 reveals, the purpose of our method is to detect and ban malicious packets using an Open vSwitch(OVS). Applying and setting the internal "rule" for the OvS are essential. Thus, this study establishes an Internal IP address of a network device in a virtual environment, as shown in Fig. 5. In addition, we propose rules for a new attack method to prevent an attack scenario.

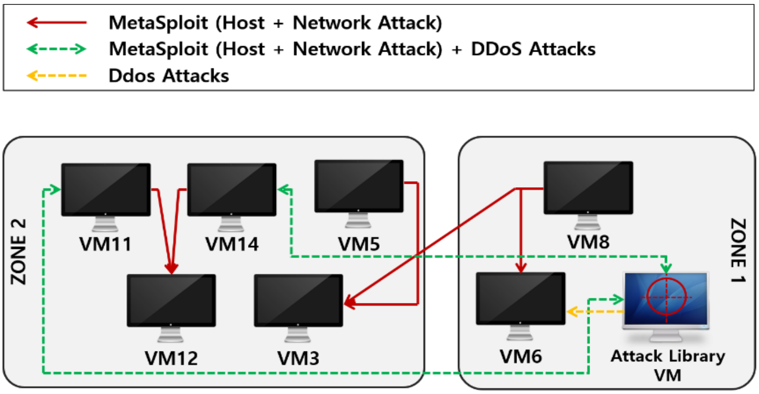


Fig. 6. DDoS attack scenarios in an UPnP environment

Each VM in Fig. 6 contains the devices that are targets. Each attack flow indicates the stack overflow vulnerabilities of UPnP and that an authority upgrade occurs during sequential attacks through DDoS. In short, DDoS interferes with the smooth communication of a home network and causes a network disconnection, thereby preventing users from receiving services. It then connects to the target device to upgraded authorization through stack overflow. As the scenario in Fig. 6 reveals, in an IoT environment, an attack on one target device can enable access to other devices within the network. In addition, because all devices operate organically in an internal IP address environment even when the zone is different, the establishment of a home network can be affected by a single device attack. Therefore, constant monitoring to check the packet modulation in each attack flow is required. In addition, an expanded rule that can actively change to respond fundamentally to the advanced Payload modulation is being required.

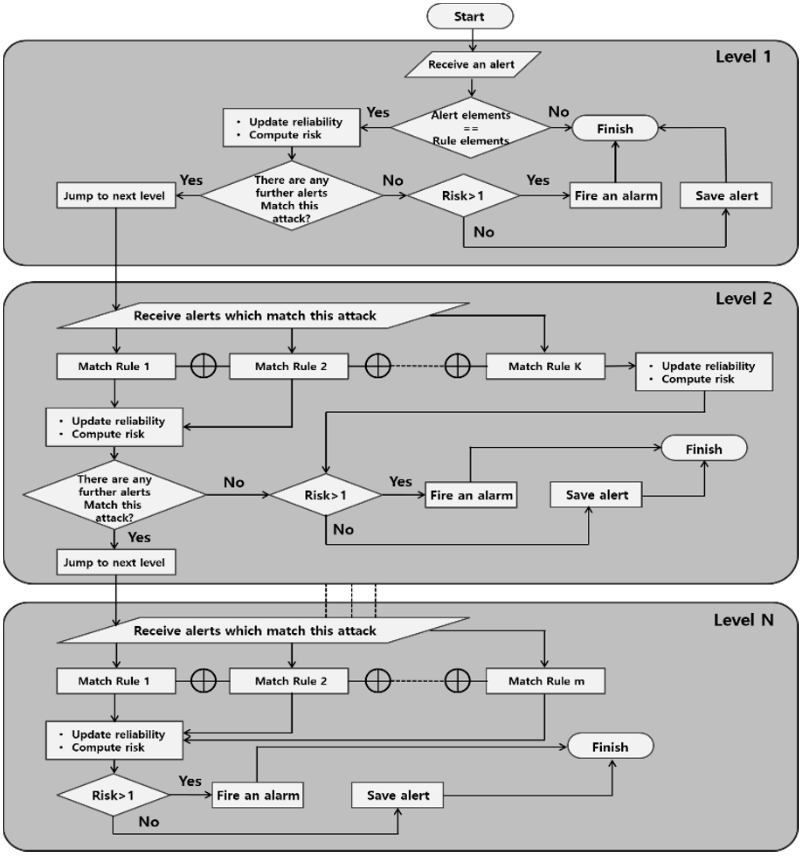


Fig. 7. Rule flow of the spread method

Fig. 7 shows the different security levels. The higher the level, the higher is the detection ratio; the lower the level, the less is the time required between detection and reaction. The rule flow of the spread method that operates within an OvS exceeds the conventional method in which rules are used for operation, and includes the aspect of artificial intelligence in that rules can be automatically created and destroyed.

4 Simulation Results

To prove the superiority of the method proposed in this study, we compared the conventional rule method used with a firewall to the spread rule method with respect to time flow and traffic.

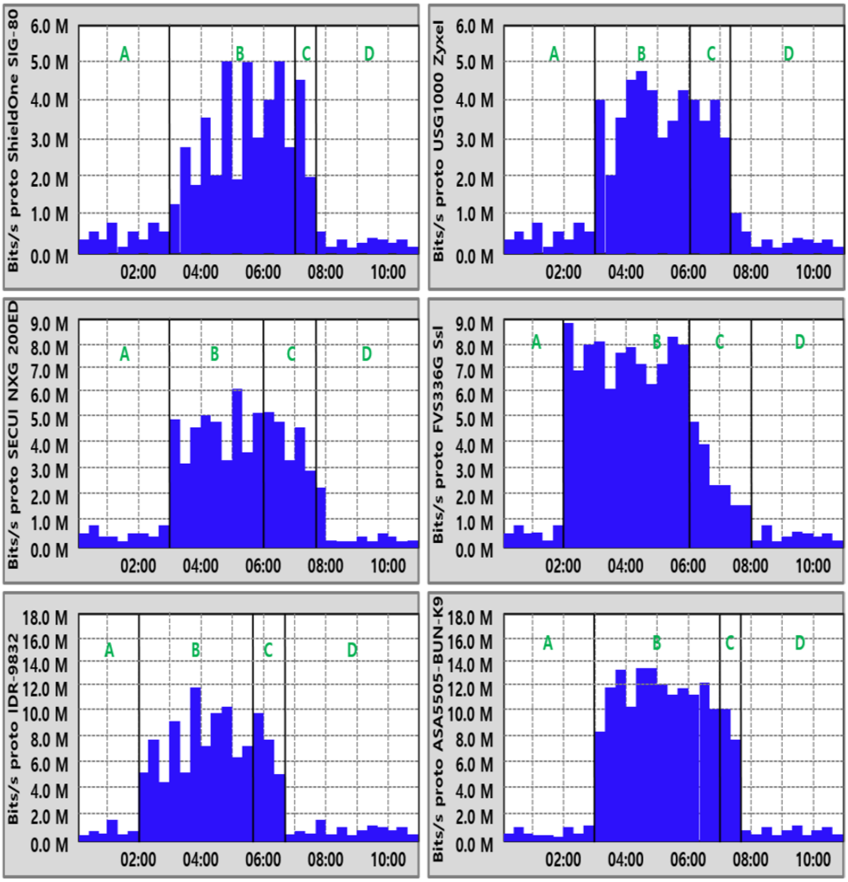


Fig. 8. Rule of spread method used in a firewall

In Fig. 8 is rule of spread method used in an existing firewall[14]. The Y-axis shows the size of traffic (during a DDoS), and the X-axis shows the time flow (time required to resolve problems following a DDoS attack). The sections shown in A, B, C, and D show a normal situation, a DDoS detection and rule creation, solution to the problem of increased traffic caused by a DDoS, and the normal state after the resolutions, respectively. Notice Device (4) in Fig. 8. This indicates the device vulnerable to a DDoS attack, whereas that marked (5) is less vulnerable. Compared to other devices, (4) is faster in detecting DDoS and requires less time to resolve. However, in Section (B) where rules are created, much resource consumption occurs and the traffic increase rate is the highest. Therefore, when traffic further increases because of the existence of many zombie PCs, the network will shut down even before it can react to the situation as shown in Section (C). In addition, DDoS detection is relatively fast in Device (5), with a low traffic increase rate and high problem resolution rate. However, Section (B) shows a dramatic change in the time required to create rules to detect and solve the problem. This means that the contents of the rules are inefficient. The same also applies to the equipment in Device (1).

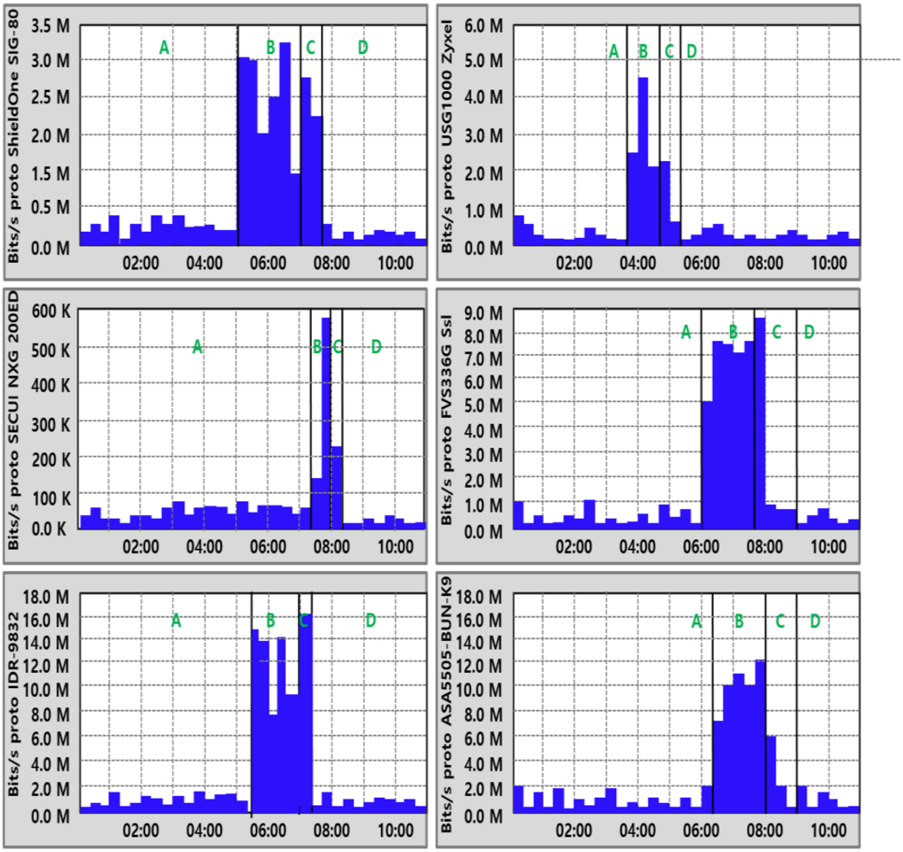


Fig. 9. Rule of spread method used in a firewall

As shown in Fig. 9, the time required to create rules and solve the problem (B-C) has been shortened. However, the time to initiate the rule creation is delayed compared to that in the conventional method. This time refers to the delay caused when processing the traffic prior to the DDoS attack, because the spread rule method processes all of the received traffic. In addition, because the rate of traffic as a result of DDoS has been reduced considerably, it can react better to a DDoS than can the conventional method. We also believe our method is more robust than the conventional method in the event the network shuts down.

5 Conclusion

This study proposed an SDN-based network intrusion detection system as well as rules to overcome security drawbacks of UPnP. Establishing a credible network environment using SDN for IoT services can resolve security drawbacks. It can also optimize network management and operation in isolation from hardware, as well as defend against DDoS attacks more efficiently. In terms of security in IoT, having a stable base requires a stable infrastructure. This is accomplished by means of software-based SDN. Future research will address diverse issues related to security policies in processing and identifying the flow of the OpenFlow standard in an SDN that considers IoT devices. These policies are necessary to solve problems that commonly occur because of software vulnerabilities.

References

1. Evans, Dave. "The internet of things: How the next evolution of the internet is changing everything." *CISCO white paper* 1 (2011): 14.
2. Kopetz, Hermann. "Internet of things." *Real-time systems*. Springer US, 2011. 307-323.
3. Plug, UPnP—Universal. "Play forum." About the UPnP Plug and Play Forum," in http://www. upnp. org. 1999
4. Warriach, Ehsan Ullah, et al. "Heterogeneous device discovery framework for the Smart Homes." *GCC Conference and Exhibition (GCC), 2011 IEEE*. IEEE, 2011
5. McKeown, Nick. "Software-defined networking." *INFOCOM keynote talk* 17.2 (2009): 30-32
6. Sezer, Sakir, et al. "Are we ready for SDN? Implementation challenges for software-defined networks." *Communications Magazine, IEEE* 51.7 (2013): 36-43.
7. Feng Jin, Ohsung Kwon, Ilchul Yoon. Analysis of UPnP-based Sensor Control Technologies Journal of Internet Computing and Services, 2015,5 pp,199-120
8. S._Y. C. Chen, Wally Kuo and Han-Chieh. Enhanced UPnP control and management in a campus network. In MUE'07:IEEE International conference on Multimedia and Ubiquitous Engineering, pages 331-335, Seoul, Korea, April 2007.
9. J.-J. Lee, C.-Y. Huang, L.-Y. Lee, and C.-L. Lei. Design and implementation of secure communication -channels over UPnP networks. In MUE, pp,307–312, 2007.
10. D. Garcia, "Universal plug and play(UPnP) mapping attacks" In DEFCON-19, 2011
11. I. AI-Mejibli and M. Colley, "Evaluating UPnP service discovery protocols by using NS2 simulator, "2nd Computer Science and Electronic Engineering Conference (CEEC), pp1-5, 2010.
12. Sun Junwen, Zheng Zhengqi. Implementation of UPnP based solution for Traversal through NAT[J]. Electronic Technology, 2009, (1), 74-75.
13. KangDa Hu, DongSheng Yin "Multimedia Resource Sharing based on UPnP protocol and embedded device". 2013 International Conference on Computer Sciences and Application
14. R. Penno, D. Wing "Universal Plug and Play(UPnP)"

A Graph-based Method for Performance Analysis of Energy Harvesting Wireless Sensor Networks Reliability

Vladimir Shakhov

Institute of Computational Mathematics and Mathematical Geophysics,
Siberian Branch of Russian Academy of Science,
Novosibirsk, 630090, Russia
{shakhov@rav.sscc.ru}

Abstract. In the near future, networked wireless devices outnumber traditional electronic appliances. Internet of Things will enable a plethora of new applications in environmental monitoring, agricultural industry, medical diagnostics, security and transportation business, in analysis of equipment performance, conditions, troubleshooting and monitoring. Nowadays, the concept of wireless sensor networks is considered as an important part of Internet of Things. Recent advances in energy harvesting technology for wireless devices have increased the lifetime and throughput of wireless sensor. However, sensor networks reliability has to be improved as well. To fill this gap we offer a method for efficient estimation of network performance and reliability.

Keywords: Wireless Sensor Networks, Energy Harvesting, Reliability.

1 Introduction

The Internet of Things (IoT) is evolved from a stage of research and innovation to a stage of market deployment. In the near future, networked wireless devices outnumber traditional electronic appliances. IoT will enable a plethora of new applications in environmental monitoring, agricultural industry, medical diagnostics, security and transportation business, in analysis of equipment performance, conditions, troubleshooting and monitoring. Cisco CEO John Chambers predicted the Internet of Things would be a $19 trillion market over the next several years. Nowadays, the concept of Wireless Sensor Networks (WSNs) is considered as an important part of IoT. Small sensors, which consist of sensing, data processing, and communicating modules, are combined by wireless channels into wireless sensor networks [1]. These networks are intended to be context aware, self-governing, flexible and reliable. WSNs have a wide range of potential customer applications such as smart home [2] etc. With the standardization of IEEE 802.15.4 and ratification of ZigBee standard, wireless sensor technologies have received a significant new surge.

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_14

As WSNs are deployed, system reliability becomes an important requirement. The sensors lifetime (and, hence, system reliability) depends on the battery capacity. However, a cost of sensor nodes is a critical issue in the design of practical WSNs. It is often economically advantageous to discard a sensor rather then sensor recharging. By this reason a battery power is usually a scare component in wireless devices. It was a serious problem for WSNs applications development. In response to growing market demand for practical WSNs, the concept of Energy Harvesting WSNs (EH-WSNs) has been produced. EH-WSNs nodes can harvest energy from the environment (solar, wind, thermal energy). They get very long lifetime due to the large number of recharge cycles. An essential progress in wireless communications, electro-mechanical and digital electronics technologies allows to deploy a large-scale EH-WSNs.

If EH-WSNs nodes increase the sensing or transmission range then the energy consumption is increased and vice-versa. An increased transmission range allows to connect bigger amount of nodes. It means the WSNs connectivity is increased, and hence, network productivity (reliability, latency etc.) is potentially improved. At the same time, the duration of energy-harvesting mode has to be increased. Therefore, sensors availability becomes potentially worse. WSNs productivity is also reduced. Thus, it is useful to get a tradeoff between the energy harvesting mode and WSNs connectivity. To fill this gap a problem statement for EH-WSNs performance improvement is formulated and investigated.

The rest of this short report is organized as follows. The basic notations are presented below. The applied network model is described. The optimizing problem for efficiency improvement of EH-WSNs is offered. Next, we describe the model for efficiency estimation of network nodes. Finally, the brief acknowledgement is provided.

2 Problem Statement

Generally WSNs topology is modeled by a UDG graph. For our purposes it is convenient to model EH-WSNs by a probabilistic UDG graph $G = (V; E)$, whose vertices set V represents the sensors and whose edges set E represents the channels between adjacent sensors. A sensor is independently randomly available with an associated probability. This probability can be interpreted as the graph node reliability. The wireless channels can be also unreliable due to harsh environment. The transmission ability of sensor can be defined as follows

$$P(e) = \begin{cases} d, x < r \\ 0, \text{otherwise.} \end{cases}$$

Where $P(e)$ is the probability of edge e existence, $e \in E$; x is the Euclidean distance between the sensor and the adjacent node; r is the transmission range of this sensor; the probability d is defined by WSNs environment, sensors parameters, MAC protocols etc.

For the number of network elements the following notations are used: $|V| = N$, $|E| = M$. Let us define EH-WSNs productivity the functional F

$$F : G(V,E) \to R^{+}, \quad \text{or } F(G)$$

For example, if the network reliability is considered then F can mean the probability of EH-WSNs structure connectivity, the mathematical expectation of disconnected pairs number etc.

A sensor node randomly comes to a the energy harvesting mode and returns to a full working mode. It depends on energy harvesting conditions. A sensor battery can be exhausted and the sensor can be switched. Therefore, a network node is randomly available. Let p be the probability of the node availability. In the considered case, if p increases then the transmission range is reduced and the number of links between sensor nodes decreases. And vice-versa, if the sensors transmission rage is reduced then the energy harvesting period and the intensity of cutoff can be reduced and p is increased.

Thus, EH-WSNs topology is described by random undirected graph G with M edges (all links in EH-WSNs) and N nodes (the number of sensors in EH-WSNs). Let p_i be the accessibility estimation of the sensor $v_i \in V$; if all nodes are homogeneous then the designation p is used. In the context of considered problems, this assumption does not reduce generality. Let us use the designation d_i for the reliability of the link $e_i \in E$; if all channels are in the same environment then the designation d is used.

The value of p is defined by network operation strategy, in other words, p depends on the energy harvesting intensity. As it was mentioned above p depends on sensors transmission range r. Therefore, the sensor node availability is a monotonically decreasing function in r,

$$r_1 > r_2 \Rightarrow p(r_1) < p(r_2)$$

In some cases it can be supposed that the links (and the graph edges) are perfectly reliable. However, the number of edges M negatively correlates with p. Any two nodes are adjacent if and only if their Euclidean distance is at most r. That is, for arbitrary $u,v \in V$, it holds that

$$(\mathrm{u},\mathrm{v}) \in \mathrm{E} \Leftrightarrow \| u,v \| \leq r.$$

and

$$r_1 > r_2 \Rightarrow M(r_1) > M(r_2).$$

The considered functions depend on concrete mechanisms of energy harvesting and transmission models. Details for transmission models can be found in the papers [3-4]. If we focus on energy consumption and the battery depletion problem then the choice of transmission range becomes equivalent to determination of energy harvesting schedule.

Now, let us formulate the problem of EH-WSNs efficiency improvement. The corresponding mathematical statement is as follows

$$\underset{\mathrm{G,r}}{\mathrm{Arg}}\{F(G(V,E(r)), p_1(r),..., p_N(r), d_1,..., d_{M(r)}) \geq L_{QoS}\}. \qquad (1)$$

Where L_{QoS} is a required level of system throughput.

In the case of homogeneous nodes and perfectly reliable links we get

$$\underset{\mathrm{G,r}}{\mathrm{Arg}}\{F(G,p) \geq L_{QoS}\}.. \qquad (2)$$

In some practical cases it is reasonable to constrain the number of neighboring sensors owing to MAC protocols specificity or interference problems. Hence, additional conditions have to be added to the problem statement.

3 Sensor Availability Model

To solve the problem (2) we need a method of node availability estimation (the value p). In this section we consider this problem. The offered model of the sensor's behavior is based on Continuous Time Discrete States Markov process. It is assumed that all nodes of WSNs are unreliable. A node of WSNs can get the stages as follows. Active stage (A) – a sensor is fully operable. Next, the stage (EH) – an energy harvesting mechanism is activated. OFF stage – depletion of battery. A sensor is failed and one waits the stage of energy harvesting. The states diagram is described on Fig. 1.

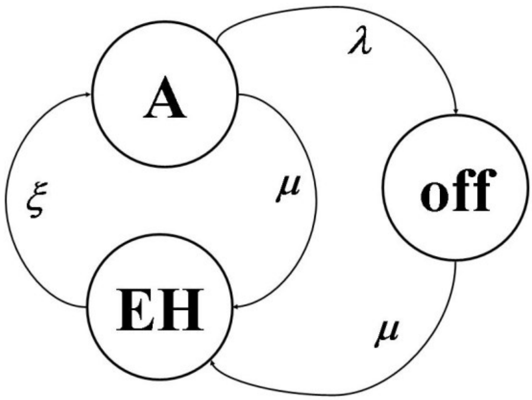


Figure 1. The states diagram

Let the intensity of battery exhausting be λ. The intensity of battery restoration is ξ. An energy source provides energy harvesting facilities with intensity μ.

Let us make the following designation: P_A, P_{EH}, P_{off} are the probabilities of fully operable state, energy harvesting state and failure state correspondingly. The index of probability corresponds to the state designation.

Using the Kolmogorov equations for a continuous-time Markov process

$$\frac{dP_A}{dt} = -P_A(\lambda + \mu) + P_{EH}\xi,$$

$$\frac{dP_{off}}{dt} = -P_{off}\mu + P_A\lambda,$$

and the normalization conditions for states probabilities, we get

$$P_A = \left(1 + \frac{\lambda}{\mu} + \frac{\lambda + \mu}{\xi}\right)^{-1},$$

$$P_{EH} = \left(1 + \frac{\xi}{\mu}\right)^{-1},$$

$$P_{off} = \frac{\lambda}{\mu} P_A.$$

The offered results allow to estimate availability of sensors. The probability of EH stage does not depend on λ. Actually, the procedure of energy harvesting is available from any stages and has got one's own schedule. It is not defined by traffic intensity or any workload.

Let us remark that

$$\lim_{\lambda \to \infty} P_{off} = \left(1 + \frac{\mu}{\xi}\right)^{-1}.$$

It means that the effect of Depletion of Battery (DoB) attack (or any other energy exhausting attacks) is limited in EH-WSNs.

4 Performance Evaluation

The offered model allows to get performance comparison between strategies of sensor availability improvement. For example, let us consider a situation of sensor battery exhausting. It can be caused by DoB attack, jamming, inefficient protocol design etc. [6,7]. Let us compare two methods of counteracting against the attack.

Table 1. Performance Comparison

IDS				EH++			
λ	μ	ξ	p	λ	μ	ξ	p
10	1	5	0,075758	10	1	5	0,075758
9	1	5	0,083333	10	2	5	0,119048
8	1	5	0,092593	10	3	5	0,144231
7	1	5	0,104167	10	4	5	0,15873
6	1	5	0,119048	10	5	5	0,17
5	1	5	0,138889	10	6	5	0,170455

First, an intrusion detection system (IDS) is activated and the intensity of battery exhausting is reduced. Second, a frequency of energy harvesting is increased (EH++). However, the energy source power is not changed. The results of performance comparison is presented in the Table 1. As we can see the choice of the counteracting method is ambiguous. if it is possible the mixed strategy has to be used.

Remark, it is generally assumed that sensors throughput is limited under energy harvesting mode. Thus, all defense mechanism can lead to QoS or information quality degradation. However, it is reasonable action under intrusions.

5 Conclusion

In this paper, a graph-based method for performance analysis of EH-WSNs is offered. The system topology is molded by random UDG graph. EH-WSNs throughput is described by the function of this graph. We consider the graph connectivity criterion. However, the approach can be extended to other appropriate functions of random graphs, where the graph represents a topology of WSNs. In this random graph the choice of probabilities for node and edges depends on transmission range, traffic intensity, intrusions details, monitoring requirements, and energy harvesting intensity. For calculation of the corresponding probabilities it is offered to use the simple Markov chain-based model. The proposed approach allows can be applied for performance comparison of EH-WSNs organization designs.

Acknowledgments. This research was supported by the Basic Research Program of the Presidium of the Russian Academy of Sciences.

References

1. I. Akyildiz, A survey on sensor networks. IEEE Commun. Magazine 40 (2002)102-114.
2. J. Byun, B. Jeon, J. Noh, Y. Kim, S. Park. An intelligent self-adjusting sensor for smart home services based on ZigBee communications, IEEE Trans. Consum. Electr., 58 (2012) 591-596.
3. Y. Zou, K. Chakrabarty, Sensor deployment and target localization based on virtual forces, in: Proceedings of the Annual Joint Conference of the Computer and, Communications (2003) 1293–1303.
4. Z. Lin, S. Zhang, and G. Yan, An incremental deployment algorithm for wireless sensor networks using one or multiple autonomous agents, Ad Hoc Networks, 11 (2013) 355-367.
5. V.V. Shakhov, Experiment Design for Parameter Estimation in Sensing Models, Springer LNCS 8072 (2013), 151-158.
6. Shakhov, V.V., Choo H.: Reliability of Wireless Sensor Network with Sleeping Nodes. In: International Conference on Computational Science 2007. LNCS, vol. 4490, pp. 530–533. Springer, Heidelberg (2007)
7. Shakhov, V.V.: Protecting Wireless Sensor Networks from Energy Exhausting Attacks. In: ICCSA 2013. LNCS, vol. 7971, pp. 184–193. Springer, Heidelberg (2013).

Part III
Mobile Data Management and Applications

High-Performance Fault-Tolerant Data Caching and Synchronization Architecture for Smart-Home Mobile Application

Ramesh Guntha, Aryadevi Remanidevi Devidas, Maneesha Vinodini Ramesh

Amrita Center for Wireless Networks and Applications, Amrita School of Engineering,
Amritapuri Campus, Amrita Vishwa Vidyapeetham University
Kollam, 690525, Kerala, India
{rameshg, aryadevird, maneesha }@am.amrita.edu;

Abstract. Mobile devices are becoming the preferred choice for internet access as they are getting increasingly powerful and affordable. But because of lack of ubiquitous high bandwidth wireless internet, many mobile applications suffer from performance and reliability issues while accessing data from the servers. While many of the applications use caching mechanism to store data locally on mobile device to improve data access performance, they are not thoroughly focusing on related areas such as two-way data synchronization, fault tolerance and recovery, offline mode support and real-time update support. And hence even many reputed applications show inconsistent and out-of-date data; especially during network and battery outages. In this paper we propose usage of Replication process, Pending data process, targeted data update broadcasts to solve the above issues in the software architecture of Smart-Home project.

1 Introduction

Mobile devices are becoming increasingly powerful in terms of processing power, storage and memory, and they are becoming increasingly affordable to common man. But the availability of ubiquitous high bandwidth wireless network is still a dream, especially in the developing countries. Apart from bandwidth, many mobile devices also suffer from poor battery backup.

Because of bandwidth issues many mobile applications suffer from performance and reliability issues while accessing data from servers. Many mobile applications often use local data caching on mobile device to avoid going to server for every data request, but they fail to thoroughly take care of related functionalities such as synchronization, replication, fault tolerance and recovery, offline-mode etc. Due to this, even many reputed applications like Twitter, Instagram, Evernote, Dropbox etc., display inconsistent or out-dated data to users under certain fault conditions, some of them do not allow offline mode activity, some of them hang under fault conditions [1], [2], [3].

The Smart-Home architecture presented in this paper shows how these issues can be addressed. The relevant data is stored on the mobile and the Replication and Pending data processes which run on the mobile device perform the data synchronization with the server. The Replication process performs the server to mobile data

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_15

synchronization. User's data updates on mobile are synchronized to server by the Pending data process. These data updates can be done in offline mode as well. The targeted data update broadcast mechanisms ensure that user sees the data updates happening to relevant entities on server in real-time. They take care of complex inter-dependent tables as well. Conflicting data updates to server are resolved based on the server time-stamp. Only the data related the particular user is considered for synchronization. The sync mechanisms are triggered on periodic-basis, based on events such as network recovery, application start, server data updates, and permission changes, thus are able to recover from network and battery outages quickly.

The rest of the paper is organized in following sections. Smart-Home application description, deep-dive into the proposed architecture, analysis of related work and conclusions.

2 Smart-Home Application

Lot of power wastage occurs due to carelessness of users. Bigger organizations can save thousands of dollars per month if they properly manage the power consumption. Smart-Home application is being developed to take care of this need.

2.1 Functionality

Smart-Home application helps in controlling the electric equipment in home/office building(s) it has both mobile and web client interfaces. The users can configure various communities, buildings, floors, rooms, common areas which they control. Under each of these locations they can add the electric equipments to be monitored and controlled. Users get the real-time status updates and the power consumption statistics for each entity they have access to. Users can control the equipment from both mobile and web applications. The updates done in offline mode on mobile would be propagated once the network becomes available. Users can also set pre-configured rules to control how much power can be consumed at an entity level.

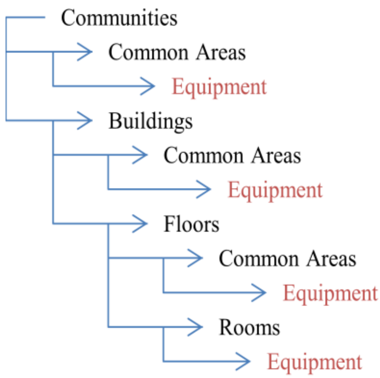


Fig. 1. Entity relations in Smart-Home application

2.2 Entity Relations

Fig. 1 describes the entity hierarchy in Smart-Home application. Community is the highest. Communities may have some common areas like streets, grounds etc. A community contains a set of buildings. Buildings may have some common areas like lobby, back-yard etc. A building contains set of floors. Floors may have some common areas like corridors, lobbies etc. A floor contains set of Rooms. Rooms do not have common areas. Finally the electrical equipment is placed under rooms and various common areas.

2.3 Roles and Permissions

Smart-Home application has two types of roles, Admin and Viewer. These roles are assigned to a particular user at a given entity level and the same privileges apply for all the child entities. Admin can control entities and view power consumption data and status. Admin also can create child entities under the entity. Viewer can only view data and status for the given entity and its children. If a user is admin for a Building, then he/she is the admin for Floors, Rooms and Equipment in that building. A given entity can have more than one user as Admin and/or Viewer, similarly, a given user can be Admin and/or Viewer for more than one entity.

2.4 Data Estimations

In a given community, such as a university campus, there are 10s of buildings, each building would have 3-4 floors, each floor would 10-20 rooms, each room would have around 10 electric equipment such as lights, fans, air conditioning equipment etc. In addition to this there could be several electric equipment like lights, water purifiers, cameras etc in common areas such as grounds, streets, lobbies, corridors etc. So for a given community we are looking at around 10,000 to 20,000 equipment. This could amount to huge data if status and power consumption details for each equipment as it's turned on/off is stored and broadcasted in real-time.

3 Smart-Home Architecture

In this section we elaborate on overall system architecture, mobile client database, replication process, pending data process, real-time data updates to mobile clients.

3.1 System Architecture

The system provides both mobile and web interfaces, supported by scalable, high performance, event-driven, and robust server architecture (Fig. 2). The data is exchanged between client and server in the light-weight and universal JSON format. On the server side high performance NGINX used as web server to handle all the media. The Application server is powered by high performance, highly flexible, extensible, robust, and heavily adopted NodeJS server. Express package is used as command router inside NodeJS, and Sequelize as the ORM (Object Relational Mapping) package to interact with MySQL database. The application logic is written in reusable modular fashion. The logic consists of periodic event code to run any scheduled tasks and broadcasting code to push the relevant data updates to clients based on user's permissions, thereby reducing bandwidth and resource consumption. The DSS module runs the various aggregation algorithms on the power consumption data and also executes various configured rules and triggers various alerts and commands to automatically control the electronic equipment.

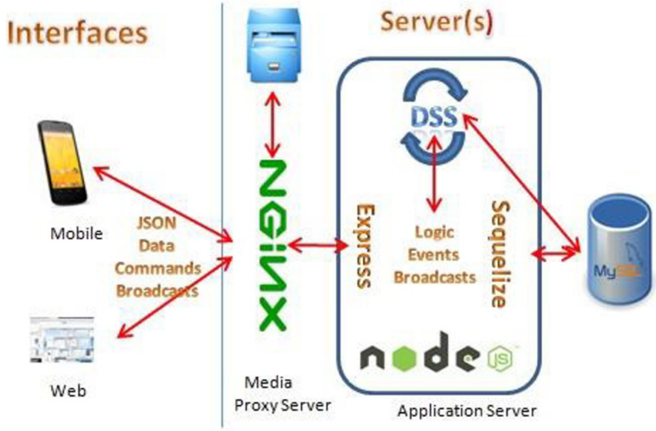


Fig. 2. System architecture of Smart-Home application

3.2 Mobile Database Design Considerations

To provide high performance user experience irrespective of network bandwidth and to support offline mode, all the relevant data for the user is cached on the mobile database. The mobile database is closely mirrored to server database so that it can support all the functionality by accessing the local database itself. In addition to application tables mobile database also has a set of Replication tables for synchronizing data from server to mobile and Pending tables for synchronizing data from mobile to server.

The "modifiedDate" Field. Every row of every table of Smart-Home application contains the last updated timestamp for the row, stored in the modifiedDate field. It plays a critical role in the two-way data synchronization.

3.3 Replication Process for Server to Mobile Synchronization

Replication process depends on Replication tables for synchronizing data from server to the mobile database. Only the relevant data to the user is replicated.

E.R Diagram of Replication Tables. The ReplicationTable table shown in the E.R diagram (Fig. 3) contains list of tables to be replicated on the mobile. The table names are listed in the replicationTableName field. The replicationParentTableId is a self referencing foreign key pointing to the parent replication table record. This relation is used to order the replication requests to server, parent first and then child, and so on. The lastDataDate field stores the "modifiedDate" of the record till which the data from server for this table is replicated on mobile. The lastReplicationDate stores the last successful replication run for this table.

The ReplicationRun table stores the information about all the Replication processes runs. The startDate field stores the date-time at the start of a run, similarly endDate stores the end date-time. The triggeredBy stores the event which triggered that particular run. The replicationStatusId stores the status of replication process. It can be either "Success", "In Progress" or "Failed".

The ReplicationRunDetail is the child table of ReplicationRun table through the replicationTableId foreign key. It stores the replication run information at the table level. The lastDataDate stores the latest modifedDate of a given table replicated from the server. The requestDate stores when the process requested data from server. The responseDate stores the response time from server. The responseStatusId can be either "Success" or "Failed". The responseNumRows indicate the total number of rows promised by server in this request. The responseActualRows indicate the actual number of rows downloaded from server. The replicationStatusId stores the status of replication process for this table. It can be either "Success", "In Progress" or "Failed".

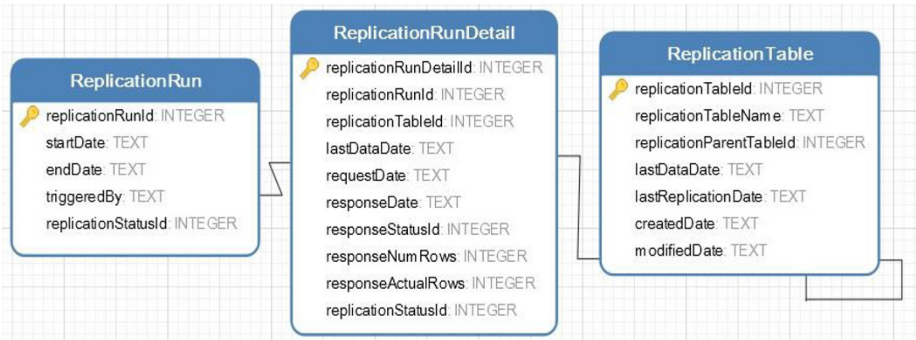


Fig. 3. E.R diagram of replication tables

Replication Algorithm. Fig. 4 shows the replication algorithm. Upon starting the replication process, it puts an entry in the Replication table, and then it finds all tables to replicate from the ReplicationTable's replicationTableName field. Then for each of the tables it makes a server call to fetch the rows which are updated later to the lastDataDate. It records the details of each table replication in the ReplicationRunDetail table. Once all the tables are replicated, it would update the ReplicationRun table.

Incremental Replication of Relevant Data. The replication process is incremental, as it fetches the rows from the server for the corresponding table, only if there are any rows which have modifiedDate later to the lastDataDate in ReplicationTable. This lastDataDate is not affected by time differences between the server and mobile, as it's the latest modifiedDate of server records from the previous replication run. In other words, lastDataDate is based on the server time.

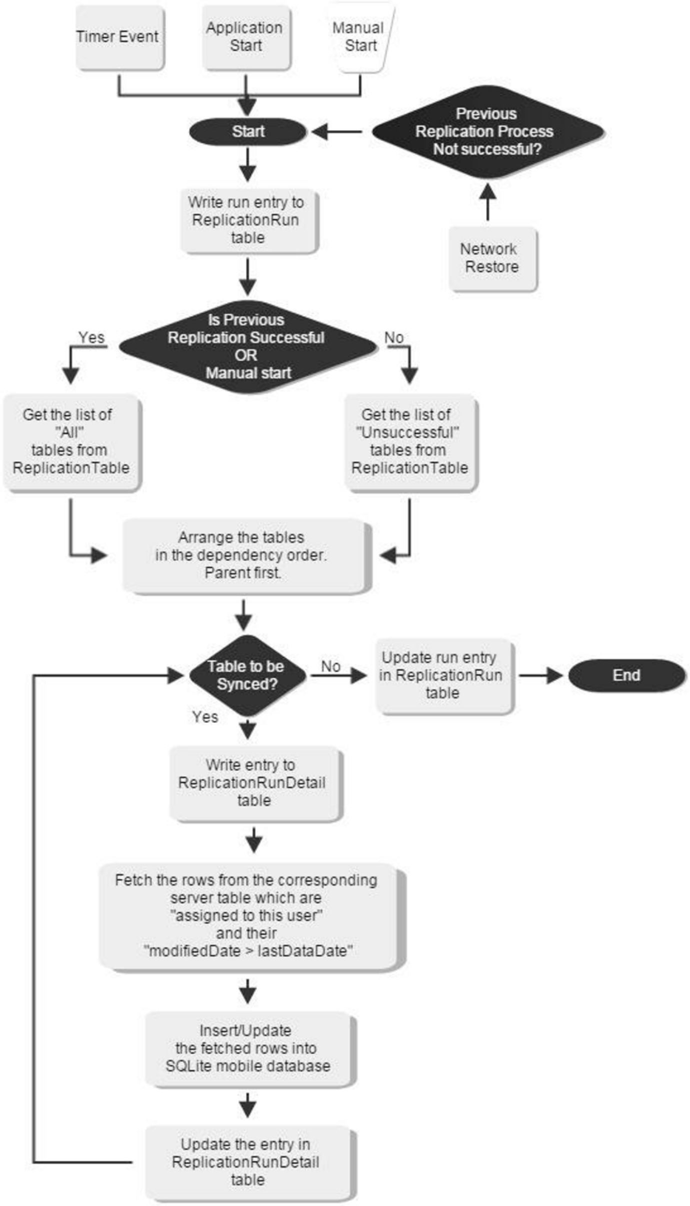


Fig. 4. Algorithm of replication process

The replication process only fetches the entities for which the user is either Admin or Viewer. When the user permission is added or revoked from a given entity, the data comes in through real-time update, and then the real-time update process adds or removes the entity from local database. In case of permission removal, the children of the entity are removed first and then the parent.

Replication of Inter-dependent Entities. Almost any database has parent-child relations between the tables enforced by foreign key constraints. So if a child record is inserted without its parent record being present, it would result in an error. So the replication process must take care of this order. The replicationParentTableId field points to the parent table record in ReplicationTable table, using this field the process replicates parent tables first and then replicates the child tables. This ensures that replication is done without violating foreign key constraints.

Replication Triggers, Fault Tolerance, and Recovery. Replication process can be triggered by periodic timer event, application start, user or network restore event. This ensures immediate recovery from network failures and battery power failures. ReplicationRun table stores whether the previous run failed or successful, so upon network recovery, the process checks this table to see if previous run got stuck, if so, it resumes the replication process. The ReplicationTable stores the table level replication status, so that it can pick-up at which table it's stuck and resume the replication process from then onwards. This ensures no wastage of network and computing resources.

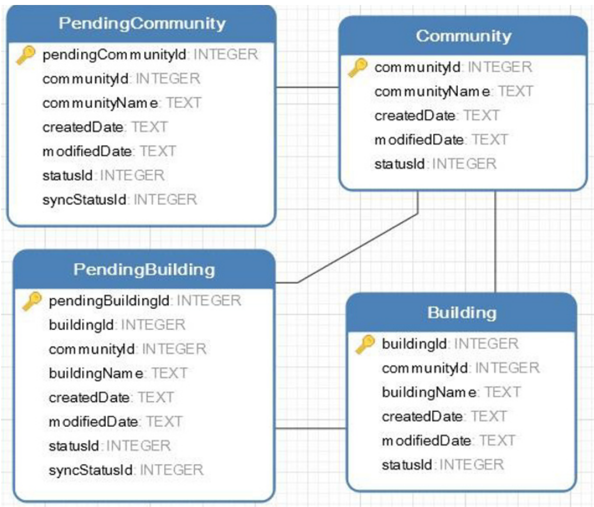


Fig. 5. E.R diagram of pending tables

3.4 Pending Table Process for Mobile to Server Synchronization

The Pending table process propagates updates made by the user on mobile to server. When the user makes an update to any entity, the update is written to the corresponding pending table. Once the update is propagated to server successfully, it is applied to the main table on the mobile. User is allowed to make any number of updates at once, but only the latest update is sent to server and all the others are marked as Overridden.

E.R Diagram of Pending Tables. The E.R diagram in Figure 5 shows pending tables for Community and Building entities. The pending table closely mirrors the corresponding main table. The Pending table has a new primary key instead of the main table's primary key. This is to allow storage of historical updates in Pending table. The syncStatusId is used to store the status of synchronization for the update. It can have any of "Pending", "In Progress", "Retry", "Success", "Failed" or "Overridden".

Lifecycle of "syncStatusId" field.

Fig. 6. Lifecycle diagram of "syncStatusId" field

When the user makes an update, the syncStatusId has the value "Pending" in the PendingTable. Once the process picks-up the change for server synchronization it is marked as "In Progress". If the network is down or server is not reachable, then the status is changed to "Retry", so that it can get picked up by later Pending process run. If the update is failed at Server for any other reason such as "Unique constraint violation" etc, it's marked as "Failed". If the server's record has a later 'modifiedDate' compared to the mobile record being synchronized, it will reject the update with an "Overridden" status. If the update goes through successfully, the syncStatusId is updated with "Success" and the changes are copied to the main table. The Pending table records with "Overridden" and "Failed" status are not retried.

Conflict Resolution. The data on mobile could become out-of-date due to network issues preventing the replication process or real-time update process. If the user updates such stale data it would be rejected by server as Overridden as the server data has later "modifiedDate" compared to that of the mobile update. Similarly, the updates made by the user might get delayed to reach the server due to network or battery outages, or some other user might have updated before this user. In this case also the server would the update as "Overridden". This ensures that conflicts are resolved with policy of rejecting stale data updates.

Synchronization of Inter-dependent Entities. As elaborated in section "Replication of Inter-dependent Entities" above, the replicationParentTableId field from ReplicationTable gives the order of synchronizing the updates to server, following which the Pending table process would avoid foreign key constraint violations.

Pending Table Process Triggers, Fault Tolerance, and Recovery. The pending table process is triggered on periodic basis, on application start, on user update and on network recovery. This ensures that updates are sent to server at earliest possible time to prevent stale data updates as much as possible. This also ensures that pending process resumes immediately after a network or power restoration, resulting in quick recovery from faults. Since pending data is copied to main table only on successful synchronization, network or battery faults, which might stop ongoing pending process in middle, would not affect application functionality, so that's how fault tolerance is achieved.

3.3 Real-Time Mobile Data Updates

Apart from the Pending data process, the mobile data gets updated by real-time targeted broadcast updates from the server. As soon as the user logs into the mobile application, it establishes a persistent stateful socket connection with server, using which the server can push fresh relevant updates that occur on the server database. Since it's an event based, push mechanism, it involves very less network resources and results in real-time data synchronization from server to mobile. When the user updates, inserts or deletes any data on the server either through web application or mobile application, the server derives all the relevant users to be notified using the access permissions, and broadcasts these updates to their mobile apps for synchronization.

4 Related Work

Research to improve mobile application performance by caching data on mobile has been going on for couple of decades [7]. But still as noted by [1], [2], [3] many reputed mobile applications are unable to display consistent and up-to-date data in case of network or battery outages. References [1], [2], [3] take a novel approach in their research to implement a database abstraction layer both on mobile and cloud server for performing data synchronization, replication which is resilient to faults. Since their approach abstracts the database completely, there is no visibility into how the data is stored, hence this approach does not give much freedom to developers and business analysts to access and analyze data from the server. It's reasonable to assume that any enterprise application's team would want to access server database for performance turning, reporting, and analyzing the data for multiple other functionalities. Also [1], [2], [3] do not outline how inter-dependent complex table structures are specified and how to enforce critical database integrity constraints like unique key, not null, or foreign key. Our architecture shows developers how to achieve the synchronization, reliability, fault tolerance and performance while still keeping the complete access to their databases and hence has much better chance of getting adopted. Reference [4] approaches the data-caching problem to store only the relevant data on the mobile in-terms of user own data and user's relationship data, which the user most likely to access on frequent basis, thereby ensuring high hit-ratio. Reference [5] provides a complete mobile-middleware-server architecture for data caching, replication and network fault recovery, but their approach is not scalable as it centralizes the replication and synchronization process on middle-ware server instead of decentralizing on to the mobile system. Due to this updates might take a lot of time to reach mobile and server under high load conditions. Reference [6] surveys various mobile data solutions and categorizes them into the segments of extended client-server systems, data caching systems and adaptive systems. Reference [7] discourages replication anywhere-anytime-anyway approach as it is not scalable because it requires eager and strict mechanisms of caching and transactions. This paper achieves the anywhere-anytime-anyway replication with lazy mechanism with much simpler architecture, which can be easily implemented. Reference [8] suggests a pre-write and pre-read mechanism and mobile host and mobile support station machines for data synchronization which is similar to our Pending data process, but our pending data process outlines mechanisms to synchronize the dependent tables and triggering mechanisms with much simpler conflict resolution mechanism which helps users to understand what went wrong when the updates get overridden or rejected.

5 Conclusion

The in-depth explanation of Smart-Home architecture in this paper demonstrates a way to implement mobile database caching so that it also takes care of two-way data synchronization, conflict resolution, dependent table synchronization, and broadcasting of real-time targeted updates. It shows that architecture is distributed and takes full advantage of storage and processing power of the mobile devices to deliver high performance user experience. It also demonstrates that architecture is fault tolerant and recovers as soon as network or power becomes available.

6 Acknowledgement

This work is funded by Amrita Vishwa Vidyapeetham University as part of “Smart Home” project.

References

1. Y. Go, N. Agrawal, A. Aranya, and C. Ungureanu. Reliable, Consistent, and Efficient Data Sync for Mobile Apps. In FAST, 2015.
2. D. Perkins, N. Agrawal, A. Aranya, C. Yu, Y. Go, H. Madhyastha, and C. Ungureanu. Simba: Tunable End-to-End Data Consistency for Mobile Apps. In Proceedings of the European Conference on Computer Systems (EuroSys ’15),Bordeaux, France, April 2015.
3. N. Agrawal , A. Aranya , C. Ungureanu, Mobile data sync in a blink, Proceedings of the 5th USENIX conference on Hot Topics in Storage and File Systems, p.3-3, June 27-28, 2013, San Jose, CA
4. Q Xing, Y Li, J Wang, Y Han, A User-Relationship-Based Cache Replacement Strategy for Mobile Social Network - Frontier of Computer Science and Technology, 2015 - ieeexplore.ieee.org
5. Ricardo A, Nuno L, Ana A, Lino F and Paulo N: Creating and optimizing client-server applications on mobile devices. In IE 2013.
6. Jing, J., Helal, A., Elmagarmid, A., “Client-server computing in mobile environments”, ACM Computing Surveys, vol. 31, n° 2, 1999.
7. Gray J., Helland P., O’Neil P., Shasha D., “Dangers of replication and a solution” , ACM Int. Conf. On Management of Data (SIGMOD), 1996.
8. Madria, S. K., & Bhargava, B. (2001). A transaction model to improve data availability in mobile computing. Distributed and Parallel Databases, 10(2), 127-160.

Cooperative Big Data Processing Engine for Fast Reaction in Internet of Things Environment: Greater Than the Sum of Its Parts

Yong-Ju Lee[1,2], Hwin Dol Park[1,2], OkGee Min[1]

[1] Big Data Intelligence Research Department, Software & Contents Research Laboratory, Electronics and Telecommunications Research Institute, Daejeon, Korea
[2] Department of Computer Software, University of Science and Technology, Daejeon, Korea
{yongju, park.hwin,ogmin}@etri.re.kr

Abstract. This paper outlines cooperative big data processing engine in Internet of Things environment. Our platform splits an analytical job into two meaningful sub-jobs. First one of two sub-jobs push away from centralized points (e.g., analysis server) to the physical Internet of Things devices (e.g., embedded devices) for filtering out the inconsequential messages and achieving the best message-response possible. Second one of two sub-jobs conducts remaining parts of complicated analytic mission in traditional servers. This approach significantly decreases the data volume that must be moved, the consequent traffic. Furthermore, it can provide faster reaction in most instances.

Keywords: Cooperative Big Data Processing, Fast Reaction, Internet of Things

1 Introduction

Over the past few years, the Internet of Things has gone from theoretical concept to our everyday lived experience. According to a new report from Juniper Research, the number of connected devices will grow from 13.4 billion in 2015 to 38.5 billion in 2020, a rise of over 285%. This explosive growth also leads to a new paradigm of edge computing. Edge computing is pushing the frontier of computing applications, data, and services away from centralized nodes to the logical extremes of a network. It enables analytics and knowledge generation to occur at the source of the data [1]. Furthermore, edge computing can provide faster response in most instances. When the fast reaction becomes more critical, such as smart surveillance, any moving object responding to surrounding conditions, latency is extremely important [2]. The proper architecture will be needed to accommodate the market demands. In this paper, we propose cooperative big data processing engine in Internet of Things environment. The central idea behind Internet of Things(IoT) is that sensors and microchips can be placed anywhere and everywhere to create a collective network that connects devices and generates data. So, we design and implement cooperative big data processing engine to break up large and complex computations into smaller tasks that can be distributed.

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_16

2 Related Works

Internet of Things (IoT) is set to become the next big thing after the introduction of Internet itself. Millions and probably billions of ‘smart’ devices are expected to connect to each other and exchange data and information over the internet. The advocates of the IoT envision nearly all aspects of our life to be covered by these smart devices. The significant increase in connected devices that’s due to happened at the hands of the Internet of Things will lead to an exponential increase in the data that an enterprise is required to manage. There are many research trends to handle tremendous sensing data. MapReduce [3], introduced by Google is one such successful framework for processing large data sets in batch-oriented approach. Nowadays, the shortcomings and drawbacks of batch-oriented data processing were widely recognized by the Big Data community. It became clear that real-time query processing and in-stream processing is the immediate need in many practical applications. Recently, Apache Spark [4] is one of the popular system in the current big data platforms. It is based on resilient distributed datasets (RDDs). This in-memory data structure gives the power to sparks functional programming paradigm. It is capable of big batch calculations by pinning memory.

3 Cooperative Big Data Processing Engine

In Fig 1a), the IoT (Internet of Things) platform diagram is the most common architectural design where the IoT gateway itself is not equipped with sensors. The IoT gateway installed on the device is responsible for collecting data from the sensors, and sending the results to the Analysis server.

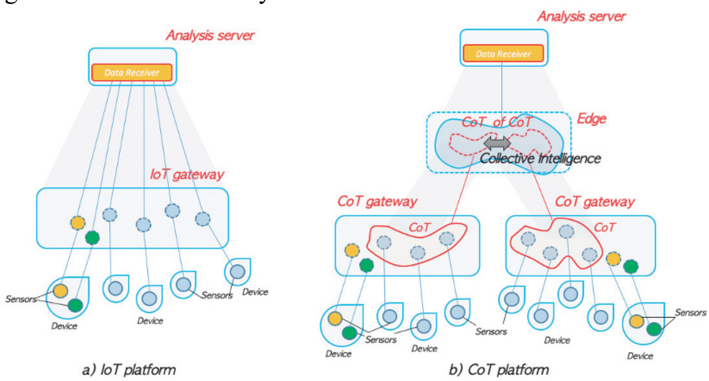


Fig. 1. Concept of two platforms (Internet of Things vs. Cloud of Things)

Fig 1b) represents the CoT (Cloud of Things) platform for handling collective intelligence from arbitrary CoTs. The CoT gateway also installed on the device is responsible for processing of information in the filed, before they’re sent to the Analysis server. Such pre-processing includes message filtering and aggregation.

Furthermore, the edge is responsible for gathering all the necessary metrics from CoT gateways.

Figure 2 shows a detailed workflow of the Cloud of Things(CoT) environment when you submit an anlysis job into our CoT system.

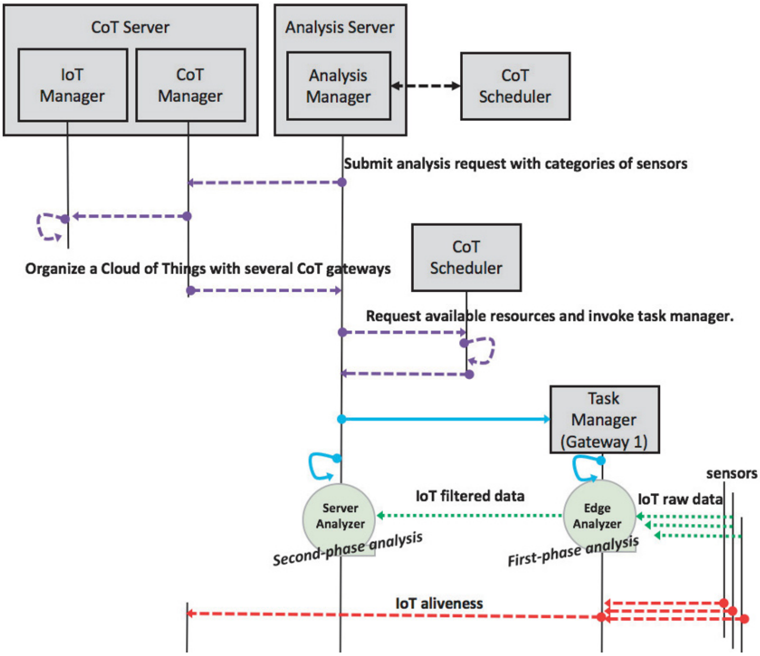


Fig. 2. Cooperative Analysis Workflow

The basic workflow contains the following steps: 1) submit an analysis request with categories of sensors. The sensors can be multiple things with a special-purpose entity. For example, suppose that a request about "*find an average temperature on the first floor*" is submitted to the Analysis Server. The Analysis Server requests the CoT creation to the CoT server "*with a sensor type (i.e., temperature sensors) and geographical locations (i.e., the first floor)*". The CoT server finds the IoT Manager to deal with sensors on the first floor and then it organizes a CoT with several CoT gateways. In this case, "*If the first floor has three rooms and a gateway in each room is existed*", three CoT gateways are chosen. The CoT Scheduler has received resource requests from the Analysis Server and invokes the task "*with the temperature average function*". The Task Manager in gateway 1 (e.g., the first room) collects the temperature data in a certain period. "*If the first room has three temperature sensors which are built in different manufacturer's sensors or different versions of a sensor*", the edge analyzer in the Task Manager receives a temperature data and then refines it with the desired value.

3.2 Cooperative Analytics Job

Fig. 3 shows an example of video analysis job that achieves two processes (i.e., face detection and face recognition). Face detection is the process of automatically locating human faces for video sequences. Once a face is detected, it can be determined if two faces are likely to correspond to the same person. As shown in Fig.3, traditional analysis job conducts face detection and face recognition in server side. Raw video from a camera is transferred into the server. However, the cooperative analysis job has split a face detection and recognition algorithm into multiple processes. These two processes are independent of each other. The face detection algorithm is performed on the edge side. The result of face detection can be transferred into the server side. The face recognition algorithm is performed on the server side. It can be used face-detected images from small bytes of data and reduced the network traffic.

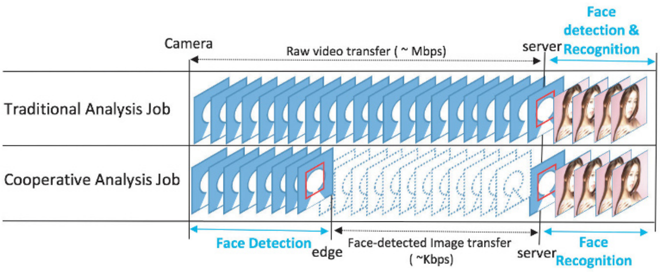


Fig. 3. Example of Video Analysis Job (Face Detection & Recognition)

4 Test Environment

We use the Raspberry Pi for the role of the CoT gateways. The Raspberry Pi [5] is a credit-card-sized, single-board computer. For this project, we were generously given several Raspberry Pis to create a little sensor networks. Fig 4 shows a comparison of traditional analysis and cooperative analysis. We made use of the face detection and face recognition algorithm, from the open source library, OpenCV [6], to implement a prototype that used the Eigenfaces [7].

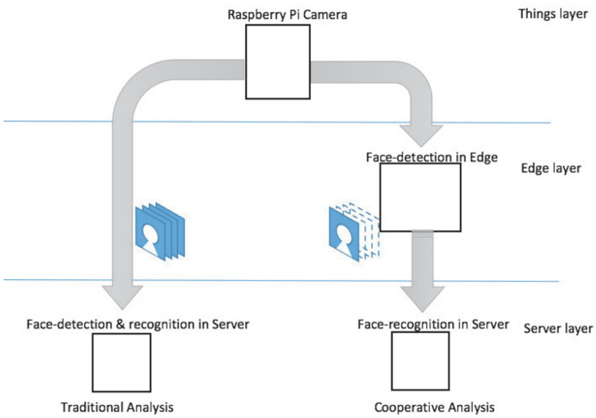


Fig. 4. Comparison of traditional analysis and cooperative analysis

Table 1 shows the total bandwidth of two analysis jobs. Traditional analysis is performed by a sequence of video frames in every 1 seconds. On the other hand, cooperative analysis is performed by a sequence of images in every 1 seconds.

Table 1. Network bandwidth of two analysis method

Method	Bandwidth	Result
Traditional analysis	~18 Mbps	A video frame in every 1 seconds
Cooperative analysis	~10Kbps	A image frame in every 1 seconds

5 Conclusions

In this paper, we suggested a new concept of IoT(Internet of Things) infrastructure, so-called Cloud of Things (CoT). In the Cloud of Things (CoT) environment, a device can be connected via IoT gateway(s) and a group of devices can be cooperated for their own benefit. Cooperative analysis in CoT environment is performed by the edge side and the sever side. It leads to filter out the inconsequential data while they are still at the edge and provide faster response. The face detection and recognition algorithm is the most common example of cooperative analysis jobs. There is growing evidence that edge processing will become more and more important in the Internet of Things. In future work, we will investigate more proper algorithm to get collective intelligence between edges.

Acknowledgments. This work was supported by the ICT R&D program of MSIP/IITP [R0126-15-1067, Development of Hierarchical Data Stream Analysis SW Technology for Improving the Realtime Reaction on a CoT (Cloud of Things) Environment].

References

1. Edge Computing, https://en.wikipedia.org/wiki/Edge_computing
2. Seven reasons edge computing is critical to IoT, http://www.thoughtthe soncloud.com/2015/07/7-reasons-edge-computing-is-critical-to-iot/
3. Dean, Jeffrey, and Sanjay Ghemawat. "MapReduce: simplified data processing on large clusters." Communications of the ACM 51.1 (2008): 107-113.
4. Zaharia, Matei, et al. "Spark: cluster computing with working sets. "Proceedings of the 2nd USENIX conference on hot topics in cloud computing. Vol. 10. 2010.
5. Raspberry Pi, https://www.raspberrypi.org
6. Bradski, Gary, and Adrian Kaehler. Learning OpenCV: Computer vision with the OpenCV library. " O'Reilly Media, Inc.", 2008.
7. Turk, Matthew, and Alex P. Pentland. "Face recognition using eigenfaces."Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91.

RANKING AND SELECTION CRITERIA OF MOBILE SERVICES USING ANALYTIC HIERARCHY PROCESS

Rohaizan Ramlan[*1], Shiau Wei Chan[1], Rohayanti Hassan[2], Rashidah Mokhtar[3] & Nur Amirah Sakeri[1]

[1]Universiti Tun Hussein Onn Malaysia, Batu Pahat, Johor, Malaysia
[2]Universiti Teknologi Malaysia, Johor Bahru, Johor, Malaysia
[3]Universiti Teknologi Mara, Segamat, Johor, Malaysia

1*rohaizan@uthm.edu.my, 1swchan@uthm.edu.my; 2rohayanti@utm.my; 3rashi271@johor.uitm.edu.my

Abstract. Mobile services in Malaysia have shown tremendous growth in subscribers since 2000. Based on previous studies, mobile services play a significant role in consumers' decisions and access to consumers' needs. Each mobile operator offers different packages and value-added services; therefore, there is huge competition among mobile service provider. The objective of this study is to propose mobile services that give the most value-added based on a range of criteria. An Analytical Hierarchy Process (AHP) approach is adopted to solve the multi-criteria decision making. A questionnaire was distributed to 101 mobile phone users between the ages of 20 to 29, who were subscribers, and had experience using mobile services. Four criteria associated with mobile services are payment mode, functionality, value-added and perceived value. The results indicated that the payment mode criterion ranked top, followed by functionality, value-added and perceived value criteria, respectively. The rankings for related sub-criteria are also shown in this study. Mobile service operators may use the results as guidelines to produce better services in order to remain competitive.

Keywords: Mobile services, Analytical Hierarchy Process (AHP), Consumers' preferences

1 Introduction

The mobile telecommunication industry is highly competitive both in the market and among customers. The mobile telecommunication industry has rapidly shifted from just delivery of voice services to highly advanced and sophisticated technologies, such as the mobile web, Global Positioning System (GPS), mobile email and location-based services [1]. The technology used by mobile services providers nowadays has

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_17

also become more advanced, and provides value-added to customers. All these point to the future trend of an expanding mobile telecommunication industry [2].

Regarding Malaysian mobile phone penetration, Forest Interactive [3] stated that, in 2015, there were 44.3 million active mobile subscribers; 18.94 per cent of this number had post-paid packages, while the majority of 81.06 per cent subscribed to pre-paid packages. These data increased tremendously from the previous year with a three million increase in active mobile subscribers. There was a 20 per cent increase in mobile penetration; the numbers of pre-paid user increased by 4.06 per cent and 3G/4G users increased by 17 per cent.

Currently, there are various mobile service providers in Malaysia. The top five mobile network operators in Malaysia in 2015 were Maxis, Celcom, Digi, U Mobile and Tune Talk [3]. Forest Interactive [3] also reported that 15 million social accounts are accessed via the mobile networks due to the range of different packages and value-added services offered by the mobile service provider, there is huge competition among them. Those mobile service providers that are able to maintain relationships and offer best value are more likely to gain customer loyalty, and assure further competitive advantage in the future [4]. Moreover, based on the results of his study, Rahman and Azhar [5] asserted that mobile service providers should focus on developing distinct brand personalities to market their brands rather than becoming involved in price wars.

The most relevant aspect to consider when designing a mobile service is customer value [2]. Users redefine the value of the way they use technology which will best fit their preferences and behaviour towards using the technology [6], meet lifestyle needs, and be of exceptional quality [7]. Nevertheless, if mobile network services can fulfill consumers' expectations, then service costs and the way their usage is charged for are, arguably, less of a concern [8]. Therefore, the aim of this study is to propose mobile services that provide most value-added based on a set of criteria. This study has important implications for mobile service providers on how to improve the packages to attract consumers to subscribe to their services, and thus retain the loyalty of consumers based on these criteria.

This paper is structured as follows. First, a review of criteria that contribute to the selection of mobile services is presented, and the basis for understanding the AHP is set out. Second, the methodology used in this study is presented and explained. The collected data are then described and the results of the analysis are presented and discussed. The final section concludes the paper.

2 LITERATURE REVIEW

2.1 Mobile Services Criteria

Several studies have been done on the selection of Mobile Service criteria in Malaysia [9]; [10]. Hassan et al. [9] studied four main criteria for mobile service provider selection using AHP. Meanwhile, Hwa et al. [10] studied purchase decisions made by generation Y relating to mobile service providers using a statistical approach. In addition to this study, a set of main criteria (with associated sub-criteria) are examined to obtain detailed results.

Previous studies identified several criteria that contribute in the selecting of mobile services. Nikou et al. [1] indicated that monthly internet, rewards, network coverage and customer care are the criteria used in mobile service selection. Meanwhile price, value-added services [10], monthly commitment, charges and rewards [9] are the main criteria in the selection of mobile service providers among Malaysian students. Other than that, connectivity, coverage, price, availability, consistent quality, packages, brand name, value-added services, company image, innovation, differentiation and advertisements are the selection attributes for generation Y [5]. Thus, main criteria used in this study are payment mode, functionality, value-added and perceived value. However, sub-criteria are adopted from Nikou et al. [1]. Each criterion is explained in detail below.

Payment Mode .

According to Nikou et al. [1], different payment methods influence significantly the user’s choice and preferences. Previous studies undertaken a few years ago in Malaysia reported that price is among the main criteria in the selection of mobile service providers by generation Y [10] and monthly commitment ranked number one among students [9]. Pricing is also the main contributory switching factor among mobile services in Ghana [11]. Therefore, payment is a key criterion for this study.

The payment packages differ between post-paid and pre-paid tariffs. In 2015, post-paid subscribers in Malaysia fell by 4.06 per cent as subscribers moved to pre-paid packages [3]. This suggests that the service fees between packages are important criteria for customers. This resulted in Jambulingam [12] suggesting that mobile service providers should reduce costs to students for them to fully benefit from a mobile learning (m-learning) environment. According to Nikou et al. [1], there are four payment methods for customers; these are usage-based charging, bundle pricing strategy, fixed price, and packet charging.

According to Nikou and colleagues [1], users are charged based on realised consumption, while bundle pricing provides a number of mobile service packages with different price categories. The advantages of bundle pricing include cost savings in production and transaction costs, and sorting consumers according to their valuation

[6]. The third sub-criterion is a fixed price. Fixed price means a fixed rate that consumers are charged every month. The packet charging is the last sub-criterion that refers to the services that are charged for in a packet-based method.

Functionality .

Functionality refers to the ability of the mobile services to allow a user to perform certain tasks [1] and to be the interface between mobile technology and the user of the mobile services [13]. Service accessibility [14], simplicity and usability have been found as crucial factors for the attribute of functionality [13], [1], [8]. The accessibility of a service is not limited to the cognitive aspects of adoption mobile services, but also includes availability and access to services in a physical sense [1].

Accessibility refers to the extent to which mobile services are accessible anytime or anywhere. Moreover, ubiquitous access to mobile services would encourage the usage of a particular service [14]. Simplicity of mobile services can be defined as when the technology used must be very simple to learn [1] and the access of the user to the mobile applications should require only minimum knowledge of technologies [13]. Meanwhile, usability occurs when a user can quickly understand how the mobile services work [8]. Asif and Krogstie [15] stated that users showed a tendency towards the service providers who provide useful services. The last is flexibility, which means the capability of the mobile services to adapt to personal profiles or requests [8]; [13].

Value-added .

Value-added services can be defined as the benefits of using the mobile services compared to other technologies [1]. The user must be convinced that by using a specific mobile application, they acquire a value that the other models do not provide [13]. Nevertheless, if a mobile service fulfills users' requirements, and improves the productivity, efficiency and effectiveness of the users, then the application can be considered as value adding [8]. The value-added sub-criteria are derived from Nikou et al. [1], which are mobility, content quality and entertainment.

Mobility is defined as the capabilities of an application to access the real-time information and communication while the user is on the move [1]. Jambulingam [12] found that the mobility has significant influences on user intention to employ m-learning due to freedom to access learning materials anywhere at any time. Quality of the content refers to the capability of an application to offer recent, accurate and timely content [8]. Chen and Hsieh [16] emphasised that content is among the most significant attributes in mobile services, while the entertainment value is the capacity of mobile services to fulfill customers' entertainment needs [1].

Perceived value (quality, cost and performance) .

Critical elements of the customer's satisfaction represent the attributes of service quality, cost and performance enhancement [1]. Price is the level of customer satisfaction with mobile services cost, while quality of service refers to how well a customer is being served. Perceived quality is the overall subjective judgement of quality relative to the expectation of quality; and perceived value represents the trade-off between costs and benefits and arises from both quality and price [17]. Meanwhile, perceived performance enhancement refers to the functionality of the service to meet users' needs and improves the users 'performance by using a particular mobile service [1].

Users are always ambiguous about the expectations and preferences of mobile operators due to the fact that human assessments of qualitative attributes are always subjective and imprecise [1] and they are reluctant to make an extra effort in complex situations, for instance choosing a service that best fits their needs [18].

2.2 Analytical Hierarchy Process (AHP)

The Analytical Hierarchy Process (AHP) is a process of developing numerical scores by which to rank each decision alternative to meet its objective. This approach was developed in the early 1970s by Saaty and the process has been used to assist numerous corporate and government decision makers. According to Hassan et al. [9], the AHP approach can formulate and analyse decisions by simplifying a multi-faceted, multi-criteria decision problem.

An AHP provides an accurate and efficient methodology with which to assess the importance of each of the needs in the hierarchy [19]. Apart from that, the AHP method provides a unique means of quantifying judgemental consistency and is simple to use and understand [20]. Moreover, the AHP is an appropriate approach because it combines all of the above-mentioned criteria into a model in a quantitative manner that can measure the importance of user requirements [21]. However, the AHP also has its weak points. The complexity on the implementation is quite inconvenient when more than one person is working on this method, and different opinions about the weight of each criterion can also complicate matters [20]. Taking all this into account, this study applied the AHP model to identify the most value-added package of mobile services.

3 METHODOLOGY

The Malaysian Communication & Multimedia Commission (MCMC) [22] reported that, as of the fourth quarter of 2015, the age range 20-24 accounted for the highest number of internet users in Malaysia followed by the 25-29 age group. Thus, in this paper, questionnaires were distributed to 101 mobile phone users between the ages of 20 to 29, who were subscribers, and had experience using mobile services in Malaysia.

The AHP includes three phases which are decomposition, comparative judgement and priority synthesis [21]. In the *decomposition* phase, the hierarchy of goal, criteria and sub-criteria are constructed as shown in Fig. 1. The next phase is *comparative judgement* at each level based on the customer's preference from the numerical ratings of pairwise comparison. The AHP questionnaire was designed using a scale of 1 to 9 suggested by Saaty [21] and shown in Table 1. Respondents were given a number of tables regarding the criteria of mobile service selection in Malaysia. Respondents were then given the numerical scale against which to rank their preferences on criteria influencing the selection of mobile services that offer them the most value-added. Scale 1 indicated the importance of both elements that have been compared, while scales 3, 5, 7 and 9 indicated the extent to which both elements are moderately important, strongly important, very strongly important and extremely important to each other. Meanwhile, scales 2, 4, 6, and 8 rank the intermediate value between two judgements. The data from the questionnaires were then transformed into comparative judgements and subjected to pairwise comparison to be analysed and to check the consistency of judgements. The last phase, *priority synthesis*, calculated a composite weight for each criterion and sub-criterion based on the preferences obtained from the previous phase. The technique in weight determination used is known as the eigenvalue method. The priority rankings for all criteria and sub-criteria are presented to form the overall results of the study.

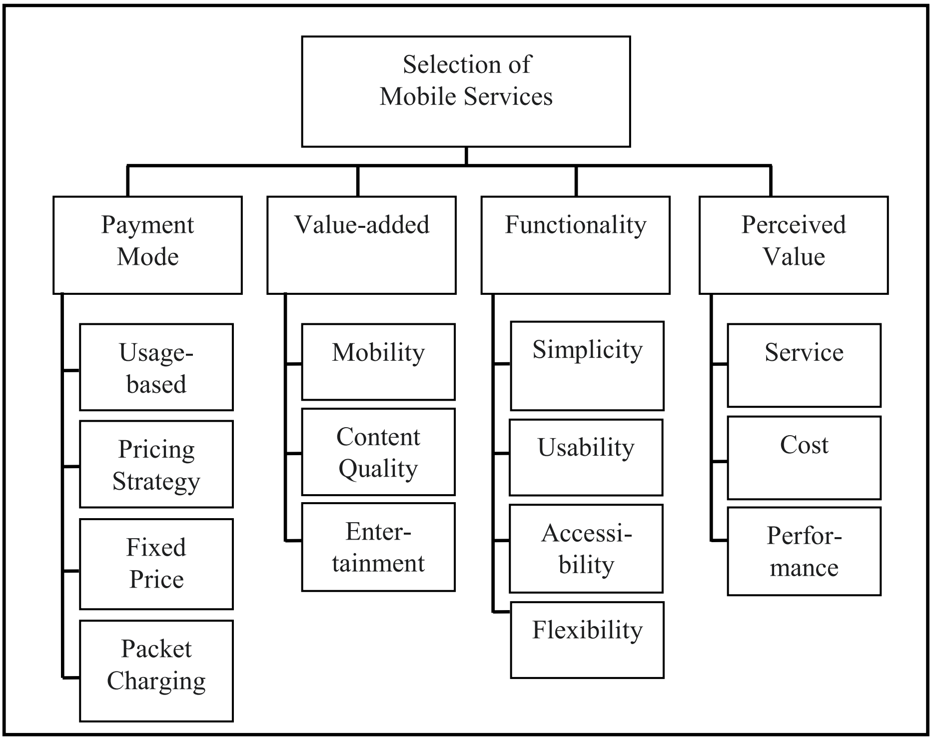


Fig. 1. AHP Framework of Mobile Services Criteria

Table 1. Questionnaire

Evaluation Criteria	Numerical Scale	Evaluation Criteria
Payment Mode	98765432123456789	Value-added
Payment Mode	98765432123456789	Functionality
Payment Mode	98765432123456789	Perceived value
Value-added	98765432123456789	Functionality
Value-added	98765432123456789	Perceived value
Functionality	98765432123456789	Perceived value

4 RESULTS AND DISCUSSION

The questionnaires were collected from 101 respondents consisting of 48 per cent male and 52 per cent female. The consistency checking was applied to all 101 questionnaires. From 101 respondents, only 52 questionnaires met the consistency ratio with less than 0.1; this shows that only 51 per cent are trustworthy and acceptable for further analysis. Meanwhile, others have been eliminated as they did not meet the consistency requirement.

Table 2 shows the average weight between all main criteria influencing the selection of mobile network operator. Results revealed that payment mode is the highest ranking (0.273), followed by functionality (0.267), value-added (0.257) and perceived values (0.242) with CR of 0.06. This shows that payment mode is the main criterion that customers apply when selecting a mobile service.

Table 2. Average weight between all main criteria for mobile services

Decision Criteria	**Priority**	**Rank**
Payment Mode	0.273	1
Functionality	0.267	2
Value-added	0.257	3
Perceived Value	0.242	4

Meanwhile, the results of the top five priority weight rankings on sub-criteria are shown in Table 3, with respondents prioritising usage-based charging with the highest weight, 0.345, followed by entertainment (0.335), mobility (0.334), perceived service quality (0.324), and content quality (0.278) with a consistency ratio CR of 0.07.

Table 3. Priority ranking and weight of sub-criteria

Priority ranking	Sub-criteria	Average weight	Category
1	Usage-based charging	0.345	Payment Mode
2	Entertainment	0.335	Value-added
3	Mobility	0.334	Value-added
4	Perceived service quality	0.324	Perceived values
5	Content quality	0.278	Functionality
6	Perceived performance	0.305	Perceived values
7	Perceived cost	0.296	Perceived values
8	Simplicity	0.279	Functionality
9	Pricing strategy	0.252	Payment Mode
10	Fixed price	0.248	Payment Mode
11	Accessibility	0.250	Functionality
12	Flexibility	0.247	Functionality
13	Usability	0.239	Functionality
14	Packet charging	0.234	Payment Mode

5 CONCLUSIONS

An AHP approach helps mobile service operators enhance their understanding of consumers' preference towards mobile service selection by ranking the criteria. The criterion payment mode is the most value-added criterion followed by functionality as the second highest, value-added as the third-ranked criterion, and perceived value as the least important criterion. This was supported by Hassan et al. [9], and Hwa et al. [10], who found monthly payment and price placed were ranked number one for mobile service selection. However, findings by Nikou et al. [1] and Nikou and Mezei [8] reported that functionality is the priority for mobile service selection in Finland, while Rahman and Azhar [5] reported that connectivity is the first ranking in mobile service selection in Pakistan. The results also show that sub-factor usage-based charging - which is the attribute of payment mode - has the highest priority weighting compared to the other sub-factors, followed by entertainment, mobility, perceived service quality, and content quality. The findings in Malaysia are dissimilar to others probably due to the ubiquitous nature of wireless connection in Malaysia as reported by Hwa et al. [10]. Mobile service providers may use these findings to improve the packages to attract consumers to subscribe to their service, and thereby sustain the loyalty of consumers.

The study cannot be generalised to the entire Malaysian population since a survey was conducted among respondents belonging to the age group 20-29. Future study should be conducted in a representative sample of the Malaysian population for a better understanding of mobile service usage. Moreover, some criteria and attributes need to improve based on packages offered by mobile service providers in Malaysia, which may help explain actual choices made by the consumers.

REFERENCES.

1. Nikou, S., Mezei, J., Bouwman, H., Liu, Y. : Factors influencing the adoption of mobile services consumers' preferences using analytic hierarchy process. (2011) 1-14.
2. Carlsson, C., Walden, P. : Further quests forvalue-added products and services in mobile commerce. In: ECIS 2002 proceedings Gdansk, Poland (2002) 715–724.
3. Forest Interactive : Malaysia Industry Statistic 2015, http://www.forest-interactive.com/malaysia-industry-statistics-2015/ (2015).
4. Roostika, R., Astuti, W. T. : The Analysis of Antecedents of Customer Loyalty in The Mobile Internet Market. In The Second International Research Symposium in Service Management, (2011)
5. Rahman, S., Azhar, S., : Xpressions of Generation Y : Perceptions of the mobile phone service industry in Pakistan. Asia Pacific Journal of Marketing and Logistics, 23(1), (2010) 91-107.
6. Bouwman, H., Carlsson, C., Molina Castillo, F. J., Walden, P. : Barriers and drivers in the adoption of current and future mobile services in Finland. Telematics and Informatics, 24(2), (2007)145–160.
7. Petrova, K., MacDonell, S. G. : Mobile services and applications: towards a balanced adoption model. The Fourth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies , (2010) 182-188.
8. Nikou, S., Mezei, J. : Evaluation of mobile services and substantial adoption factors with Analytic Hierarchy Process (AHP). Telecommunications Policy, 37(10), (2013) 915-929.
9. Hassan, N., Ahmad, N., Aminuddin, W. M. W : Selection of Mobile Network Operator Using Analytic Hierarchy Process (AHP). Advances in Natural & Applied Sciences, 7(1) (2013) 1-5.
10. Hwa, E. T. B., Lee, E. L. W., Cheng, R. : Generation y and choice of mobile service provider: A study on their purchasing decisions in choosing a mobile service provider. In 2nd International Conference on Business and economic Research Proceeding (2011) 1083-1101.
11. Bediako, D. E., Deh, I. Y., Asuamah, S. Y. : Demographic Effect In The Switching Decisions Of Mobile Service Users. International Journal of Innovative Research and Development, 2(8) (2013) 422-429.
12. Jambulingam, M. : Behavioural intention to adopt mobile technology among tertiary students. World applied sciences journal, 22(9), (2013) 1262-1271.
13. Buyukozkan, G. : Determining mobile the commerce user requirements using an analytic approach. Computer Standards and Interfaces, 31 (1), (2009)144 -152.
14. Zarmpou, T., Saprikis, V., Markos, A., Vlachopoulou, M. : Modeling users' acceptance of mobile services. Electronic Commerce Research, 12(2), (2012) 225-248.
15. Asif, M., Krogstie, J. : Role of Personalization in Mobile Services Adoption. In Proceedings of the International Conference on Multimedia and Human Computer Interaction. International ASET, (2013) (59-1)-(59-10).

16. Chen, P. T., Hsieh, H. P.. Personalized mobile advertising: Its key attributes, trends, and social impact. Technological Forecasting and Social Change, 79(3), (2012) 543-557.
17. Awang, Z. H., Jusoff, K. : The effects of corporate reputation on the competitiveness of Malaysian telecommunication service providers. International Journal of Business and Management, 4(5), (2009)173-178.
18. Tétard, F., & Collan, M. : Lazy user theory: A dynamic model to understand user selection of products and services. In System Sciences, 2009.HICSS'09 42nd Hawaii International Conference on. IEEE. (2009) 1-9.
19. Hepler, C., Mazur, G. : The Analytic Hierarchy Process: Methodologies and Application with Customers and Management at Blue CrossBlue Shield of Florida. In Transactions of the 19th US and 13th International Symposia on QFD. (2007) 137-149.
20. Tahriri, F., Osman, M. R., Ali, A., Yusuff, R. M. : A review of supplier selection methods in manufacturing industries. Suranaree Journal of Science and Technology, 15(3), (2008) 201-208.
21. Saaty, T. L. : Decision making with the analytic hierarchy process. International journal of services sciences, 1(1), (2008) 83-98.
22. Malaysian Communication & Multimedia Commission (MCMC) : Pocket book of Statistic Q4 2015. Retrieved on February 2016 http://www.skmm.gov.my/skmmgovmy/media/General/pdf/CM-Q2-2015-BI-(pdf).pdf (2015).

Data Quality of ERP systems in Mobile Environment

Porntida Kaewkamol

College of Arts, Media and Technology, Chiang Mai University
Huay Kaew Rd., Muang, Chiang Mai, Thailand

Porntida.k@cmu.ac.th

Abstract. Data can be regarded as an important asset of the organisation. Valid data at the right time is able to support appropriate decisions of employees from operational through executive level. As a result, there is an effective flow of the working processes across the organisation. Enterprise Resource Planning Systems (ERP Systems) is the core internal application that integrates and optimises business processes. It also establishes the main database to centralise data and information of the enterprise. Recently, the growth of wireless and mobile technology has led ERP system to become more accessible via mobile devices. However, input data from the users might be in different forms since there are various mobile platforms. Consequently, there might be an impact on data quality of the systems and might further provide negative effects for business. This paper, therefore, aims to identify and propose data quality dimensions for mobile ERP systems. This paper has the methodology that primarily focused on the study of literature reviews. Despite a limitation regarding the practical result, this paper contributes a guideline for data quality dimensions that should be considered to utilise the mobile ERP system for business.

1. Introduction

Enterprise Resource Planning Systems (ERP Systems) have been implemented in many industries, for instance, automotive, public sector, and education industry. Recently the global ERP system market increased 3.8% from 2012 to 2013 [11]. Similarly, SAP AG, one of key ERP system vendors, mentioned in the 2014 annual report that the revenue of software and software-related service rose 6% from 2013 to 2014 [17]. These figures represent the growth of ERP system implementation over the past few years.

Nevertheless, ERP system might not always be an appropriate business solution, as there are the reports of ERP implementation failure in some cases [3] [4] and [13]. According to the reason behind this, system quality is one of critical success factors (CSFs) which can lead ERP system implementation projects to become successful or failed [3], [4] and [13]. Besides, system quality can generally be implied to cover any quality-related issues of the system, for example, quality of business process execution, quality of the outputs or reports that are generated from the system and also quality of data and information within the ERP system. This paper, therefore, focuses mainly on the study of system quality, in particular data quality, to investigate and identify data quality dimensions of mobile enterprise resource planning system. This paper also proposes data quality dimensions that should be considered as a guideline for the successful ERP system deployment for the business.

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_18

2. Literature Reviews

2.1 ERP Systems and the Challenges of mobile ERP Systems

Enterprise Resource Planning (ERP) is an approach for streamlining the internal business activities of the enterprise [7]. It also integrates functional and cross-functional processes for effective internal management [16]. Additionally, it enables seamless flow of data and information across the organisation. ERP can also support decision making process with timely and reliable information [8].

An ERP system is an application which consists of various modules and each module is developed for supporting particular business functions of the company [12]. A general ERP system typically provides modules of Procurement, Production, Sales and Distribution, Finance and Accounting, and Human Resources [16]. Besides, the ERP systems typically establish the central database. Users from each department of the company consequently have the ability to access data of other related business functions in order to make the right decisions and complete their tasks [6]. Furthermore, the central database encourages shared information across the enterprise. ERP system is, therefore, the backbone of internal data, information, and knowledge [7].

With the advances in wireless and mobile technologies, vendors of ERP systems have recently enhanced the capability of the systems to be operated in mobile environment. However, there are still some challenges on how to effectively design and develop the mobile ERP systems. Information presentation is one of the key technical issues for mobile access to ERP systems. Mobile devices are compatible with various browsers which support several data formats. The systems should, hence, be able to support many markup languages, for example, WML, XHTML or HTML [1].

2.2 ERP Failure Factors and Data Quality Issue

ERP system implementation might not always be successful. Some researchers investigated the failure factors from 52 research papers and summarised them into nine failure factors of ERP system implementation [3]. These factors , from the most to least critical, were excessive customisation, a lack of change management, poor data quality, dilemma of internal integration, poor understanding of business implications and requirements, lack of top management support, limited training, misalignment of IT with business, and hidden costs [3].

Since poor data quality was one of top three most critical failure factors, data quality is directly related to the successful of ERP systems. Complete and accurate data at the operational level can lead to the appropriate business decisions and this tends to bring competitive advantages to the business [3]. In the other hands, data misfit between - organisational requirements and the ERP package in terms of data format can cause an increase in operating time and cost to fix the errors [3].

2.3 Data in ERP Systems

Data can be implied as the basis of information technology system. Every activity in business process requires data which is created in a previous step and, similarly, that piece of data is also used to generate further data to be used in the following steps. According to SAP ERP system, a wide range of data is stored in three groups: organisational data, master data and transactional data [16].

Organizational data is basically a structure of an organisation. The SAP ERP system allows a company to independently group related data with regards to the company's purpose, for example, grouping by sales organisation, by sales area, and by distribution channel. These groupings are called organisational structures [5]. The example of organisational data can be shown in Fig. 1:

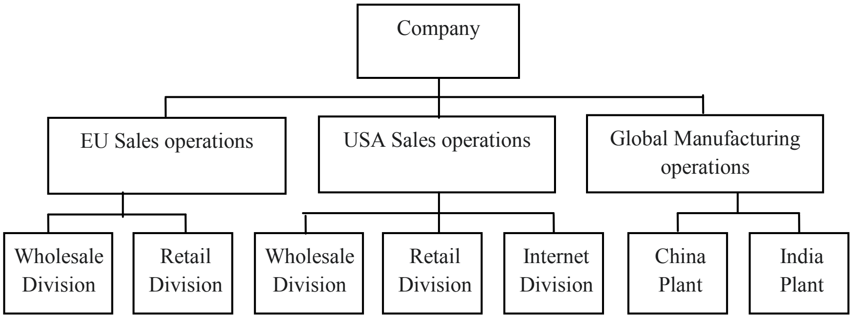


Fig. 1: Organisational data [Adapted from 16]

Master data is the main entities that a company is related to, for example, customers and suppliers [16]. It refers to data in the system that remains fairly unchanged, generally created once and used multiple times [2]. Additionally, it is stored in the central database, and is accessible by other modules of the ERP system [5]. Additionally, since master data is shared by various business units, it should be stored for only one version to reserve its integrity [16].

Transaction data refers to the day-to-day business activities of the organisation. It is also connected to the aforementioned data in ERP systems, organisational and master data, to define the context of that particular transaction. [2] [16]. In other words, each transaction record involves in a value, a timeframe, and other related data in the system [2]. Examples of transaction data can be shown in Table 1.

Table 1: Examples of Transaction data (Adapted from [2] which refers to Samaranayake, 2008)

Transaction Data	**Associated master data**
Planned order	Material master and bill of materials (BOM)
Production order	Material master, BOM, operations routing, cost centre
Purchase order	Material master and vendor master
Sales order	Material master and customer master
Maintenance order	Equipment, task list, work centre

2.4 Data Quality Dimension

According to the study of the quality in order fulfillment process, four variables were defined to measure the quality of information in each step, they included: timeliness, accuracy, convenience, and reliability [9].

In time means that data or information is stated when it is needed to make a decision corresponds to the state of the system. Accuracy refers to the data which is corrected or revised by its owner in order to provide appropriate decisions to execute working processes onwards. Convenience focuses on the feasibility to get an access when that data is needed, and reliability means the probability that the particular piece of data remain stable or unchanged [9].

Besides, the importance of quality is a key issue for both information and knowledge. A person who is in charge of creating that piece of data should also create it with quality. However, it also depends on the point of the time and the purpose to use that data. In other words, quality is related to and considered by its circumstance [14]. Additionally, other researchers mentioned ten of the information quality dimensions which were completeness, concision, reliability, timeliness, validity, accessibility, sufficiency (appropriate amount), credibility, relevance, and understandability [12].

Complete means that data should be comprehensive to plan and process the task. Concise refers to the situation that data can be used directly without prior revision, concise covers data format, data content and data structure. Reliable is the correctness of data [12].

Timely determines whether data is delivered in time. Valid means that data measures what it should measure. Accessible refers to data that can be accessed when it is needed [12].

Appropriate amount refers to the situation that data can be used appropriately without editing. Credible means data is accepted and believable. Relevant means data is proper for the task, and understandable means data is easy to use and edit [12].

Furthermore, Haug, Arlbjorn, and Pederson (2009) studied data quality dimensions from various papers. They mentioned the four dimensions from Ballou and Pazer (1985) including accuracy, timeliness, completeness, and consistency. Another data quality dimensions from Wand and Wang (1996) were mentioned, they consists of complete, unambiguous, meaningful, and correct. Interestingly, there are some quality dimensions that were differently defined by Shanks (1999) and they are syntactic, semantic, pragmatic, and social. However, the meaning or attributes of these are similar to prior data quality dimensions. Syntactic concerns correctness and consistence of the form or symbols for that data. Semantic means completeness and accuracy of the meaning of data at particular point in time. Pragmatic refers to usefulness and usability. Social means a shared understanding of the representation of that data [2].

Moreover, Zhu, Madnick, Lee, and Wang (2014) mentioned dimensions of data quality that have been categorised into four groups: intrinsic, contextual, representational, and accessibility [10]

Intrinsic data quality can be accuracy, objectivity, and any other data quality attributes which reflects the quality of that piece of data. Contextual highlights that a context of a particular task should also be considered for data quality. Examples of attributes in the contextual group are timeliness and completeness. Attributes in representational group, such as interpretability, and attributes in accessibility group, such as access security, involve in the role of system and tools which provide user interaction with data [10]

3. Methodology

Since this paper focuses principally on concepts and theories of data quality for enterprise resource planning systems, the methodology is primarily focused on the literature reviews of relevant papers. Main topics to study include the concepts of enterprise resource planning systems and their failure factors, data in ERP system, and data quality dimension. These studies were done to analyse and identify data quality dimensions that should be considered for a successful ERP implementation project.

4. Framework and Discussion

With regards to literature reviews, data quality dimensions for mobile enterprise resource planning can be discussed and identified in Fig. 2:

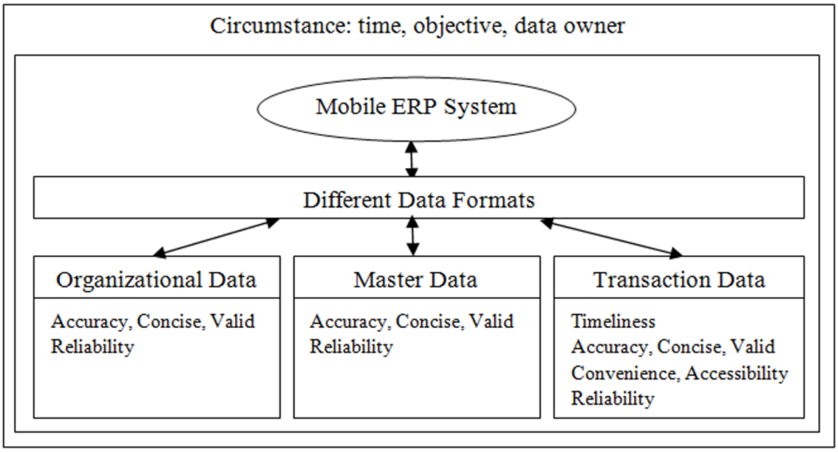


Fig. 2: Framework of data quality dimension in mobile ERP systems

According to Fig. 2, data quality dimensions in mobile ERP systems mainly consist of four aspects:

- Accuracy, Concise, and Valid
- Reliability
- Timeliness
- Convenience, Accessibility

These dimensions are related in all three data categories of ERP systems which are organisational data, master data, and transaction data. However, they are related and should be considered with a different degree. Timeliness and convenience or accessibility need to be carefully considered for transaction data, since this type of data involves in the day-to-day business activities and can be changed often from mobile devices. It therefore needs to be conveniently easy-accessed at the right time so an employee, who is in charge of each business process, is able to promptly use data when it is needed to operate and make proper decision on their working processes and likely to maintain the seamlessly business flows.

Besides, data quality dimension in terms of accuracy, concise, valid and reliability should be considered for all three data categories in ERP systems. This is due to the fact that data in the system needs to be accurate, concise, valid, and reliable in every step of business processes although they are inserted into the system via different data formats from various mobile devices. If there is no correctness in any kinds of data categories, employees and also executives might have to retrieve and make business decisions based on wrong sets of data which likely to provide negative effects for business in the long run.

However, this framework is not divided data quality dimension into each business functions or business processes as it needs further investigation under real business context. Since data quality of each process also depends on the time that the data is utilised, its objective to be used, and its owner or the person who needs that data or who is responsible for that particular process, data quality also needs to be considered by its circumstance in the holistic view.

5. Conclusion

Data is a value asset for enterprises since it is an initial input to be processed as the required output. In business, almost all working process basically needs data for the decisions on what task should be done and how it should be done. Furthermore, data can also be used to project and shape a future direction of an enterprise.

Data in mobile enterprise resource planning systems can be categorised into three main groups: organizational data, master data, and transaction data [16]. Each group of data is relevant to data quality dimension in different ways. Organizational data and master data are rarely changed so they need to be accurate, concise, valid, and reliable. Transaction data is generally changed frequently; it is therefore needs to be convenient to access at the right time or at the time it is needed.

However, there are other factors that should be considered together with these data quality dimensions. The context when using a particular piece of data also plays an important role for data quality. Circumstance in terms of the time when data is used, the objective that data is used, and the owner who uses that data, should not be overlooked since it also affects the quality of data.

References

1. Dabkowski, A. and Jankowska, A. M.: Comprehensive Framework for Mobile ERP System, Proceedings of the 14th International Workshop on Database and Expert Systems Applications, IEEE Computer Society, (2003) 890-894

2. Haug, A., Arlbjorn, J. S. and Pedersen, A.: A classification model of ERP system data quality, Industrial Management and Data Systems, Vol. 109 (8). Emerald, (2009) 1053-1068

3. Momob, A., Roy, R. and Shehab, E.: Challenges in enterprise resource planning implementation: state-of-the-art, Business Process Management Journal, Vol. 16 (4). Emerald, (2010) 537-565

4. Venugopal, C. and Rao, K. S.: Learning from a failed ERP implementation: a case study research, International Journal of Managing Projects in Business, Vol. 4 (4). Emerald, (2011), 596-615

5. Monk, E. F. and Wagner, B. J.: Concepts in Enterprise Resource Planning: Fourth Edition, Cengage Learning, (2013)

6. Jacobs, F. R. and Whybark, D. C.: Why ERP?: A Primer on SAP Implementation, McGraw-Hill Higher Education, (2000)

7. Norris, G., Hurley, J. R., Hartley, K. M., Dunleavy, J. R., Balls, J. D. and Dunleavy, J.: E-Business and ERP: Transforming the Enterprise, Canada, John Wiley & Sons Inc., (2000)

8. Saloner, G. and Spence, A. M.: Creating and Capturing value: Perspectives and Cases on Electronic Commerce, New York, John Wiley & Sons Inc., (2002)

9. Forslund, H.: Measuring information quality in the order fulfillment process, International Journal of Quality and Reliability Management, Vol, 24 (5). Emerald, (2007), 515 – 524

10. Zhu, H., Madnick, S., Lee, Y. W. and Wang, R. Y.: Data and Information Quality Research: Its Evolution and Future, Computing Handbook: Information Systems and Information Technology, 3rd Edition, Taylor and Francis Group, (2014), 16-1 – 16-20

11. Columbus, L.: Gartner's ERP Market Share Update Shows The Future Of Cloud ERP Is Now, Forbes, (2014), http://www.forbes.com/sites/ louiscolumbus/2014/05/12/gartners-erp-market-share-update-shows-the-future-of-cloud-erp-is-now/ Accessed from January 2016

12. Gustavsson, M. and Wanstrom, C.: Assessing information quality in manufacturing planning and control processes, International Journal of Quality and Reliability Management, Vol, 26 (4). Emerald, (2009), 325 – 340

13. Al-Mashari M. and Al-Mudimigh A.: ERP implementation: lessons from a case study, Information Technology & People, Vol. 16 (1), Emerald, (2003) 21-33

14. Lillrank, P.: The quality of information, International Journal of Quality and Reliability Management, Vol. 20 (6), Emerald, (2003) 691 -703

15. Magal S. R and Word, J.: Integrated Business Processes with ERP Systems.USA, John Wiley & Sons Inc. (2012)

16. SAP AG,: 2014 Annual Report, SAP website, (2014) http://www.sap.com/bin/sapcom/en_us/downloadasset.2015-03-mar-20-01.sap-2014-annual-report-pdf.html / Accessed from January 2016

m-PACS gateway platform for mobile-based diagnostic medical image service

Yongsoo Kim[1], Woon Cho[2], Daewon Chung[3], Jihoon Cho[4], Joonhyeon Jeon[5,*]

Department of Electronics and Electrical Engineering, Dongguk University-Seoul, 30, Pildong-ro 1gil, Jung-gu, Seoul 100-715, Korea
[1]isamokan@dgu.edu; [2]chowoon177@dgu.edu; [3]jung1362@dgu.edu; [4]hoon@dgu.edu; [5]memory@dgu.edu

Abstract. This paper describes an efficient and effective gateway platform, which provides a mobile-based picture archiving communication system (*m*-PACS) for diagnostic medical image service available via android-based mobile phone. The proposed *m*-PACS platform plays a role as a gateway to provide a standard way of communicating between mobile phone and hospital's PACS, and also allows trustworthy and resilience service of mobile applications on demand. This *m*-PACS platform has a three-layered structure which consists of management layer, functional layer and network layer. Especially, the functional layer includes the following features for mobile-based diagnostic image processing: header syntax conversion for single standard format, resolution transformation on JPEG/JPEG2000, text and avatar search, and etc. Finally, a mobile service system is developed for the use of the *m*-PACS platform in the android framework, and this is shown to provide a useful and efficient solution for a mobile-based diagnostic image service.

Keywords: PACS, mobile PASC, healthcare system, diagnostic image service.

1 Introduction

In today, the development of IT-based diagnostic technologies is greatly accelerated and advanced diagnostic image equipments (CT, MRI, PET, etc.) are provided to accurately diagnose patients who have various diseases. Hospitals, through the use of a Diagnostic Medical Image service System (DMIS) based on these digital diagnostic devices, can provide better health care to patients [1-2]. In general, DMIS consists of Hospital Information System (HIS), Order Communication System (OCS), Electronic Medical Record (EMR) and Picture Archiving Communication System (PACS). In particular, PACS plays a very important role for the storage and transmission of diagnosis medical images made by various diagnostic medical devices. Thus, PACS is widely used in medical fields and absolutely needed for the digital health information services [2]. Nevertheless, in the case of the existing PACS, there need constraints of space (i.e., consultation room) and time (i.e, appointment) for taking doctor's diagno-

* Corresponding author.

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_19

sis of medical image. Furthermore, since diagnostic medical images in PACS are made in various formats dependent on medical devices, their own methods to display and handle these diagnostic images are also additionally required [3]. For this reason, new diagnostic medical services based on advanced network technology are also required more and more. This paper provides a platform of mobile-based PACS (*m*-PACS), available via Android-based mobile phone, for diagnostic medical image service.

2 Diagnostic medical image service system based on *m*-PACS

Fig. 1 shows a proposed *m*-PACS platform for mobile-based diagnostic medical imaging services and the related functional diagram is indicated in Fig. 2. The proposed *m*-PACS platform plays a role as a middleware to provide a standard way of communicating between mobile phone(on wireless network) and hospital's PACS(on wired network). It provides a three-layered structure which consists of management layer, functional layer and network layer, and also allows mobile applications to convert / compress / store / manage / transform diagnostic medical images. The proposed *m*-PACS system is summarized based on Fig. 1 and Fig. 2 as follows.

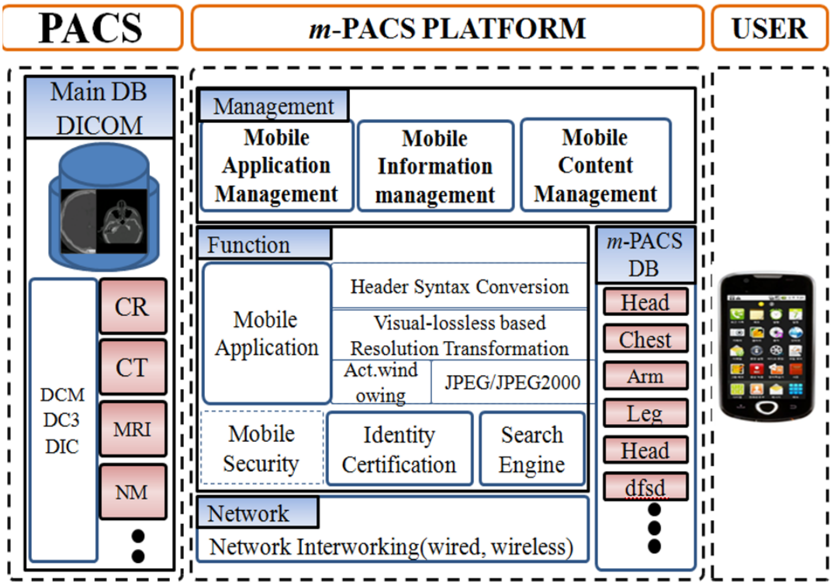


Fig. 1. System diagram for mobile-based diagnostic medical image service

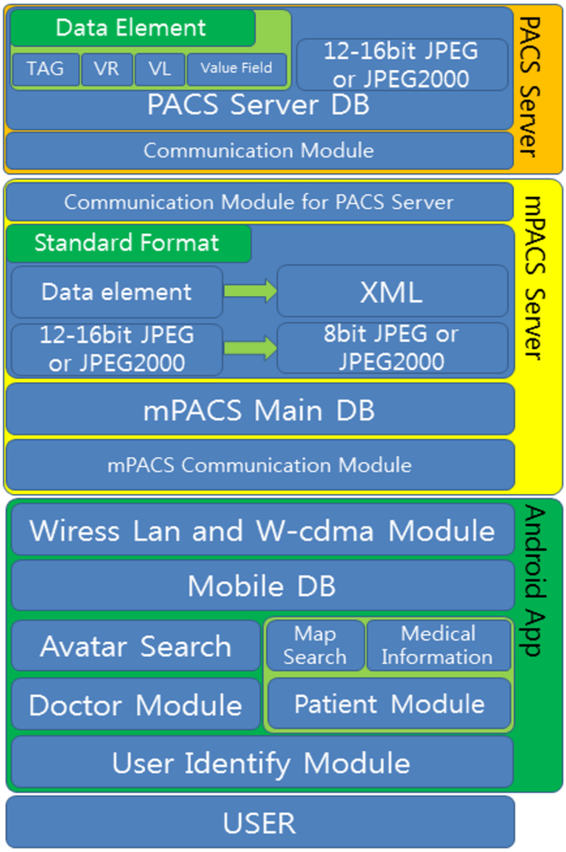


Fig. 2. Functional structure of *m*-PACS platform

2.1 Hospital's PACS operation for *m*-PACS

Hospital's PACS manages and stores patient's diagnostic medical images generated from various diagnostic medical equipments and it also delivers data streams in real-time, upon users request. Operation between PACS and *m*-PACS is as follows:

① Authentication module: PACS provides a login function for *m*-PACS in order to access database of hospital PACS server when the request occurs from *m*-PACS.

② Module of providing patient's diagnostic data to *m*-PACS: PACS searches the diagnostic medical data & information of patient enrolled in database and transmits the corresponding data & information to *m*-PACS.

2.2 *m*-PACS operation as a gateway

The *m*-PACS system provides a layered structure platform for communicating between mobile phone and hospital's PACS. It also performs various functional modules for mobile quality of service, which manage / convert / compress / store / transform diagnostic medical images. Operation of *m*-PACS as a middleware between mobile phone and hospital's PACS is as follows:

① Authentication module: Authentication module with wireless network interworking. Then, user is allowed to access the *m*-PACS.

② Network interworking modules: Communication between user and *m*-PACS is requested when the user required image is not registered in *m*-PACS.

③ Single standard conversion module: It is a function that converts 12-bit DICOM image format [3, 4] into 8-bit format with *m*-PACS standard header in order to display the PACS's diagnostic medical image in mobile devices.

④ Classification and storage module: It provides the ability to save to *m*-PACS, classified according to the imaging method by patient name, ID, attending physician, hospital personnel and imaging apparatus.

⑤ Avatar search engine module: It is a GUI to provide a search conveniently and rapidly of the patient's diagnosis.

2.3 User wireless device: smartphone

User is able to features of the integrated platform through the installation of android application. Functional operation on mobile phone is as follows:

① Login module: Giving a grant access to the platform for user's authentication.

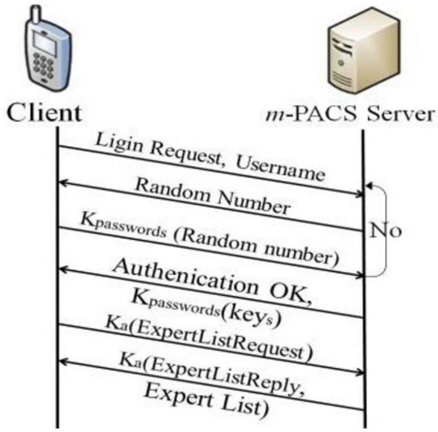
② Diagnostic image processing module for Doctors: Providing functions that consist of image download, avatar search, and image processing (i.e., zooming, rotating, retrieving and etc).

③ Patient's hospital service module: Providing a function of various health services- emergency service and search hospitals/pharmacies.

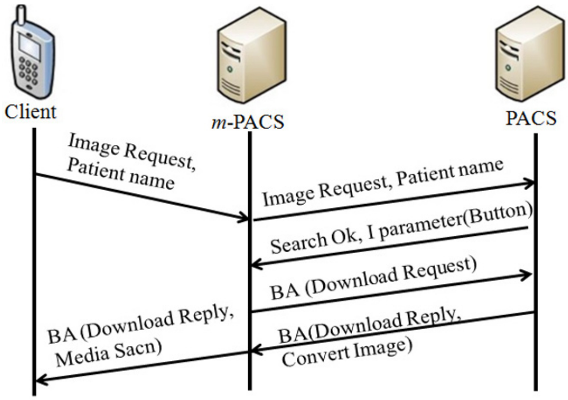
3 Functions of *m*-PACS platform

3.1 Network interworking

When making mobile-based diagnostic medical image services, users are firstly connected to the *m*-PACS platform by using the mobile device. Then, it proceeds with login using personal information in order to validate a user's access. After permission is granted, the user can requires a patient's diagnostic image through using the avatar search or the regular search in the *m*-PACS. The *m*-PACS, if not storing in database, requests hospital's PACS for the patient's diagnostic image. The PACS searches the desired diagnostic image in own database (DICOM original image) and transmits to the *m*-PACS. Finally, the *m*-PACS transmits to the user and stores in own database after recording the header information for the avatar search and performing the processing step to read in the user device. Fig. 3 shows the login and request flow charts between user, *m*-PACS and PACS.



(a)

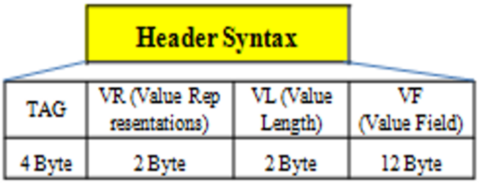


(b)

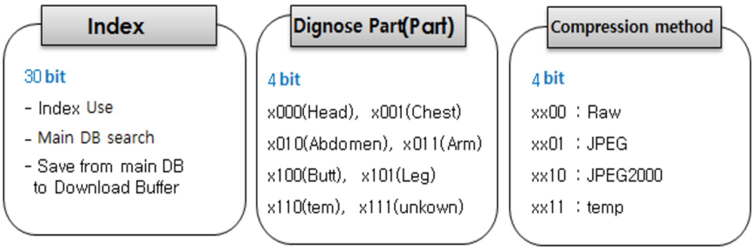
Fig. 3. (a) Login and (b) Request flow charts

3.2 8-bit conversion, Header syntax, and avatar search

Diagnostic medical images stored in the hospital PACS database (DICOM original image) have 12 or 16 bit resolutions and also include independent headers dependent on manufacturers and devices. Therefore, it is impossible to represent these images on the user device using 8-bit displayer [4]. In addition, because of the complex header information retrieval, a lot of problems occur in processing diagnostic medical images (restore, display and etc.). In order to solve these problems, the *m*-PACS platform provides the 8-bit representation method using a human visual function [5, 6] for mobile quality of service and allows a single-header-syntax standardization for replacing DICOM header of diagnostic medical images (Fig.4). Fig. 5 shows the 8-bit conversion of the 12 bit diagnostic medical images. After 8-bit conversion and re-compression (:JPEG2000 [5]), each of imaging region in the *m*-PACS database is stored by indexing of the standard single header.

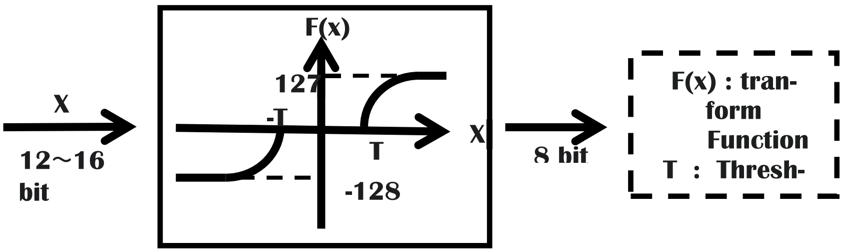


(a)

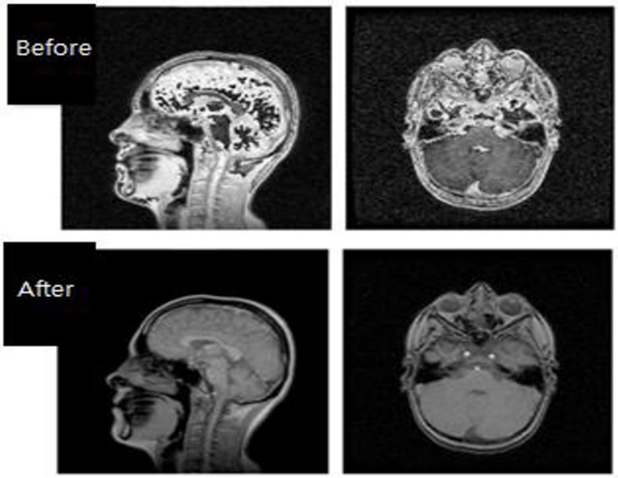


(b)

Fig. 4. (a) DICOM header and (b) single header syntax



(a)



(b)

Fig. 5. (a) 8-bit conversion method and (b) results

The *m*-PACS platform also provides the avatar and the text searches via the mobile device. Especially, the avatar search (Fig. 6 and Fig. 7) is a method that considers of user conveniences and it allows to the rapid search of each diagnostic medical image parts in the device.

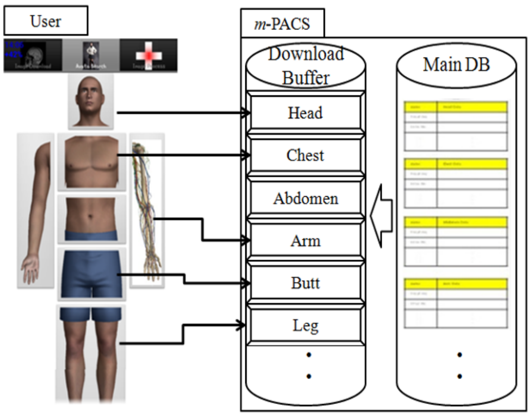


Fig. 6. User GUI and *m*-PACS DB indexing for avatar search

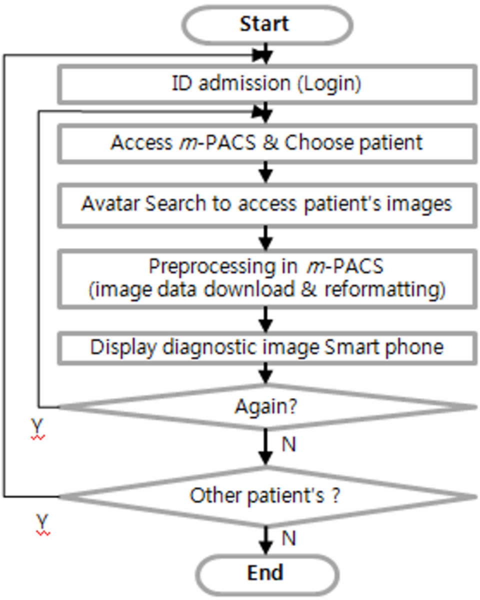


Fig. 7. Flow chart for avatar searching on mobile device

4 Implementation of *m*-PACS platform

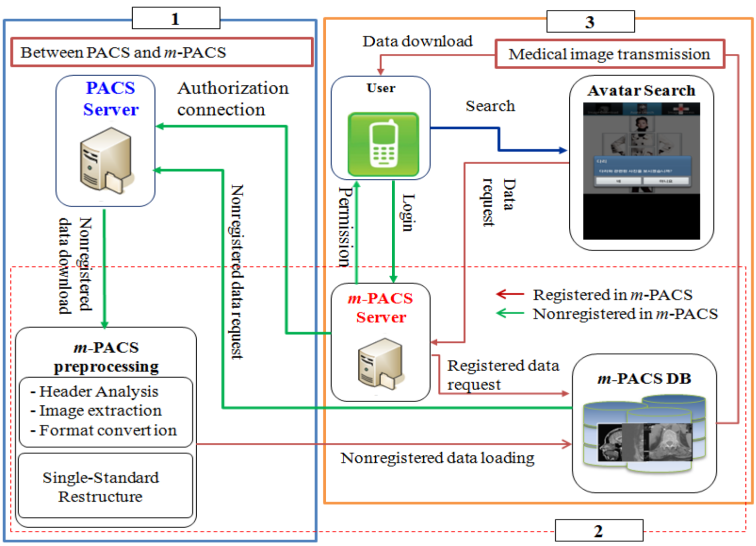
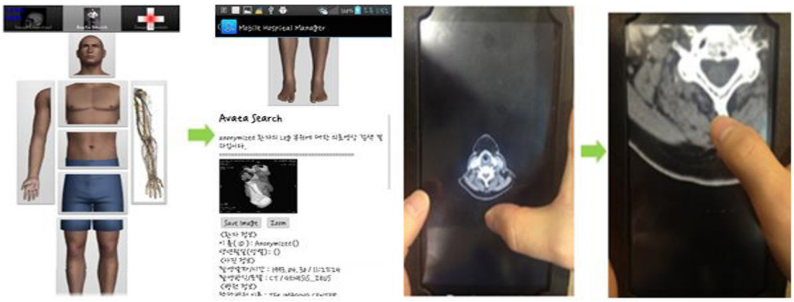


Fig. 8. Overall service flows

To demonstrate the proposed *m*-PACS platform shown in Fig. 1 and Fig. 2, a diagnostic medical image service system is developed and the overall signal & data flow diagram is indicated in Fig.8. The Galaxy Note 4 by Samsung [7] is also used as an android-based mobile device for mobile-based diagnostic medical imaging service. A driven running example is indicated in Fig. 9, which shows the login screen, avatar searching screen and zooming screen, respectively. Diagnostic images could be also searched in combination ID, physician, hospital personnel, imaging device, and the conditions of the imaging method. If patient's diagnostic images requested via mobile device exist in the *m*-PACS database, user can select the enable diagnostic part to display them on the screen and save. Then, it also provides an image processing functions such as enlarge/ reduce the diagnostic medical image.

(a)



(b)

Fig. 9. (a) Login screen and (b) Avatar searching and zooming screens

5 Conclusion

In this paper, we have presented an integrated *m*-PACS platform to be able to easily provide diagnostic medical image service via mobile devices, without changing the PACS installed in the hospital. The proposed *m*-PACS platform provides a new insight into the design of mobile-based medical image platforms.

References

1. Xinhua, C., Huang, H.K.: Current status and future advances of digital radiography and PACS. Engineering in Medicine and Biology IEEE. 19 (2000) 80-88
2. Silva, L.A., Pinho, R., Ribeiro, L.S., Costa, C., Oliveira, J.L.: A Centralized Platform for Geo-Distributed PACS Management. Int. J. Digit. Imaging. 27 (2014) 165-73
3. Treichel, T., Gessat, M., Prietzel, T., Burgert, O.: DICOM for Implantations-Overview and Application. Int. J. Digit. Imaging. 25 (2011) 352-358
4. Archie, K.A., Marcis, D.S.: DicomBrowse: Software for Viewing and Modifying DICOM Metadata. Digit. Imaging. 25 (2012) 635-645
5. Bong, J., Yang, G., Jeon, H.: Diagnostic Medical Image Compression Method Using Visually Lossless Threshold on JPEG2000. KICIS. 34 (2009) 671-680
6. Kanoun, O., Salim, M., Mezghani, S.: Medical Images Adaptive Compression With Statistical Model for Transmission and Archiving; Application to MRI modality. ICTTA. 1 (2006) 11457-1462
7. WIKIPEDIA, http://en.wikipedia.org/wiki/Samsung_Galaxy_S_II.

Part IV
Mobile Software

Mobile Learning Development for Supporting Computer Programming Skills

Krittawaya Thongkoo

Chiang Mai University, Chiang Mai, Thailand
krittawaya.t@cmu.ac.th

Abstract. This is so difficult for the first year students to start to write computer program in the subject of writing basic computer program due to most of students don't have a good basic of sciences and mathematics including lack of interesting in the subject processes related to writing Programming Assistance Instrument, which is web application that can help to motivate students to learn about writing fundamental program efficiency. This will help to increase learning skills and related to idea for students to write computer program very well. This research has proposed that the first year students, which are learning in Modern Management and Information Technology (MMIT) Course of College of Arts, Media and Technology should receive the contribution in writing computer program with this Programming Assistance Instrument that the main objective of this research is the contribution of knowledge include helping MMIT students to understand the idea of writing basic computer program in order to have motivation to learn and write further other computer programs. The expected result that should be received is the students who have use Programming Assistance Instrument will get knowledge, understanding and motivation in writing computer program better than students who studied with the old method.

Keywords: collaborative learning, inquiry-based learning, mobile learning

1 Introduction

Nowadays, the demand on industry's labor are increasingly associated with the develop web program, developing program on mobile and on cloud computing. So, this semester will try to generate students to match with the need of the market in order to increase the proportion of jobs for the students by increasing subjects, which is related to those technologies in elective subject for the students who is interested to learn more.

However, due to current students who are interested in the program are a minority because the students don't have enough knowledge of the basic of sciences and mathematics to help analyze or interpret the commands, which are related to writing of program. For that reasons, it is very important to encourage learning skills and knowledge which is needed for students by starting from introductory computer programming subject in order to be a basic writing program for the first year students.

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_20

This research will focus on the knowledge and skills of writing program for the first year students especially for MMIT students, College of Arts, Media and Technology. This will help to assign the criteria for the skills selection and the knowledge which is related to writing program that should have in Programming Assistance Instrument.

2 Literature Review

2.1 Collaborative Learning

Collaborative Learning is a situation in which student and faculty learn or attempt to work something together in order to create knowledge as science's teaching (Pedagogy) that has the center to create the mutually meaning and is the process that is full with knowledge and more enhancements [1]. Moreover, Buffy [2] has said that the knowledge and the truth are found in everywhere and just wait for discovering by trying of human with the social method and relying on the support collaborative learning. The knowledge is the thing that people created by speaking and agreement.

Collaborative Learning is the learning, which can exchange idea in the small group. That is not an increasing of the interesting of participants but this will support analysis thinking, exchanging of knowledge between learners. This will be the opportunity to foster by debates together, responsibility for their own learning [3].

In conclusion, Collaborative Learning is learning how to use a form of social learning that there was talk among the group of people in order to create knowledge with ourselves from the relationship between each other by supporting the learning that is happen from the cooperation, dependence and helping each other as most as possible.

2.2 Inquiry-based Learning

The searching of knowledge is the form of learning and teaching according to the creating knowledge's theory (constructivism). It said that it is the process that students must search, find, survey, inspection and search with other methods until the students understand and percept with the meaning of knowledge. This will be created as knowledge of the students and keep as the information in the brain for a long time. This can be brought to use in any confront situations.

The Searching of knowledge is asking the question, which can find and communicate the answer. This is related to the asking of question, surveying of the information, analysis of the result conclusion, invention, exchanging opinions and communication of explanation [4][5].

The conclusion of Inquiry-Based Learning is the process that the scientist has studied the explanation of natural phenomenon, which is laid down on the basic of the evidence or several reasons and another meaning is the process that the students use to search the answer as systematically in order to explain the various situations, which prefer to study.

2.3 Mobile Learning (m-Learning)

The principle of the management of studying and teaching with m-Learning is the students are able to take lessons from the field and placed on the mobile browsing at anytime and anywhere. They can send and receive the information when needed and when it has a signal from telecommunication network. Moreover, it must be worked in two ways, in order to change the lesson, send homework or analysis scores from exercises as well.

The students and teachers will use the important instrument such as mobile instrument, which is convenient for shift and can connect with computer network without signal cable in real-time such as Notebook Computer, Portable Computer, Tablet, PC and Cell Phones for learning and teaching activities.

The type of learning process m-Learning consists of five steps:

First Step:	Students and instruments are ready.
Second Step:	Connecting to network and find the necessary content for study.
Third Step:	If found the content, go to Fourth Step. If not, return to Second Step.
Fourth Step:	Operate learning without using the network.
Fifth Step:	Got the learning result according to the objective.

<u>The relationship of the learning form m-Learning and management of study</u>

The transition of learning condition from holding teachers as the center will be changed to interact directly with the students. This will encourage people to communicate with friends and teachers more and more. This can receive the information, which has no specification name, in order that the students who are not confident, more assertive foster learning real.

3 Methodology

3.1 Data Collection

Initially, the questionnaire will be distributed to the first year students in Computer Programming subject for 100 persons. All students are studying at Modern Management and Information Technology Course, Chiang Mai University. This questionnaire will ask about personal information, basic knowledge including basic writing program skills. The results of the questionnaire will be analyzed to develop a system that responds directly to the needs of the students.

3.2 System Analysis & Design

The necessary for a better understanding of basic computer programming for the first year students is basic programming knowledge (syntax, principles and concepts) and code comprehension including code generation. From the above information, Programming Assistance Instrument was developed as shown in Fig.2 by using the concept of collaborative learning as shown in Fig.3 and inquiry-based learning, in order to stimulate student learning with an emphasis on group work. The three main stakeholders with the system design for this system are combined; Admin, Teachers and Students as shown in Fig.1.

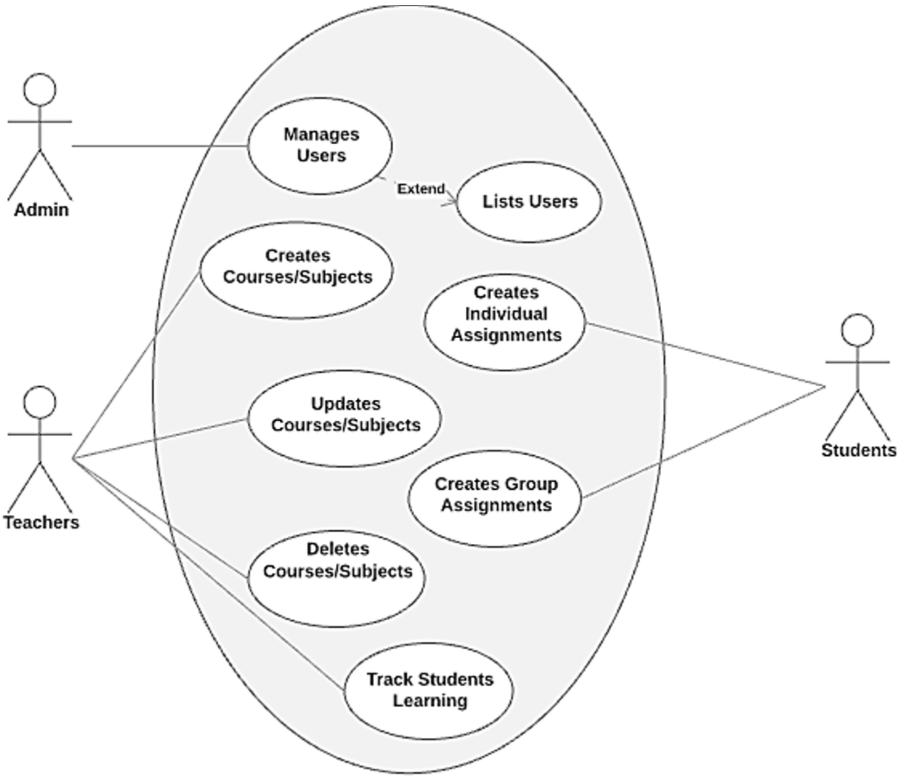


Fig.1 Use Case Diagram of m-Learning System

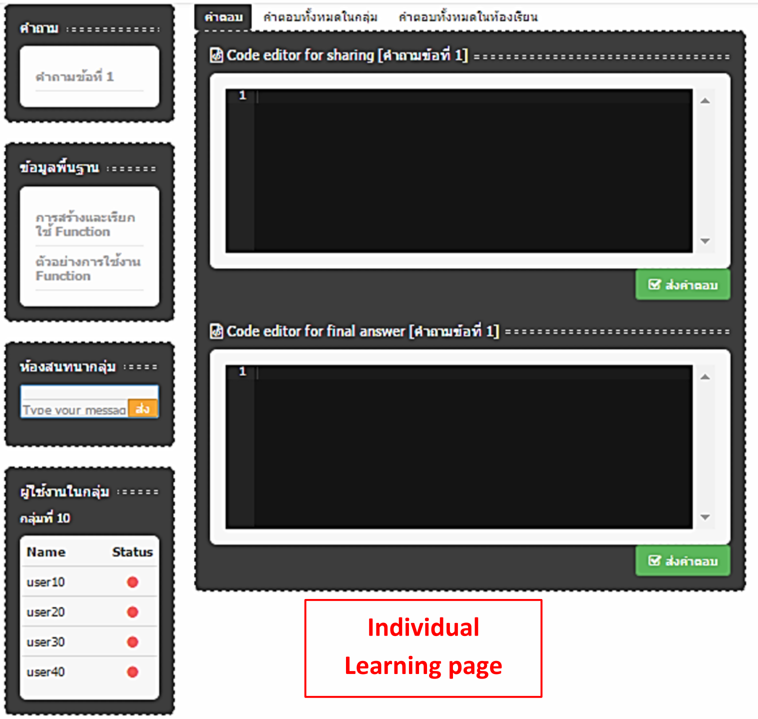


Fig.2 m-Learning Interface

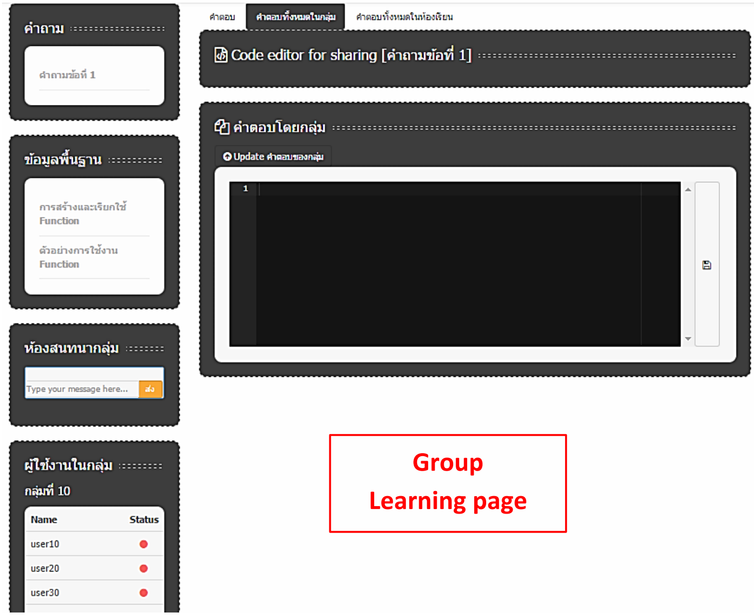


Fig.3 Example of collaborative learning Interface

4 Conclusion

This research was presented the system approach in educational technology for the development of knowledge and skills in basic of computer programming for the first year students. The proposed tool is intended to help facilitate the learning process and its application in practice of computer programming. This application is designed for the user who has low or minimal skills in writing program and it has enough features to enable students to work effectively as a tool for self-learning. For these reasons, this application is particularly useful for the first year students of the course MMIT.

References

1. Matthews, R.S., Collaborative learning: Creating Knowledge with students. In R.J.Menges,M. Weimer, & Associates (Eds.), Teaching on solid ground : Using scholarship to improve practice (1996)
2. Buffee, K.A., Collaborative Learning: Higher education, interdependence, and the authority of knowledge. Baltimore, MD:Johns Hopkins University Press (1993)
3. Johnson, R.T. & Johnson, D.W., Action research: Cooperative learning in the science classroom. Science and Children (1986)
4. Wu and Hsieh,H.-K. Wu, C.E. Hsieh, Developing sixth grader's inquiry skills to construct explanations in inquiry-based learning environments. International Journal of Science Education (2006)
5. Wu and Hsieh,H.-K. Wu, C.E. Hsieh, The effects of inquiry-based learning on elementary students' conceptual understanding of matter, scientific process skills and science attitudes. Procedia - Social and Behavioral Sciences (2010)

A GPU Powered Mobile AR Navigation System

Mengshen Zhao and Byunghyun Jang
*Department of Computer and Infor*mation Science
The University of Mississippi
University MS, USA 38677
Email: `mzhao@go.olemiss.edu` *and* `bjang@cs.olemiss.edu`

Abstract. This paper presents a real-time Augmented Reality Navigation system (ARNavi) on Android mobile phone that exploits the parallel computing power of mobile GPUs. Unlike conventional navigation systems, our proposed ARNavi augments navigation information onto live video streaming from device camera in real-time. The contributions of this paper are two-fold. First, we propose a fast and accurate lane detection and mapping algorithms. Second, we demonstrate that real-time augmented reality navigation can be achieved by taking advantage of CPU-GPU heterogeneous computing technology on a mobile processor. We achieve up to 18 FPS for 640×360 resolution camera streaming. It is more than 2.6× speedup over CPU only execution.

Keywords: GPGPU; OpenCL; OpenCV; mobile GPU; augmented reality; navigation application;

1 INTRODUCTION

GPS navigation became prevalent on mobile devices such as smartphones. Conventional GPS systems overlay direction and other useful information on a pre-built map and user perceptually relates the map with real scene. In this paper, we propose a new navigation system called real-time Augmented Reality Navigation system (ARNavi) on mobile Android devices. Unlike conventional GPS navigation systems, our proposed ARNavi overlays navigation information directly onto live streaming real scenes coming from device camera. This AR assisted navigation improves user experience as users

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_21

do not have to relate the conventional map with real scene. Further, this could be displayed on car's windshield using head-up display technology (which is our future work). A major technical difference between conventional and our proposed system is in image processing and computer vision technology without hardware support. Our proposed system requires detecting lanes on streaming images from camera, matching GPS data, and drawing lanes and other navigation information on screen. We present new algorithms to achieve these task quickly and accurately.

Performance is another challenge in our proposed system as image processing and computer vision tasks are compute-intensive and they have to be done in real time. We take advantage of the parallel computing power of mobile GPU that modern mobile phone offers. We offload and accelerate compute intensive parts of the application on mobile GPU using OpenCL [1] while CPU is executing other parts (which is known as CPU-GPU heterogeneous computing). This allows us to speed up the application by more than 2.6 times compared to CPU only execution, reaching up to 18 FPS on Samsung Galaxy Note 4 [2].

2 Lane detection

The first main task is to detect lanes from incoming streaming frames after camera calibration and perspective transformation. 3D to 2D perspective transformation (shown in Figure 1) is required to map GPS data. We developed a light-weight lane detection algorithm based on a simple property of road that lanes are parallel and symmetric multiple lines. In our proposed algorithm, lines are said to be a lane if they are the longest parallel lines within certain range in width. This property greatly reduces the complexity of the task as it requires to loop through pixels only once. Our proposed lane detection algorithm consists of four steps: distance calculation, histogram computation, finding max subarray, and lane reconstruction.

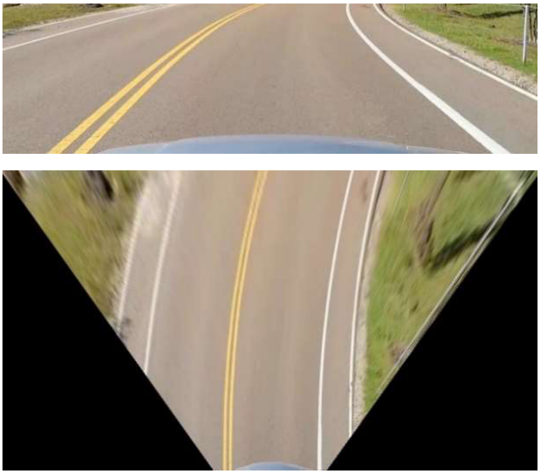


Fig. 1. 3D to 2D perspective transformation.

A. Distance calculation

Distance calculation takes the result of edge detection [3] (not described in this paper) as input and calculates the distance and order between two adjacent white pixels in each row. Examples of distance segments are shown as highlighted lines in the Figure 2. A binary image is scanned and the value of each pixel in each row is inspected. Since input (edge detection) is a binary image, pixel values are either 0 (black) or 255 (white). Whenever a white pixel is detected, the distance from previous white pixel to current white pixel is calculated and stored in 2D array of the same size as edge detection image.

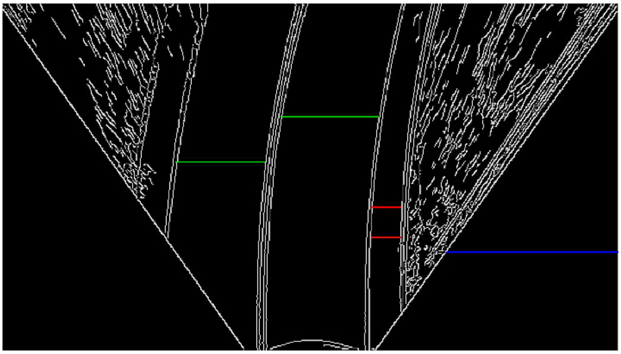


Fig. 2. Distances segmentation on edge detection image after perspective transformation.

B. Histogram computation

Histogram computation counts and returns the frequency of all distance segments in the image. It iterates 2D array generated by previous distance calculation and places each data in a corresponding bin. The bin spans from 0 to 640, representing the minimum and maximum length that can be found in a single row. Figure 3 shows an example which highlights target distances in black-dotted box. In United States, the width of road is between 2.7 to 4.9 meters. In the homography we obtained, one meter is equal to 30 pixels. Thus, the width of road varies from 81 to 147 pixels. If the length of a segment is within this range, we determine it is a lane.

C. Finding MFAS (Most Frequently Appear Segments)

This step selects a target range from histogram data. Histogram bins are scanned from the first bin and the value is added to a temporary variable and compared with current variable. If temp variable is larger than the current, the value of current variable is updated, implying that the data is contiguous. Thus, the value of the temporary end point is increased by 1. If temp variable is equal (the minimum distance is 0) to the current, the value of current variable is set to zero. This means the data loses continuity. Because current value is zero, we update both temporary staring point and end point to next position and start to look for new contiguous data. Finally, we compare and assign the larger summed value to the maximum. This algorithm runs in linear time O(n).

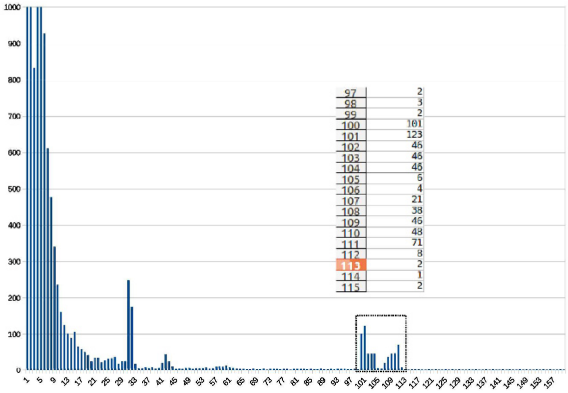


Fig. 3. Histogram and range of MFAS (Most Frequently Appear Segment).

D. Lane reconstruction

This process consists of two tasks: lane drawing and perspective retransformation. For lane drawing, we simply iterate through distance image and change the color of the pixel to green whenever the value falls into the range. Perspective retransformation is simply to perform an inverse perspective transformation on the reconstructed lane image (Figure 4) and overlay to the original input image (Figure 6). OpenCV provides an API that converts the homography matrix H to inverse matrix H−1.

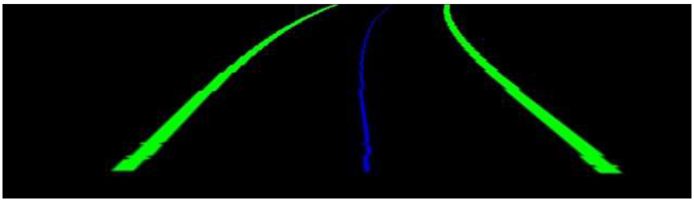


Fig. 4. Reconstructed lane in 3D perspective.

Fig. 5. Final lane detection overlaid to live camera image.

3 GPS Data Matching and Mapping

We use NMEA (National Marine Electronics Association) standards protocol [4, 5, 6] for GPS data. We implement our own simple parser to get positioning coordinates (i.e., latitude and longitude) from the strings (called *sentences* in protocol terminology) returned by the protocol. Our GPS data matching and mapping algorithm has three critical variables: way points, geo points, and prediction points. We use free Google Map database service [7] to get geometric coordinates for origin, destination, and route.

Way points are a series of geometric coordinates that represent a route on a map. Intervals between two adjacent points are not uniform in length so that each connection between two way points becomes close to straight line. Therefore corners and joints have more way points while straight roads have less. We use GPS Visualizer website [8] to get all waypoints. By calculating the spatial relationship between two waypoints, some useful information such as moving direction, angle to North are obtained. These data play a critical role in the following steps.

Geo points refer to actual geographic coordinates of given locations, which are obtained by calculating the angle, θ, from two directions. One direction is between two way points and the other is to true North. The surface of the earth is curved so this problem is similar to calculating the distance on a sphere. For simplicity, two main way points are denoted as *WP1*,*WP2*. The value of *lat1*, *lon1*, *lat2*, *lon2* are corresponding latitude and longitude of WP1, WP2. In our experiment, the minimum distance between two sub way points is set to one meter for accuracy.

Prediction points are geometric points being computed based on the coordinates of existing points. Prediction points change along with the car's moving direction and current position. In our experiment, we use a prediction points matrix whose size is 12 by 15 and distance between adjacent points is 1 meter both horizontally and vertically. This prediction points can be visualized as a matrix dot placed in front of car (see Figure 4). Further, the matrix is divided into three smaller matrices: one center column matrix and two size matrices of 12×7 to simplify computations. In the matrix, only the coordinates of point *End_P* is known, which is the current location of the car.

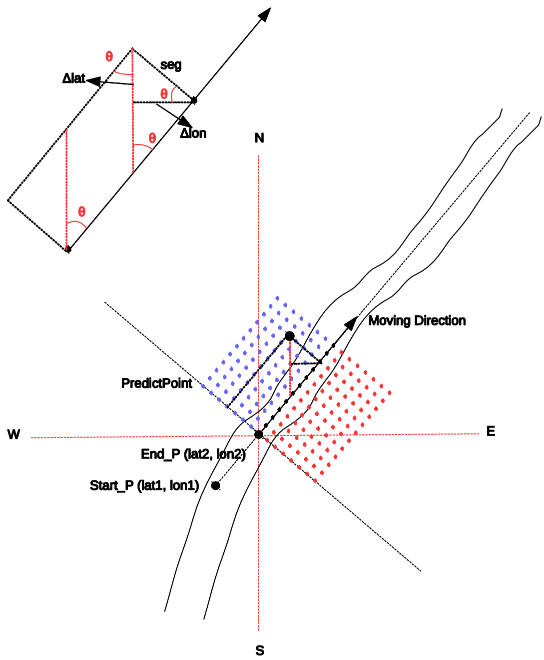


Fig. 4. Relationship between θ and geo points.

The coordinates of other points are calculated by calculating the angle θ as car starts moving. Figure 4 also shows how θ is used to predict the points. Details of *Δlon, Δlat, seg* calculation and other equations are omitted due to page limit.

4 Matching Geo Points and Prediction Points

This process consists of two stages. First, each sub-way point is subtracted by all prediction points and the prediction points that have minimum difference in latitude and longitude in each row are stored in an array *minidx*. This array has 12 elements, one for each row. The selected points which are the closest points to the sub way points are called matched points. The complexity of this process is O(n2).

Next step is to verify whether or not the matched points are valid points. A prediction point is valid if the absolute value of its value is less than a certain threshold. In our experiment, the threshold is set to 0.00002, or 2 meter in length. A flag is set for each valid point. This flag plays an important role in the drawing process as it decides which drawing mechanism to invoke. Figure 5 shows the result of matching geo points and prediction points.

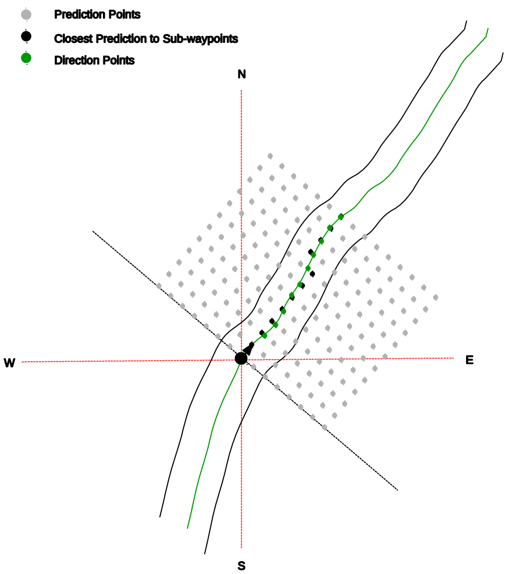


Fig. 5. Matching geo points and prediction points.

5 Drawing

The last step is to overlay the results on the real scene. This process relies on the following three key variables.

- *valid_count* keeps track of the number of valid prediction points in each row. Total twelve points, one in each row.
- *valid_y* keeps track of how many of these valid points has a detected lane in the same row. Total twelve pixels, one in each row.
- *center_line_y_selected* stores the *y* pixel coordinates of the valid prediction points.

The drawing function is invoked when *valid_count* and *valid_y* are more than 5, and the lane is drawn based on the values in *center_line_y_selected* array. If any of these are less than 5, the lane will not be drawn. The drawing function simply loops though the output frame and changes the value of pixel. Figure 6 shows the final result.

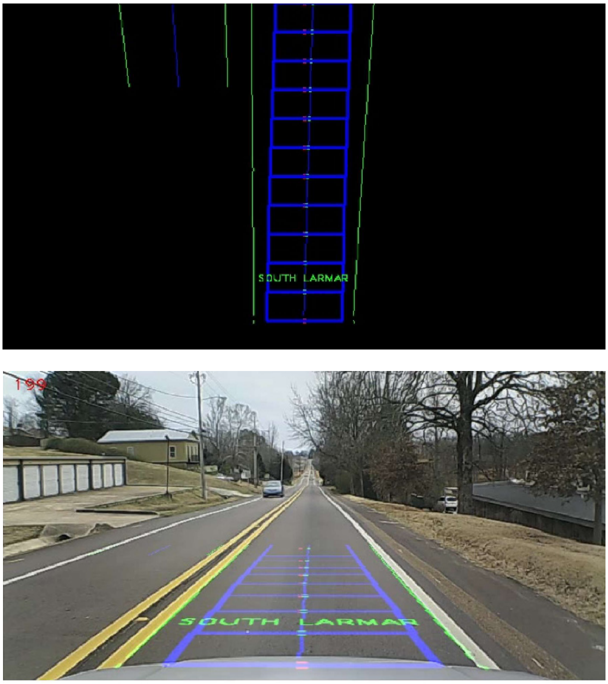


Fig. 6. Matched Points-Detected Lane.

Figure 7 shows the overall task flow of our proposed ARNavi system. Our implementation is mixed with Java, C/C++, and OpenCL code running both on CPU and GPU. We leverage Java JNI (Java Native Interface) framework [9] to merge Java and C/C++ code. We also port and use OpenCV framework and data structures on Android system. Data type *Mat* must be converted to *ocl::oclMat* to use OpenCV's OCL module. Figure 7 also includes the names of high level data structures and functions used in our implementation.

6 Mobile GPU Acceleration

In our proposed ARNavi, image processing algorithms are the most time consuming tasks. These compute intensive processes account for about 80% of total execution time. Therefore, we target to offload those tasks to a mobile GPU to speed up.

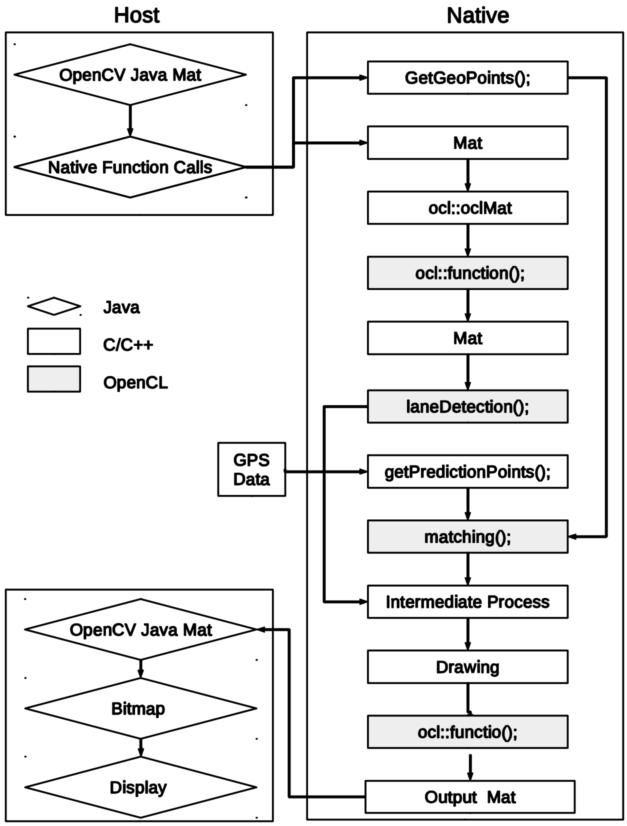


Fig. 7. Overall task flow of ARNavi system.

OpenCV development team has ported commonly used image processing and computer vision functions to OpenCL (*ocl* module) [10]. As long as libOpneCL.so library is properly linked and OpenCL compatible GPU is present in the system, applications can use GPU accelerated functions provided by OpenCL. However, image data (i.e., *Mat* structure) type must be converted to GPU version (i.e., *ocl::Mat*) before launching kernels. There is also data copy overhead between CPU and GPU.

Our profiling shows that image processing and computer vision tasks including lane detection and prediction and way points matching are top time consuming functions. We ported and optimized these functions in OpenCL. Lane detection algorithm calculates the length between two horizontally adjacent pixels, current and previous white pixels. Thus, the computation possesses inherit data dependences between pixels (e.g., loop carried dependence in serial implementation). Therefore, we let each thread process an entire row by iterating from the beginning to the end. This thread mapping results in workgroups, each of which consists of a fixed number of consecutive threads.

As our test input size is 360 (row) by 640 (col), there are 360 threads executing each row (640 elements) at the same time.

7 Experiments and Results

Our test platform is Samsung Galaxy Note 4. Table I lists the hardware and software specifications of our test platform. Serial and parallel implementations share the same framework and code base except that the parallel version runs parts of the code on a GPU. While we use a system wall clock timer to measure overall execution time of the application, we also use OpenCL event profiler for accurate GPU kernel time measurement. All data read/write overheads are included in GPU execution time.

Device	Samsung Galaxy Note 4
OS	Android 4.2
AP	Qualcomm Snapdragon 805
CPU	Qualcomm Krait 450 Quad-core at 2.7GHz
GPU	Qualcomm Adreno 420 at 600 MHz
OpenCL	1.2 Full
OpenCV	Open4Android 2.4.10

Table 1. Test platform specifications.

	Total exec. time (FPS)	GPU kernels
CPU	0.138 (7.25)	0.108
GPU	0.053 (18.87)	0.023
Speedup	2.6	4.7

Table 2. Performance comparison between CPU and GPU execution (one frame execution time in seconds).

Table II compares the per-frame execution time of CPU only and GPU-powered execution. It also shows the execution time of GPU kernels. Overall, GPU-powered execution runs about 2.6× faster than CPU-only execution, reaching up to 18.87 FPS.

While all offloaded computations are accelerated on GPU, the only exception is lane detection kernel whose execution time increases from 0.002071 to 0.002595 second. Our analysis shows that data copy overhead between host and device offsets the benefits of kernel acceleration. Out of 0.002595 seconds, about 28% accounts for data copy overhead. Other GPU kernels suffer less from this overhead as either the amount of data copy is smaller or relative kernel execution time is larger.

8 CONCLUSION

This paper presents a GPU powered mobile augmented reality navigation system, called ARNavi. We have developed and implemented our own fast and accurate lane detection & matching algorithms. We also demonstrate that the compute intensive parts of the application are accelerated by mobile GPU to achieve higher performance without any hardware support. With GPU acceleration, the application runs 2.63× faster. Further algorithm and GPU kernel optimizations remain as future works and our proposed system can be applied to car's augmented reality head-up display. More details not included in this paper can be obtained by contacting the authors.

9 ACKNOWLEDGEMENT

This work is supported in part by National Science Foundation (NSF) grant CCF-1337138.

REFERENCES

1. K. Group, "OpenCL." [Online]. Available: https://www.khronos.org/opencl/
2. "Samsung," http://www.samsung.com/us/explore/galaxy-_note-_4-_features-_and-_specs/.
3. J. Canny, "A computational approach to edge detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PAMI-8, no. 6, pp. 679-698, Nov 1986.
4. "NMEA 0183 Standard," http://www.nmea.org.
5. "GPS - NMEA sentence information," http://http://aprs.gids.nl/nmea/.
6. "NMEA Reference Manual," https://www.sparkfun.com/datasheets/GPS/NMEA%20Reference%20Manual1.pdf.
7. Google, "Google Maps." [Online]. Available: https://www.google.com/maps
8. "GPS Visualizer." [Online]. Available: http://www.gpsvisualizer.com/
9. Oracle, "Java Native Interface." [Online]. Available: http://docs.oracle.com/javase/7/docs/technotes/guides/jni/
10. OpenCV developers, "OpenCL module introduction." [Online]. Available: http://docs.opencv.org/modules/ocl/doc/introduction.html

A Simulator to Study the Effects of Color and Color Blindness on Motion Sickness in Virtual Reality Using Head-Mounted Displays

Dmitri A. Gusev, David M. Whittinghill, Justin Yong

Purdue University
Columbus, IN, U.S.A., West Lafayette, IN, U.S.A.
dgusev@purdue.edu, dmwhittinghill@purdue.edu,
yongj@purdue.edu

Abstract. We have created an Android-based virtual reality (VR) simulator to investigate the effects that color and color blindness have on visually induced motion sickness in virtual reality using Samsung Gear VR. We simulate a patterned optokinetic drum or sphere spinning for test subjects to look at. By choosing specific colors and using color blindness filters, we can investigate the impact on a colorblind person who looks at a spinning optokinetic drum or sphere that seemingly shows one flat color. The amount of the time that elapses before motion sickness symptoms occur can be recorded and compared among observers with normal color vision, colorblind observers, and test subjects who have a color blindness filter enabled. We believe that our new tool will contribute to improvements in accessibility of VR games and simulations developed for modern head-mounted displays, including those designed to hold a mobile device, such as a smartphone or tablet.

1 Introduction

Development for virtual reality (VR) has increased in recent years to be used with the upcoming releases of the Oculus Rift and HTC Vive head-mounted displays, as well as the already released Samsung Gear VR. As head-mounted VR becomes increasingly accessible to consumers, developers will need to be ready to utilize VR for a wide variety of applications. For example, VR can be used to virtually prototype products in order to analyze their ergonomics, constructability, and aesthetics before production [1]. VR can also be used in the medical field to provide a means of treatment for patients, such as therapy for those with post-traumatic stress disorder [2]. The entertainment industry can utilize VR to create immersive experiences for consumers. No matter the use, VR will often cause its users to feel discomfort due to simulator sickness.

Simulator sickness is a major issue for VR developers. As soon as a user feels any discomfort, the VR application will no longer be usable. Two publications disagree as to whether color plays a role in the phenomenon of motion sickness in VR. Bonato et al. [3] studied visually induced motion sickness (VIMS) using an *optokinetic drum*, a

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_22

rotating instrument to test vision in which a test subject is seated facing the cylinder's interior wall. They reported that motion sickness onset was faster and more severe when human observers were shown chromatic stripes (white, red, yellow, black, green, and blue) in an optokinetic drum than it was in two other experiments that had only involved black, white, and gray stripes. On the other hand, So and Yuen [4] concluded that changes of color did not affect the levels of VIMS inside a virtual environment in which four different chromatic pairs were employed. Hence the problem remains open as to whether color is a significant contributing factor to motion sickness in VR, and if it is, then which VIMS symptoms are impacted the most.

Given that the total frequency of color vision defects in the general population is reported to be up to 8% in males [5] and at least 0.5% in females [6], it is also important to find out if motion sickness in VR affects colorblind people less than it does viewers with normal trichromatic vision. The most common inherited color defects are the red-green ones [7], and more than 75% of red-green color blind individuals have difficulties with daily tasks [8]. If a colorblind person were presented with two colors that look exactly the same due to the color deficiency, would the person still succumb to motion sickness inside an optokinetic drum?

This paper describes the design and explains the decisions made during our development of a simulator for Samsung Gear VR helmets compatible with Samsung Galaxy Note 4 and S6 smartphones. The simulation is intended to determine the effect that color in general, and specifically color as viewed by a colorblind person, has on motion sickness in modern head-mounted displays designed to hold an Android-based mobile device. The effect can be quantified using the simulator sickness questionnaire developed by Kennedy et al. [9].

The software engineering aspects of the design and implementation of the simulator will be explained in the next section. After that, the color science angle of the problem related to color selection will be covered. Then we will describe the controls and provide screenshots of the simulator in action on an Android smartphone. Finally, the conclusions will be drawn, and the future work directions will be discussed.

2 Design and Implementation

We developed a simulation that contains two optokinetic drums and two optokinetic spheres. Two different textures, one with a checkerboard pattern and one with a striped pattern, are applied to each type of object. A color blindness filter is applied to the camera to simulate color blindness. The test subject is placed in the center of a rotating drum or sphere. The choice of object that the subject is inside of and the rotation speed of the object can be modified with controls on the Samsung Gear VR helmet.

2.1 Unity

According to John Carmack [10], 90-95% of developers for the Oculus Rift VR platform were using Unity, a cross-platform game engine developed by Unity Technologies, as of March 2015. We chose to develop in Unity because it provides all the tools necessary to quickly produce the simulation. Meshes can be quickly added, and C# scripts can be used to add behavior with a few lines of added code. The attributes of objects can be swiftly edited to fit the needs of the final simulation. The colors chosen for the patterns on the drums and spheres are a special pair of red and green. The shades of red and green could be quickly changed so that when they are viewed through the color blindness simulator, the colors would become indistinguishable from each other. The pair of colors were selected to have the same luminosity levels. The details of the color selection process will be covered in the next section. The optokinetic drum in the simulation is modeled after a physical optokinetic drum and behaves similarly. Likewise, the optokinetic sphere is modeled after a physical optokinetic sphere.

We examined two third-party solutions for color blindness simulation in Unity — Simulate Color Blindness 1.1 by Drash [11] and Color Blindness Simulator for Unity 1.0.0 by Gulti [12]. The latter solution simulated only two types of color blindness — protanopia or deuteranopia. We chose the former product that offered 9 simulation modes:

1. Normal color vision.
2. *Protanopia* — a variety of red-green color blindness such that the people affected by it are less sensitive to red light; to these people, red tones appear beige or grey, and green tones look beige or grey, like the red ones.
3. *Protanomaly* — a condition less severe than protanopia; the L-cones responsible for sensitivity to red light are not missing, but merely defective.
4. *Deuteranopia* — another variety of red-green color blindness; the people affected by it are less sensitive to green light, because the cones of the type more sensitive to the medium wavelengths are missing.
5. *Deuteranomaly* — a condition less severe than deuteranopia; the green sensitive cones are not missing, but their peak of sensitivity is shifted toward that of the red sensitive cones.
6. *Tritanopia* — a very rare variety of color blindness, approximately 1 in 10,000 people are afflicted by it; they confuse blue with green and violet with yellow.
7. *Tritanomaly* — a milder form of blue-yellow color blindness than tritanopia; here the short-wavelength cones (S-cones) are present but have some kind of mutation.
8. *Achromatopsia* — the complete lack of the perception of color in a subject, who sees only black, white, and shades of grey.
9. *Achromanomaly* — a milder or transitional form of achromatopsia; this condition is more commonly known as *incomplete achromatopsia*, or *dyschromatopsia*.

More detailed data on prevalence of different types of color blindness are available elsewhere [13].

2.2 Samsung Gear VR

We developed for the Samsung Gear VR because it is a high quality consumer head-mounted display based on the Oculus Rift technology [10]. It proved fairly straightforward to deploy the simulation from Unity to Samsung Gear VR by cross-compiling the Unity project to Android for one of the supported smartphones, such as Samsung Note 4, Samsung Galaxy S6, or Samsung Galaxy S6 Edge. The developer must remember to include the latest version of the OVR Mobile SDK as an asset into the Unity project in order to support the interface with the helmet controls. Figure 1 shows a screenshot of the Unity project running a simulation of an optokinetic drum with a checkerboard pattern in the normal color vision mode.

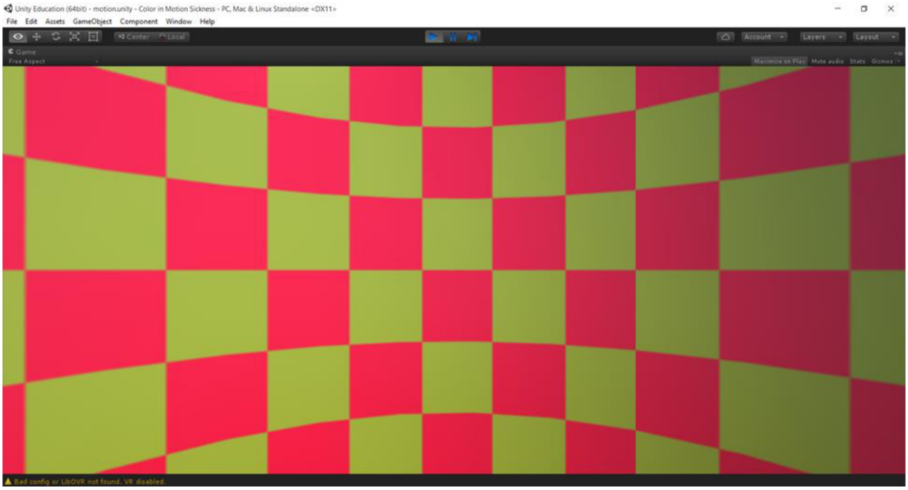


Fig. 1. Running an optokinetic drum simulation in Unity at the development stage

Utilizing the touchpad that is integrated into the headset, the user can easily change the settings of the simulation, which may help protect the system's integrity in case of a sudden onset of motion sickness. Section 4 will provide more details on how the controls work and what the mobile device displays.

Another advantage of Samsung Gear VR over Oculus Rift SDK 2 from our specific standpoint is that the test subject is no longer tethered to a desktop or laptop PC by cables, which allows for a cleaner experiment.

But how should we choose colors for the design of the simulated optokinetic drums and spheres? The answer will be given in the next section.

3 Color Selection

We chose colors that would have the same luminosity levels after being modified by the simulated color blindness filter. We start by picking the first color that will be shown when seen with no color deficiency. We compute the simulated color using the equation

$$\begin{bmatrix} R_S \\ G_S \\ B_S \end{bmatrix} = \boldsymbol{\varphi}_{CVD} \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \qquad (1)$$

where R_S, G_S, and B_S are the simulated color components, $\boldsymbol{\varphi}_{CVD}$ is the transformation matrix to achieve the correct color adjustments, and R, G, and B are the original color components. To simulate deuteranomaly, Machad et al. [14] recommend

$$\boldsymbol{\varphi}_{CVD} = \begin{bmatrix} 0.367322 & 0.860646 & -0.227968 \\ 0.280085 & 0.672501 & 0.047413 \\ -0.011820 & 0.042940 & 0.968881 \end{bmatrix}. \quad \textbf{(2)}$$

The luminance of a color can be given by its Y value in the XYZ color space [15]. We use the standard sRGB as our initial RGB color space. To convert from RGB to XYZ, RGB values must be made linear. According to Lindbloom [16], the linear components, r, g, and b, represented by **v**, are calculated from the original R, G, and B components, represented by **V** according to the formula

$$\boldsymbol{v} = \boldsymbol{V}^{\gamma}. \quad \textbf{(3)}$$

The gamma of sRGB can be approximated by $\gamma = 2.2$ [17]. The linear rgb values are multiplied by a transformation matrix, based on the RGB color space and reference white, to be converted to XYZ. Using the sRGB color space and reference white D_{65}, as recommended by Lindbloom [18], the conversion from rgb to XYZ is given by the formula

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124564 & 0.3575761 & 0.1804375 \\ 0.2126729 & 0.7151522 & 0.0721750 \\ 0.0193339 & 0.1191920 & 0.9503041 \end{bmatrix} \begin{bmatrix} r \\ g \\ b \end{bmatrix}. \quad \textbf{(4)}$$

In particular, the luminance of an RGB color is

$$Y = 0.2126729 \cdot R^{2.2} + 0.7151522 \cdot G^{2.2} + 0.0721750 \cdot B^{2.2}. \quad \textbf{(5)}$$

We can find colors that have the same luminance as the color that has been modified by the color deficiency filter. Using R_{1S}, G_{1S}, and B_{1S} as the simulated first color and R_{2S}, G_{2S}, and B_{2S} as the simulated matching color, we get the condition

$$\begin{gathered} 0.2126729 \cdot (R_{1S})^{2.2} + 0.7151522 \cdot (G_{1S})^{2.2} + 0.0721750 \cdot (B_{1S})^{2.2} = \\ 0.2126729 \cdot (R_{2S})^{2.2} + 0.7151522 \cdot (G_{2S})^{2.2} + 0.0721750 \cdot (B_{2S})^{2.2}. \end{gathered} \quad \textbf{(6)}$$

After choosing a second simulated color, the normal vision color can be calculated by multiplying the simulated RGB values by the inverse of the $\boldsymbol{\varphi}_{CVD}$ matrix,

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \boldsymbol{\varphi}_{CVD}^{-1} \begin{bmatrix} R_S \\ G_S \\ B_S \end{bmatrix}. \quad \textbf{(7)}$$

4 Controls and the Interface

The simulation was originally designed to be viewed on an Oculus Rift SDK 2 connected to a desktop or laptop computer. A few modifications had to be made to allow the simulation to run on a Samsung Gear VR holding an Android-based mobile device. The controls for the simulation had to be changed so that they utilized the Samsung Gear VR's touchpad. To increase the rotation speed of the cylinder or sphere, the user will swipe up on the touchpad. Swiping down will decrease the speed of rotation. Swiping forward and backward will cycle through the cylinders and spheres that the user will be in the middle of. Pressing the back button will cycle through the different color blindness filters.

The resulting simulator interface is visually minimalistic, so nothing distracts the observer from viewing the color patterns. Figures 2-4 feature screenshots taken while running the simulator on a Samsung Note 4 Android-based smartphone under different conditions. Samsung Gear VR combines two side-by-side views for the 3D effect.

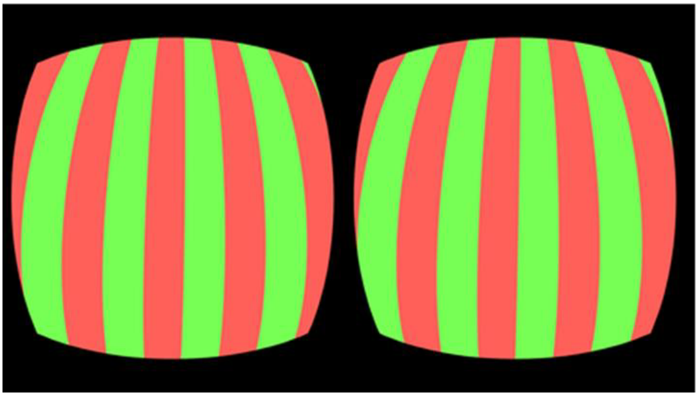


Fig. 2. Optokinetic drum simulation, normal color vision, stripes

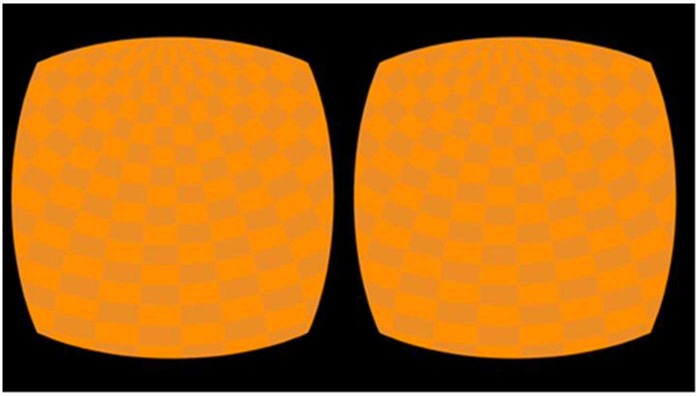


Fig. 3. Optokinetic sphere simulation, protanomaly, the checkerboard pattern

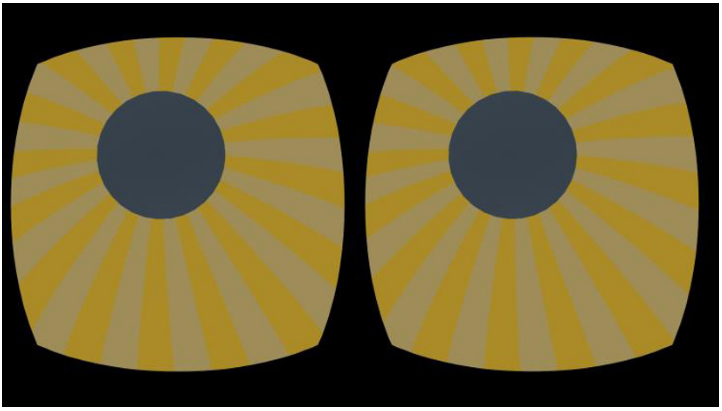


Fig. 4. Optokinetic drum simulation, deuteranomaly, stripes

5 Conclusions and Future Work

We created a simulation for the Samsung Gear VR to investigate what effects color and color blindness have on simulator sickness. By selecting two colors that become indistinguishable when viewed through a color blindness filter, we can observe the differences that a colorblind person would experience while in virtual reality. Because the use of an optokinetic drum or sphere induces motion sickness on test subjects, comparing the length of exposure until VIMS symptoms occur between those who view the simulation with the color blindness filter disabled and those who have the color blindness filter enabled will provide insight into the effect that color blindness has on motion sickness in VR.

We plan to further extend our experiment to other VR systems, such as Oculus Rift head-mounted display and the CAVE projection-based immersive environment with walls. The effect of rotation in different directions should be investigated as well. We intend to include both colorblind individuals and those with normal color vision in our observer pool. In order for our results to be as general as possible, we intend to use realistic virtual environments in addition to those specifically designed to induce motion sickness. We anticipate that the research will impact the design process of modern VR games and simulations so as to broaden participation of people with such disabilities as motion sickness susceptibility and color blindness. We will seek to open-source part of the software produced by this research.

Acknowledgments. The authors thank Dr. Brad Duerstock, Dr. Cary Supalo, and Mr. George Takahashi of Purdue University for their help and advice. The authors are also indebted to Dr. Reiner Eschbach (Monroe Community College), Dr. Ioannis Papapanagiotou (Netflix), and Mr. Thomas Westin (Stockholm University) for their support and inspiration.

References

1. Craig, A., Sherman, W. R. and Will, J.D.: Developing Virtual Reality Applications: Foundations of Effective Design. Morgan Kaufmann, London (2009).
2. Difede, J. and Hoffman, H.: Virtual Reality Exposure Therapy for the World Trade Center Post-Traumatic Stress Disorder: A Case Report. CyberPsychology & Behavior 5 (2002) 529–535
3. Bonato, F., Bubka, A. and Alfieri, L.: Display Color Affects Motion Sickness Symptoms in an Optokinetic Drum. Aviat. Space Env. Med. 75 (2004) 306–311
4. So, R. H. Y. and Yuen, S. L.: Comparing Symptoms of Visually Induced Motion Sickness Among Viewers of Four Similar Virtual Environments with Different Color. In: Shumaker, R. (ed.): Virtual Reality. HCII 2007, LNCS 4563. Springer-Verlag, Berlin Heidelberg (2007) 386–391
5. Deeb, S. S.: The Molecular Basis of Variation in Human Color Vision. Clin. Genet. 67 (2005) 369–377
6. Sharpe, L. T., Stockman, A., Jägle, H. and Nathans, J.: Opsin Genes, Cone Photopigments, Color Vision, and Color Blindness. In: Gegenfurtner, K. R. and Sharpe, L. T. (eds.): Color Vision: From Genes to Perception. Cambridge University Press, Cambridge (1999) 3–52
7. Pease, P. L.: Color Vision. In: Benjamin, W. J. (ed.): Borish's Clinical Refraction. 2nd edn. Butterworth-Heinemann, Oxford (2006) 289–355
8. Steward, J. M. and Cole, B. L.: What Do Color Vision Defectives Say About Everyday Tasks? Optom. Vis. Sci. 66 (1989) 288–295
9. Kennedy, R.S., Lane, N.E., Berbaum, K.S., & Lilienthal, M.G.: Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. Int. J. Aviat. Psychol. 3 (1993) 203–220
10. Carmack, J.: The Dawn of Mobile VR. GDC 2015, San Francisco, CA (2015) Retrieved from: http://www.gdcvault.com/play/1022304/The-Dawn-of-Mobile
11. Drash: Simulating Colorblindness in Unity (2013) Retrieved from: http://blog.crunchywood.com/
12. Gulti Co., Ltd.: Color Blindness Simulator for Unity (2014) Retrieved from: https://www.assetstore.unity3d.com/en/#!/content/19039
13. Colblindor: Types of Color Blindness (2010) Retrieved from: http://www.color-blindness.com/2010/03/09/types-of-color-blindness/
14. Machad, G. M., Oliveira, M. M., and Fernandes, L. A.: A Physiologically-Based Model for Simulation of Color Vision Deficiency (2009) Retrieved from: http://www.inf.ufrgs.br/~oliveira/pubs_files/CVD_Simulation/CVD_Simulation.html
15. Kerr, D. A.: The CIE XYZ and xyY Color Spaces (2010) Retrieved from: http://dougkerr.net/Pumpkin/articles/CIE_XYZ.pdf
16. Lindbloom, B.: RGB to XYZ (2001) Retrieved from: http://www.brucelindbloom.com/index.html?Eqn_RGB_XYZ_Matrix.html
17. Lindbloom, B.: RGB Working Space Information (2001) Retrieved from: http://www.brucelindbloom.com/index.html?WorkingSpaceInfo.html#Specifications
18. Lindbloom, B.: RGB/XYZ Matrices (2001) Retrieved from: http://www.brucelindbloom.com/index.html?Eqn_RGB_XYZ_Matrix.html

An Artificial Intelligence Approach to Dyscalculia

Filipa Ferraz[1], José Neves[2*], António Costa[2] and Henrique Vicente[2, 3]

[1] Departamento de Informática, Universidade do Minho, Braga, Portugal
filipatferraz@gmail.com
[2] Centro Algoritmi, Universidade do Minho, Braga, Portugal
{jneves; acosta}@di.uminho.pt
[3] Departamento de Química, Escola de Ciências e Tecnologia,
Universidade de Évora, Évora, Portugal
hvicente@uevora.pt
* Corresponding author: phone: +351-934201337; fax: +351-253604471;
e-mail: jneves@di.uminho.pt

Abstract. *Dyscalculia* stands for a brain-based condition that makes it hard to make sense of numbers and mathematical concepts. Some adolescents with dyscalculia cannot grasp basic number concepts. They work hard to learn and memorize basic number facts. They may know what to do in mathematical classes but do not understand why they are doing it. In other words, they miss the logic behind it. However, it may be worked out in order to decrease its degree of severity. For example, *disMAT*, an *app* developed for *android* may help children to apply mathematical concepts, without much effort, that is turning in itself, a promising tool to dyscalculia treatment. Thus, this work focuses on the development of an *Intelligent System* to estimate children evidences of dyscalculia, based on data obtained on-the-fly with *disMAT*. The computational framework is built on top of a *Logic Programming* framework to *Knowledge Representation and Reasoning*, complemented with a *Case-Based* problem solving approach to computing, that allows for the handling of incomplete, unknown, or even contradictory information.

Keywords: Dyscalculia · Knowledge Representation and Reasoning · Logic Programming · Case-Based Reasoning · Similarity Analysis.

1 Introduction

Dyscalculia can be defined as a mathematical learning disability that affects the ability to perform operations and make the proper use of arithmetic. Frequently described as the dyslexia or blindness for numbers, the dyscalculia is hard to be well-diagnosed, despite the incidence on 6 to 7% in the population [2].

Besides affecting the realization of simple calculations with two digits and basic operations (i.e., addition, subtraction, multiplication and division), dyscalculia also influences tasks such as to distinguish left from right, to tell the time, or even to count money/cash. Since this specific developmental disorder can be reflected in various

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_23

areas of mathematic, dyscalculia may be set in six sub-areas taking into account the most affected areas [3], namely:

- *Lexical dyscalculia* – troubles in reading mathematical symbols;
- *Verbal dyscalculia* – troubles in naming mathematical quantities, numbers and symbols;
- *Graphic dyscalculia* – troubles in writing mathematical symbols;
- *Operational dyscalculia* – troubles in performing mathematical operations and calculus;
- *Practognostic dyscalculia* – troubles in enumerating, manipulating and comparing real objects and pictures; and
- *Ideagnostic dyscalculia* – troubles in mental operations and understanding mathematical concepts.

Furthermore, dyscalculia can also be classified according to the state of neurological immaturity [4, 5], i.e.:

- A *former state*, related with the individuals that react favorably to therapeutically intervention;
- A *second one*, associated to the individuals who have other learning disabilities; and
- A *last one*, linked to the individuals that feature an intellectual deficit caused by a neurological injury(ies).

Dyscalculia is irreversible, i.e., the disorder cannot be treated, but can be worked out in order to decrease its degree of severity. Therefore, the therapeutics must use attractive methods to help the children to deal with the mathematical issues [6]. In order to meet this challenge an *android app*, *disMAT*, was developed in order to improve the arithmetic skills of children through simple tasks that look at their weak spots, like measures, without much effort, obligation nor awareness, turning this support system into a promising tool [1]. This *app* is also attractive with respect to the fact that children nowadays are keeping up with the technologic era, carrying their tablets and smartphones everywhere to entertainment.

This paper addresses the theme of dyscalculia and describes an attempt to diagnosis this disorder using a Case-Based Reasoning (CBR) approach to problem solving [7, 8]. The *app disMAT* was applied to a group of children, and some parameters were recorded aiming to build up a knowledge base. To set the structure of the information and the associate inference mechanisms, a computational framework centered on a *Logic Programming* (*LP*) based approach to knowledge representation and reasoning was used. It handles of unknown, incomplete, or even contradictory data or knowledge [9].

2 Background

2.1 Knowledge Representation and Reasoning

LP has been used for knowledge representation and reasoning in different areas, like Model Theory [10, 11], and Proof Theory [12, 13]. In the present work the proof theoretical approach is followed in terms of an extension to *LP*. An *Extended Logic Program* is a finite set of clauses in the form:

$$\{$$

$$p \leftarrow p_1, \cdots, p_n, not\ q_1, \cdots, not\ q_m$$

$$?\,(p_1, \cdots, p_n, not\ q_1, \cdots, not\ q_m)\ \ (n, m \geq 0)$$

$$exception_{p_1}$$

$$\ldots$$

$$exception_{p_j}\ \ (0 \leq j \leq k),\ \ being\ k\ an\ integer\ number$$

$$\} :: scoring_{value}$$

where "*?*" is a domain atom denoting falsity, the p_i, q_j, and p are classical ground literals, i.e., either positive atoms or atoms preceded by the classical negation sign $\neg$ [12]. Under this formalism, every program is associated with a set of abducibles [10, 11], given here in the form of exceptions to the extensions of the predicates that make the program. The term $scoring_{value}$ stands for the relative weight of the extension of a specific *predicate* with respect to the extensions of the peers ones that make the overall program.

In order to evaluate the knowledge that can be associated to a logic program, an assessment of the *Quality-of-Information* (*QoI*), given by a truth-value in the interval [0, 1], that stems from the extensions of the predicates that make a program, inclusive in dynamic environments [14, 15]. The objective is to build a quantification process of *QoI* and *DoC* (Degree of Confidence), being the latter a measure of one's confidence that the argument values or attributes of the terms that make the extension of a given predicate, with relation to their domains, fit into a given interval [16]. The *DoC* is evaluated as depicted in Fig. 1 and computed using $DoC = \sqrt{1 - \Delta l^2}$, where Δl stands for the argument interval length, which was set in the interval [0, 1]. The scheming procedure is better understood looking at. Thus, the universe of discourse is engendered according to the information presented in the extensions of such predicates, according to productions of the type:

$$predicate_i - \bigcup_{1 \leq i \leq m} clause_j\left(\left(QoI_{x_1}, DoC_{x_1}\right), \cdots, \left(QoI_{x_m}, DoC_{x_m}\right)\right) :: QoI_i :: DoC_i \tag{1}$$

where U and m stand, respectively, for *set union* and the *cardinality* of the extension of $predicate_i$. QoI_i and DoC_i stand for themselves [9].

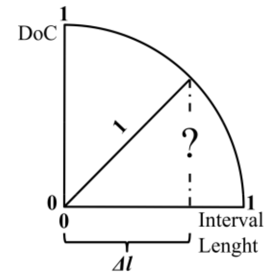


Fig. 1. Evaluation of the Degree of Confidence.

As an example, let us consider the logic program given by:

$$
\begin{aligned}
&\{\\
&\quad \neg f_1\left((QoI_{x_1}, DoC_{x_1}),\ (QoI_{y_1}, DoC_{y_1}),\ (QoI_{z_1}, DoC_{z_1})\right)\\
&\qquad\qquad\qquad \leftarrow not\left((QoI_{x_1}, DoC_{x_1}),\ (QoI_{y_1}, DoC_{y_1}),\ (QoI_{z_1}, DoC_{z_1})\right)\\
&\quad f_1 \underbrace{\left((QoI_{\perp}, DoC_{\perp}), (QoI_2, DoC_2), \left(QoI_{[1,\,1.75]}, DoC_{[1,\,1.75]}\right)\right)}_{attribute`s\ values} :: QoI :: DoC\\
&\qquad \underbrace{[3, 6] \qquad\quad [0, 3] \qquad\qquad [0.5, 2]}_{attribute`s\ domains}\\
&\quad exception_{f_1}\left(\left(QoI_{[5,\,6]}, DoC_{[5,\,6]}\right), (QoI_{\perp}, DoC_{\perp}), (QoI_{0.8}, DoC_{0.8})\right) :: QoI :: DoC\\
&\quad exception_{f_1}\left((QoI_4, DoC_4), \left(QoI_{[0,\,2]}, DoC_{[0,\,2]}\right), (QoI_{\perp}, DoC_{\perp})\right) :: QoI :: DoC\\
&\quad exception_{f_1}\left((QoI_{\perp}, DoC_{\perp}), (QoI_{1.5}, DoC_{1.5}), \left(QoI_{[1,\,2]}, DoC_{[1,\,2]}\right)\right) :: QoI :: DoC
\end{aligned}
$$

$\} :: 1$ (*once the universe of discourse is set in terms of the extension of only one predicate*)

where $\perp$ denotes a null value of the type unknown. It is now possible to split the abducible or exception set into the admissible clauses or terms and evaluate their QoI_i. A pictorial view of this process is given below (Fig. 2) as a pie chart.

2.2 Case-Based Reasoning

A *CBR* methodology for problem solving stands for an act of finding and justifying the solution to a given problem based on the consideration of similar past ones, by reprocessing and/or adapting their data or knowledge [8, 9]. In *CBR* – the cases – are stored in a *Case Base*, and those cases that are similar (or close) to a new one are used in the problem solving process. There are examples of its use in *The Law* with respect to dispute resolution [17], in *Medicine* [18], among others. The typical *CBR* cycle presents the mechanism that should be followed to have a consistent model. The first stage consists in the initial description of the problem. The new case is defined and it is used to retrieve one or more cases from the repository. At this point it is important to identify the characteristics of the new problem and retrieve cases with a higher degree of similarity to it. Thereafter, a solution for the problem emerges, on the *Reuse* phase, based on the blend of the new case with the retrieved ones. The suggested solution is reused, i.e., adapted to the new case, and a solution is provided [7, 8]. However, when adapting the solution it is crucial to have feedback from the user, since automatic adaptation in existing systems is almost impossible. This is the *Revise* stage, in which the suggested solution is tested by the user, allowing for its correction, adaptation and/or modification, originating the test repaired case that sets the solution to the new problem. The test repaired case must be correctly tested to ensure that the solution is indeed correct. Thus, one is faced with an iterative process since the solution must be tested and adapted while the result of applying that solution is inconclusive. During the *Retain* (or *Learning*) stage the case is learned and the knowledge base is updated with the new case [7, 8].

Despite promising results, the current *CBR* systems are neither complete nor adaptable enough for all domains. In some cases, the user is required to follow the similarities method used by the system, even if it do not fit into their needs [9]. Moreover, other problems may be highlighted. Moreover, the existent *CBR* systems have limitations related to the capability of dealing with unknown, incomplete and contradictory information.

3 Methods

The data was taken from the evaluation of 148 (one hundred forty eight) children of a primary school in the north of Portugal who played the *disMAT app*. For each participant was recorded the age, the number of game levels completed, the minimum score obtained, as well as the maximum score, the response time in each of the three levels and the classification of understanding and doing difficulties through the game. This section demonstrates how the information comes together and how it is processed.

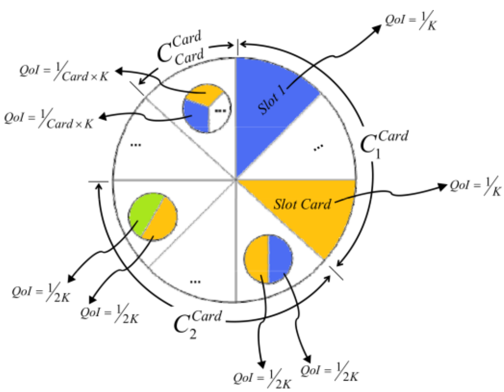


Fig. 2. QoI's values for the abducible set of clauses referred to above, where the clauses cardinality set, *K*, is given by the expression $C_1^{Card} + C_2^{Card} + \cdots + C_{Card}^{Card}$.

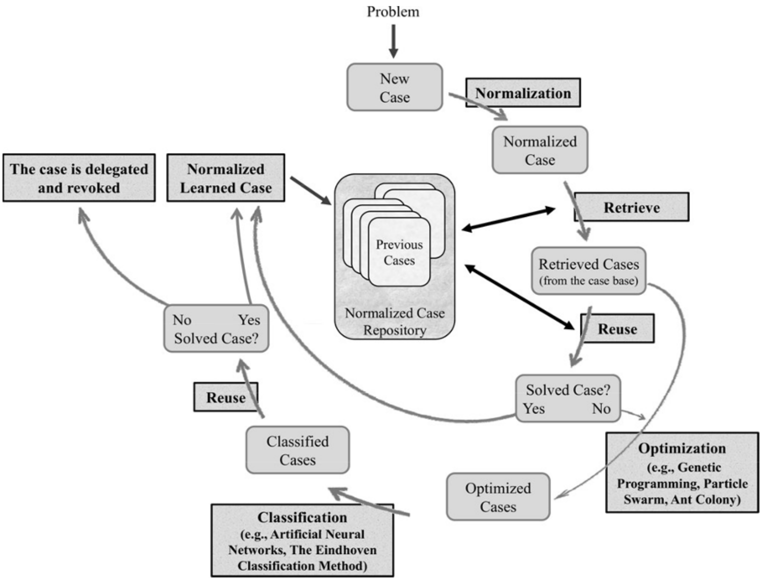


Fig. 3. An extended view of the CBR cycle [9].

3.1 A Logic Programming approach to Data Processing

The knowledge database is given in terms of the extensions of the relations depicted in Fig. 4, which stand for a situation where one has to manage information about children evidences of dyscalculia. The knowledge base includes features obtained by both objective and subjective methods. Under this scenario some incomplete and/or unknown data is also present. For instance, the *Level 2 Response Time* in case 1 is unknown, which is depicted by the symbol $\perp$, while *Level 3 Response Time* ranges in the interval [0, 6]. The *Understanding* and *Doing Difficulties* columns ranges in the interval [0, 2], wherein 0 (zero), 1 (one) and 2 (two) denote, respectively, *easy*, *medium* and *hard*.

Applying the algorithm presented in [9, 16] to the fields that make the knowledge base for dyscalculia diagnosis (Fig. 4), excluding of such a process the *Description* ones, and looking to the DoC_s values obtained, it is possible to set the arguments of the predicate *dys*calculia_*diag*nosis (*dys_diag*) referred to below, that also denotes the objective function with respect to the problem under analyze.

$$dys_{diag}: Age\ ,\ L_{evels}C_{ompleted},\ Min_{imum\ Score},\ Ave_{rage\ Score}, Max_{imum\ Score},$$
$$L_{evel}1\ R_{esponse\ Time},\ L_{evel}2\ R_{esponse\ Time},\ L_{evel}3\ R_{esponse\ Time},$$
$$U_{nderstanding}D_{ifficulties},\ D_{oing}D_{ifficulties} \rightarrow \{0,1\}$$

where 0 (zero) and 1 (one) denote, respectively, the truth values *false* and *true*.

	Dyscalculia Diagnosis											
Attributes of the Feature Vector:	#	Age	Nº of Levels Completed	Minimum Score	Average Score	Maximum Score	Level 1 Response Time	Level 2 Response Time	Level 3 Response Time	Understanding Difficulties	Doing Difficulties	Description
Feature Vector Attributes:	1	6	0	0	5	10	[12, 30]	$\perp$	[0, 6]	2	2	Description 1
	2	7	2	90	190	290	[12, 21]	[0, 11]	[6, 16]	1	0	Description 2
	…	…	…	…	…	…	…	…	…	…	…	…
	148	6	$\perp$	0	60	120	[12, 30]	[11, 24]	[0, 6]	$\perp$	1	Description 148
Feature Vector Domains:		[5, 10]	[0, 3]	[0, 300]	[5, 300]	[10, 300]	[12, 60]	[0, 30]	[0, 20]	[0, 2]	[0, 2]	

Fig. 4. A fragment of the knowledge base to predict children evidences of dyscalculia.

Exemplifying the application of the algorithm presented in [9, 16], to a term (case) that presents feature vector ($Age = 8$, $L_{evels}C_{ompleted} = 2$, $Min_{imum\ Score} = 20$, $Ave_{rage\ Score} = 120$, $Max_{imum\ Score} = 220$, $L_{evel}\ 1\ R_{esponse\ Time} = [12, 20]$, $L_{evel}\ 2\ R_{esponse\ Time} = 15$, $L_{evel}\ 3\ R_{esponse\ Time} = 0$, $U_{nderstanding}\ D_{ifficulties} = \perp$, $D_{oing}\ D_{ifficulties} = \perp$), one may have:

$$\{$$

$$\neg\, dys_{diag}\left((QoI_{Age}, DoC_{Age}),\ \cdots, (QoI_{L1R}, DoC_{L1R}),\ \cdots, (QoI_{DD}, DoC_{DD})\right)$$

$$\leftarrow not\ dys_{diag}\left((QoI_{Age}, DoC_{Age}),\ \cdots, (QoI_{L1R}, DoC_{L1R}),\ \cdots, (QoI_{DD}, DoC_{DD})\right)$$

$$dys_{diag}\ \underbrace{\left((1,\ 1),\ \ \cdots,\ \ (1,\ 0.99),\ \ \cdots,\ \ (1,\ 0)\right)}_{\substack{attribute`s\ quality-of-information \\ and\ respective\ confidence\ values}} :: 1 :: 0.80$$

$$\underbrace{[0.6, 0.6]\quad \cdots \quad [0, 0.17] \quad \cdots \quad [0, 1]}_{attribute`s\ values\ ranges\ once\ normalized}$$

$$\underbrace{[0, 1]\quad \cdots \quad [0, 1] \quad \cdots \quad [0, 1]}_{attribute`s\ domains\ once\ normalized}$$

$$\} :: 1$$

This approach allows the representation of the case repository in a graphic form, showing each case in the Cartesian plane in terms of its *QoI* and *DoC*.

4 Results and Discussion

4.1 Sample Characterization

A total of 148 children were enrolled in this study with an age average of 8.6 years, ranging from 5 to 10 years old. The gender distribution was 59.2% and 40.8% for female and male, respectively.

4.2 Case-Based Approach

Contrasting with other problem solving methodologies (e.g., those that use Decision Trees or Artificial Neural Networks), relatively little work is done offline [14]. Undeniably, in almost all the situations, the work is performed at query time. The main difference between this new approach and the typical CBR one relies on the fact that all the cases have their arguments set to the interval [0, 1] having in consideration their domains. On the one hand, this stands for a new approach to Speculative Computation [15, 16], understood as an estimate of one's confidence that an unknown value of a given attribute falls within a given range. Consequently, it is irrelevant if all attributes are independent or not. On the other hand, the Case Base will be given in terms of triples in the form:

$$Case = \{< Raw_{data}, Normalized_{data}, Description_{data} >\}$$

where Raw_{data} and $Normalized_{data}$ stand for themselves, and $Description_{data}$ is made on a set of strings or even free text, which may be analyzed with String Similarity Algorithms.

When confronted with a new case, the system is able to retrieve all cases that meet such a structure and optimize such a population, i.e., it considers the attributes *DoC*'s value of each case or of their optimized counterparts when analysing similarities among them. Thus, under the occurrence of a new case, the goal is to find similar cases in the *Case Base*. Having this in mind, the algorithm given in [9, 16] is applied to a new case that presents feature vector ($Age = 6$, $L_{evels}C_{ompleted} = \bot$, $Min_{imum\ Score} = 30$, $Ave_{rage\ Score} = 150$, $Max_{imum\ Score} = 240$, L_{evel} 1 $R_{esponse\ Time} = [12, 20]$, L_{evel} 2 $R_{esponse\ Time} = [11, 18]$, L_{evel} 3 $R_{esponse\ Time} = [6, 15]$, $U_{nderstanding}$ $D_{ifficulties} = 0$, D_{oing} $D_{ifficulties} = 0$, *Description* = *Description new*), with the results:

$$\underbrace{dys_{diag_{new}}\big((1,1),(1,0),\cdots,(1,0.97),(1,0.89),\cdots,(1,1),(1,1)\big) \ :: 1 :: 0.88}_{new\ case}$$

Thus, the *new case* can be depicted on the Cartesian plane in terms of its *QoI* and *DoC*, and through clustering techniques, it is feasible to identify the clusters that intermingle with the new one (symbolized as a star in Fig. 5). The *new case* is compared with every retrieved case from the cluster using a similarity function *sim*, given in terms of the average of the modulus of the arithmetic difference between the arguments of each case of the selected cluster and those of the *new case* (once *Description* stands for free text, its analysis is excluded at this stage). Thus, one may have:

$$\underbrace{\begin{array}{l} dys_{diag_1}\big((1,1),(1,0.98),\ \cdots,\ (1,1),(1,1),\cdots,\ (1,1),(1,0)\big) :: 1 :: 0.84 \\ dys_{diag_2}\big((1,1),(1,1),\ \cdots,\ (1,0),(1,1),\cdots,\ (1,1),(1,0.95)\big) :: 1 :: 0.89 \\ \qquad\qquad\qquad\qquad\vdots \\ dys_{diag_j}\big((1,1),(1,0.92),\ \cdots,\ (1,0),(1,0),\cdots,\ (1,1),(1,0)\big) :: 1 :: 0.72 \end{array}}_{normalized\ cases\ from\ retrieved\ cluster}$$

Assuming that every attribute has equal weight, the dissimilarity between $dys_{diag_{new}}^{DoC}$ and the $dys_{diag_1}^{DoC}$, i.e., $dys_{diag_{new\to1}}^{DoC}$, may be computed as follows:

$$dys_{diag_{new\to1}}^{DoC} = \frac{\|1-1\| + \|0-0.98\| + \|1-1\| + \cdots + \|1-0\|}{10} = 0.17$$

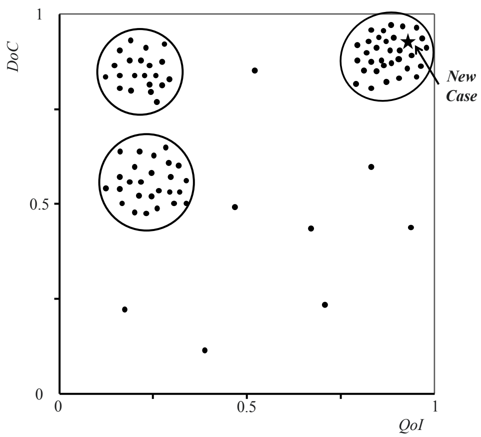


Fig. 5. A case's set split into clusters.

Thus, the similarity for $dys_{diag}{}_{new \to 1}^{DoC}$ is 1 – 0.17 = 0.83. Regarding *QoI* the procedure is similar, returning $dys_{diag}{}_{new \to 1}^{QoI} = 1$.

With these similarity values it is possible to get a global similarity measure, i.e.,

$$dys_{diag}{}_{new \to 1}^{QoI, DoC} = \frac{1 + 0.83}{2} = 0.92$$

These procedures should be applied to the remaining cases of the retrieved cluster in order to obtain the most similar ones, which may stand for the possible solutions to the new problem.

5 Conclusions

This work presents an intelligent decision support system to estimate children evidences of dyscalculia based on the use of *android app disMAT*. It is centred on a formal framework based on *LP* for *Knowledge Representation and Reasoning*, complemented with a *CBR* approach to problem solving that caters for the handling of incomplete, unknown, or even contradictory information. Under this approach the cases' retrieval and optimization phases were heightened and the time spent on those tasks shortened in 12.3%, when compared with existing systems. On the other hand, the overall accuracy was around 86.4%. Additionally, under this approach the users may define the weights of the cases' attributes on-the-fly, letting them to choose the most appropriate strategy to address the problem (i.e., it gives the user the possibility to narrow the search space for similar cases at runtime).

Acknowledgments. This work has been supported by COMPETE: POCI-01-0145-FEDER-007043 and FCT – Fundação para a Ciência e Tecnologia within the Project Scope: UID/CEC/00319/2013.

References

1. Ferraz, F., Neves J.: A brief look into dyscalculia and supportive tools. In Proceedings of the 5th IEEE International Conference on E-Health and Bioengineering (EHB 2015), pp. 1–4, IEEE Edition (2015)
2. Berch, D., Mazzocco, M.: Why Is Math So Hard for Some Children? The Nature and Origins of Mathematical Learning Difficulties and Disabilities. Paul H. Brookes Publishing Co., Baltimore (2007)
3. Kosc, L.: Developmental Dyscalculia. Journal of Learning Disabilities 7, 164–177 (1974)
4. Romagnoli, G.: Dyscalculia: A Challenge in Mathematics. CRDA, São Paulo (2008)
5. Barkley, R.: Guidelines for Defining Hyperactivity in Children." In Lahey, B., Alan E. Kazdin, A. (eds.) Advances in Clinical Child Psychology, vol. 5, pp. 137–180, Springer, New York, (1982)
6. Rubinsten, O., Henik, A.: Developmental Dyscalculia: heterogeneity might not mean different mechanisms. Trends in Cognitive Sciences 13, 92–99 (2008)
7. Aamodt, A., Plaza, E.: Case-based reasoning: Foundational issues, methodological variations, and system approaches. AI Communications 7, 39–59 (1994)
8. Richter, M.M., Weber, R.O.: Case-Based Reasoning: A Textbook. Springer, Berlin (2013)
9. Neves J., Analide, C., Fernandes, B., Freitas, M., Vicente, H.: A Logic Programming approach to Case-Based Reasoning (In preparation)
10. Kakas, A., Kowalski, R., Toni, F.: The role of abduction in logic programming. In: Gabbay, D., Hogger, C., Robinson, I. (eds.) Handbook of Logic in Artificial Intelligence and Logic Programming, vol. 5, pp. 235–324. Oxford University Press, Oxford (1998)
11. Pereira, L., Anh, H.: Evolution prospection. In: Nakamatsu, K. (ed.) New Advances in Intelligent Decision Technologies – Results of the First KES International Symposium IDT 2009, Studies in Computational Intelligence, vol. 199, pp. 51–64. Springer, Berlin (2009)
12. Neves, J.: A logic interpreter to handle time and negation in logic databases. In: Muller, R., Pottmyer, J. (eds.) Proceedings of the 1984 annual conference of the ACM on the 5th Generation Challenge, pp. 50–54. Association for Computing Machinery, New York (1984)
13. Neves, J., Machado, J., Analide, C., Abelha, A., Brito, L.: The halt condition in genetic programming. In: Neves, J., Santos, M.F., Machado, J. (eds.) Progress in Artificial Intelligence. LNAI, vol. 4874, pp. 160–169. Springer, Berlin (2007)
14. Machado J., Abelha A., Novais P., Neves J., Neves J.: Quality of service in healthcare units. In Bertelle, C., Ayesh, A. (eds.) Proceedings of the ESM 2008, pp. 291–298. Eurosis – ETI Publication, Ghent (2008)
15. Lucas, P.: Quality checking of medical guidelines through logical abduction. In: Coenen, F., Preece, A., Mackintosh A. (eds) Proceedings of AI-2003 (Research and Developments in Intelligent Systems XX), pp. 309–321. Springer, London (2003)
16. Fernandes, F., Vicente, H., Abelha, A., Machado, J., Novais, P., Neves J.: Artificial Neural Networks in Diabetes Control. In Proceedings of the 2015 Science and Information Conference (SAI 2015), pp. 362–370, IEEE Edition (2015)
17. Carneiro, D., Novais, P., Andrade, F., Zeleznikow, J., Neves, J.: Using Case-Based Reasoning and Principled Negotiation to provide decision support for dispute resolution. Knowledge and Information Systems 36, 789–826 (2013)
18. Ping, X.-O., Tseng, Y.-J., Lin, Y.-P., Chiu H.-J., Feipei Lai, F., Liang J.-D., Huang, G.-T., Yang, P.-M.: A multiple measurements case-based reasoning method for predicting recurrent status of liver cancer patients. Computers in Industry 69, 12–21 (2015)

Using Mobile Game to Enhance the Learning Motivation and Performance in Higher Education

Kannika Daungcharone

College of Arts Media and Technology, Chiang Mai University,
239 Huaykeaw Rd, Suthep, Muang, Chiang Mai, Thailand, 50200
E-mail: kannika.d@cmu.ac.th

Abstract. The changes of technology are affecting today's lifestyle. Mobile applications are widely used in daily life, business function, and education. There also has been an emerging of developing educational models which suit to new technologies. Therefore, learners and lecturer in this era in particular those who involved with education institutions and related organizations need to concern. This paper presents an idea of developing the teaching and learning model in higher education for improving learners' motivation and performance by using mobile game since mobile technologies is equipped with advance features and most learners have and use them on a regular basis. It would be useful if learners can use mobile technologies as a tool for helping them to be more educated.

1 Introduction

Nowadays, the modern technology, in particular computer and smartphone, can be seen as a part of students' life since they have some experiences with computer games, search engines, social networks, text processors, instant messaging and etc. Therefore, educational institutions and related organizations need to be concerned with the novel educational models that should be developed to suit with the new lifestyle of learners. As learning in higher education requests students to study in-depth of theories so as to apply them to real problems, modern technology can simplify various difficult theories and abstract idea which learners feel hard to understand.

In essence, it is now accepted the educational system attempt to move further into the 21st century by introducing a new set of cognitive and non-cognitive skills considered to be principally important for success in academic contexts as well as on the students' job. The skills could enclose with primarily higher-order thinking skills and complex cognitive processes. They could relate creativity, problem solving, and information, communication technology (ICT) literacy, and etc. Such skill is very important issues considered by various researchers and policy makers. [1]. Consequently, almost institutions have attempted to encourage their

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_24

lecturers to modify the learning style by introducing the students to build self-learning process in order to improve their ability to understand, remember, and be applied effectively.

Among various methods and techniques, some lecturers introduce game plays as a good role for learning style in 21st century as game impact effectively on the performance and motivation for studying. Regarding the research conducted by Mazeyanti [2], Game-Based Learning (GBL) is retrieved 162 million hits on a Google search and 2.37 million documents, categorized as scholarly papers, on a Google Scholar search. This information confirms that GBL continues to obtain large interest and attention from lecturers as well as practitioners including educators, commerce, the military and health services. As the increasing of Net generation, GBL can be seen as an alternative learning and training style to motive and train for student in this generation.

The purpose of this paper is to represent the involved factors supporting mobile games that can enhance the learning motivation and performance in higher education. In particular, this novel idea will be specifically applied to the computer programming subject for developing a suitable tool to support students to understand the difficult contents.

2 Literature Review

This paper is stimulated by various theories and researches in order to describe and support the concept. The main concept is categorized by employing three main matters as shown below:

2.1 Motivation and Performance of Learning

As technology plays a good role in terms of changing the learning style for this day, the traditional learning style would be disappeared soon. It is predicted that new techniques and modern tools would be the main tools for supporting learning style for the future.

Particularly, games are now used to be as techniques and tools in classroom for augmenting the interest of learners in nowadays since most designed games for educational purposes always create player enjoyment. Based on Michail's work [3], educational related games effect students in three main areas including learners' entertainment (enjoyment), learners' acceptance (intension to use), and learner's emotion (happiness) which the encourage learners' performance (see Fig. 1)

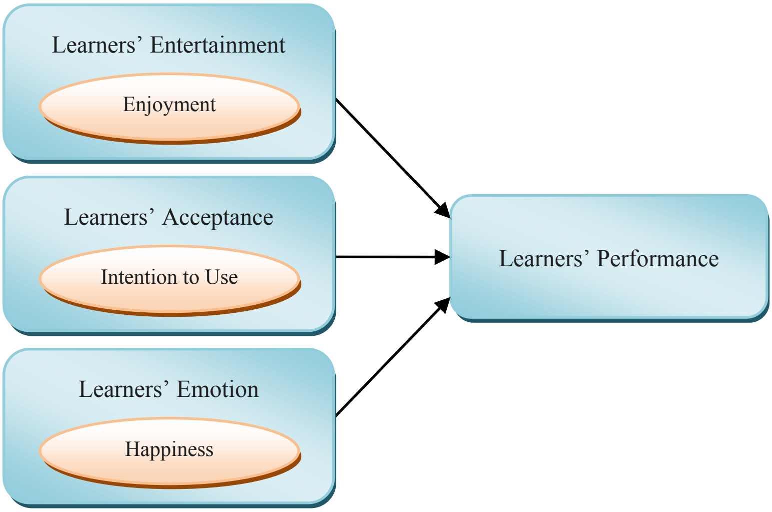


Fig. 1. Four characteristics of game to get the performance of learner [3]

Further, Eric [4] assert the technology acceptance model (TAM) can be adapted to value the degree to which students accepted the educational game by employing the four constructs: perceived usefulness, perceived ease-of-use, attitude towards usage, and intention to use. This issue is also supported by the research conducted by Hainey et al [5]; they conclude that game affect the learners' intrinsic motivation since game contains six individual attributes (see Fig. 2) and the detail of each attribute is explained in the following list:

- Challenge: to create a contest, fight, or competition
- Fantasy: to create imagination that related with responsibilities
- Control: to create power effecting or directing people's behavior
- Curiosity: to create a strong desire to know or learn
- Competition: to create the activity or condition of competing
- Recognition: to create the action or process of recognizing

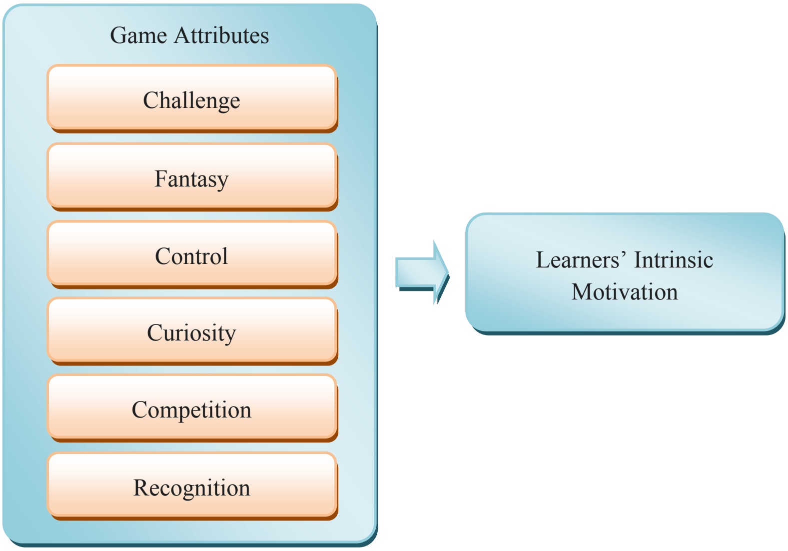


Fig. 2. Six individual attributes of game affect to learners' intrinsic motivation [5]

2.2 Digital Game-based Learning

According to Anissa et al [6], Digital Game-Based Learning (DGBL) can benefit the educational objective since it can entertain the learners. Specifically, DGBL can generate a balance between gaming and learning elements which then produce two main important foundations: fun/entertainment and educational components. As such, many lecturers employ the DGBL to support their studies for both learning and motivation. Regarding the DGBL, it can be categorized into two main types: special purpose games created for educational purpose, and commercial-off-the-shelf games created for entertainment purposes, that are being deployed in an educational content.

Besides, Lynceo [7] asserts that the use of educational games should be employed by using different learning theories. Indeed, the educational game should be based on behavioral concept by perceiving learning as an associative process; this then can generate an important role in changing observed behavior. This conception of learning is confirmed by gamers seeking to exercise concepts and/or abilities by repetitive practice. However, games based on constructivist conceptions shows other arguments; this means that games should have the emphasized aspects focusing on the importance of practical and real experience in the construction of knowledge by employing the Theory of Experiential Learning. In other words, games must have interactive revealing the effects of particular actions allowing new strategies to be established and tested. Further, games should produce a safe environment for experimentation which then the consequences are not transferred to the real world. Regarding the recent computational progress, the games are now accomplished with providing new involvements that are rich, complex, equipped with interactivity, and comparable to real life. These experiences can be seen as a source for the construction of novel knowledge, which is not simply transmitted but is obtained as a result of reflection within such environment.

Based on the research conducted by Eric [4], the way to develop a suitable DGBL for learners is to adapt the Bloom's Taxonomy established by Anderson and Krathwohl. Actually, this concept is used for scoping the design standard for the educational game. Specifically, the knowledge dimension of such taxonomy can be divided into four levels from concrete to abstract: factual knowledge, conceptual knowledge, procedural knowledge, and meta-cognitive knowledge. In addition, cognitive processes can be categorized into six main levels ranking from low complexity to high: remember, understand, apply, analyze, evaluate, and create. There are represent in Fig. 3.

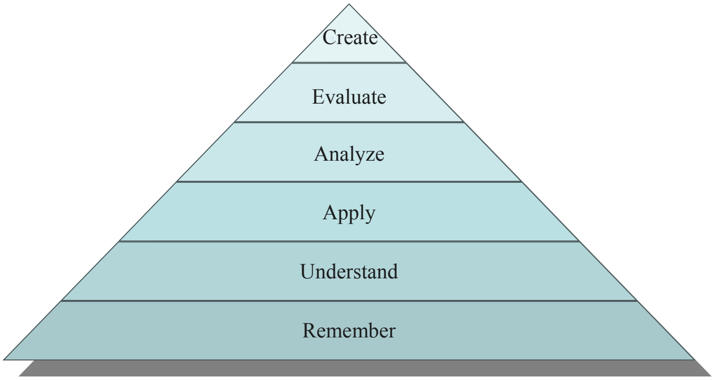


Fig. 3. The cognitive level of Bloom's Taxonomy theory [4].

2.3 Mobile Games

Mobile technology, such as Tablet and Smartphone, has become popular device for new generation globally. The mobile technology has been used by all level of students in classrooms. As far as the success of Smartphone and Tablet computers is concerned, it generates a significant growth of the Internet applications which specially developed for those devices. Also, the mobile devices are now affordable and functional; this then has been seen as partly explanation for the attractiveness of mobile devices in education. While there are various positive results of using mobile technology on learning, it still absences of appropriate guidelines for new curricula and pedagogy for supporting and assessing the use of mobile technology in educational institutes [8].

Besides, Birgit [9] asserts that mobile devices being used to engage learners and to support learning have been widely acknowledged by several educational practitioners and scientists. This means that an extensive range of practical studies across different areas and application scenarios has confirmed the usefulness for these technologies in terms of the process of teaching as well as learning. Indeed, mobile devices can be employed for several learning contexts and their functions are used in many ways. One of various formats frequently used for taking advantage of the mobile technology is using mobile games for learning. Specifically, mobile games can be applied for both commercial and scientific use since they have been specifically developed for several target groups and learning environments.

As Yao-Ting [10] stated, mobile devices also have gradually been presented into educational contexts over the past 2 decades. This mobile technology has bought most people to own their individual small technology devices containing their own computing power, such as laptops, personal digital assistants (PDAs), tablet personal computers (PCs), cell phones, e-book readers and etc. These portable devices also integrate with wireless communication function allowing one-to-one learning tool which can be considered as the great potential tools for learning for both traditional classrooms and outdoor informal education. In terms of promoting innovation in education by employing information technology, mobile computing should support not only traditional lecture-style teaching, but also throughout convenient information gathering and sharing. This then can encourage innovative teaching methods which could be cooperative learning, exploratory learning outside the classroom, game based learning and etc. As far as mobile technologies are concerned, it generates great potential to enable more innovative educational methods which not only help subject content learning, but may also simplify the development of communication, problem-solving, creativity, and other high-level skills among students.

Likewise, Janet [11] declares that mobile devices have become progressively popular with our daily lives. New version of these devices integrate with advanced features that make people more convenient; these features are also more affordable and continually become available making our lives easier. As such, these advances technologies have influenced educators and researchers to utilize these devices to support their teaching and learning. One benefit for educational purpose using these mobile devices is to transform how student learn by changing the traditional classroom environment into environment being more interactive and engaging. It also enables lecturers to conduct their teaching without being restricted by time and place, enable students to learn after class over or outside the classroom, gives lecturers ability to connect with their students on a more personal level by using devices that they use on a regular basis. In this case, mobile learning generates highly motivation for students since mobile environments positively affect on learning outcome by supporting them to improve their understandings. Therefore, mobile applications used for learning must be consist of six features including technology-based scaffolding, location-aware functionality, visual/audio representations, digital knowledge-construction tools, digital knowledge-sharing mechanisms and differentiated roles.

3 Conceptual Framework

From various literature reviews which are presented above, there are related and supported issues that mobile games base on DGBL theory can increase the learners' motivation and performance in learning. Those previous concepts can be integrated and then constructed to be a novel conceptual for framework for using mobile game to enhance the learning motivation and performance in higher education which is shown in Fig. 4.

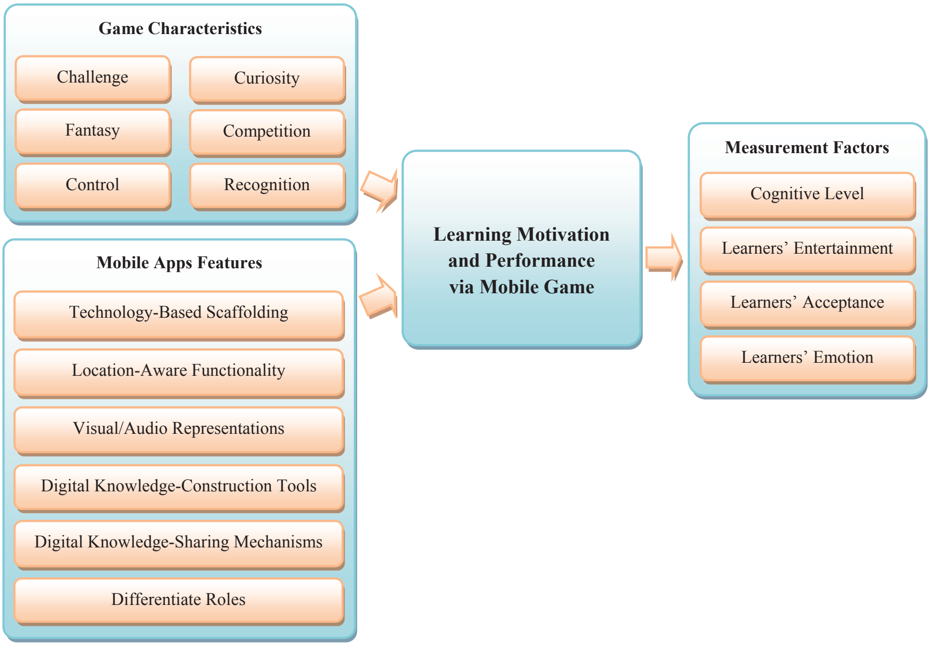


Fig. 4. The conceptual framework.

4 Conclusion

The conceptual framework in Fig. 4 clarifies an idea that using mobile game as a tool for higher education can raise the learners' motivation and performance. Specifically, the attributes enhancing such motivation and performance relies on mobile game applications in which this should include the features of challenge, fantasy, control, curiosity, competition, recognition, technology-based scaffolding, location-aware functionality, visual/audio representations, digital knowledge construction tools, digital knowledge-sharing mechanisms, and differentiate roles. For measure the learning motivation and performance via mobile games, this can be completed by measuring the perspective of learners toward learners' cognitive level, entertainment, acceptance and emotion.

References

[1] Samuel Greiff, Christoph Niepel and Sascha Wustenberg. 21st century skills: International advancements and recent developments. Thinking Skills and Creativity 18 (2015) 1-3.

[2] Mazeyanti Mohd Ariffin, Alan Oxley and Suziah Sulaiman. Evaluating Game-Based Learning Effectiveness in Higher Education. Social and Behavioral Sciences 123 (2014) 20–27.

[3] Michail N. Giannakos. Enjoy and learn with educational games: Examining factors affecting learning performance. Computers & Education 68 (2013) 429–439.

[4] Eric Zhi Feng Liu and Po-Kuang Chen. The effect of Game-Based Learning on Student's Learning Performance in Science Learning – A case of Conveyance Go. Social and Behavioral Sciences 103 (2013) 1044-1051.

[5] Thomas Hainey, Wim Westera, Thomas M. Connolly, Liz Boyle, Gavin Baxter, Richard B. Beeby and Mario Soflno. Students' attitudes toward playing games and using games in education: comparing Scotland and the Netherlands. Computer & Education 69 (2013) 474-484.

[6] Anissa All, Elena Particia Nunez Castellar and Jan Van Looy. Assessing the effectiveness of digital game-based learning: Best practices. Computer & Education 92-93 (2016) 90-103.

[7] Lynceo Falavigna Braghirolli, Jose luis Duarte Ribeiro, ndreas Dittmar Weise and Morgana Pizzolato. Benefits of educational games as an introductory activity in industrial engineering education. Computers in Human Behavior 58 (2016) 315-324.

[8] Marta Gomez Domingo and Antoni Badia Gargante. Exploring the use of educational technology in primary education: Teachers' perception of mobile technology learning impacts and applications' use the classroom. Computer in Human Behavior 56 (2016) 21-28.

[9] Birgit Schmitz, Roland Klemke, Jaap Walhout and Marcus Specht. Attuning a mobile simulation game for school children using a design base research approach. Computer & Education 81 (2015) 35-48.

[10] Yao-Ting Sung, Kuo-En Chang and Tzu-Chien Liu. The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. Computer & Education 94 (2016) 252-275.

[11] Janet Mannheimer Zydney and Zachary Warner. Mobile apps for science learning: Review of research. Computers & Education 94 (2016) 1-17.

Pre-extraction of Features and Environment Variable-based Database Filtering for Fast Image Matching on Mobile

Sung Soo Hwang, Sung Ho Hong

School of Computer Science & Electronic Engineering, Handong Global University, Republic of Korea
sshwang@handong.edu, 21100778@handong.edu

Abstract. Image matching refers to a technique which finds the best matching image from the database to an input image. Currently, feature-based image matching is widely used, and several features have been proposed to reduce the computational cost of image matching. This paper presents a method to further reduce the computational cost of image matching so that image matching can be applicable to mobile services. To this end, instead of storing images to manage the database, we extract features from images in the database beforehand and keep the features. In this way, the computational cost of the actual image matching is reduced since feature extraction is performed on the input image only. Furthermore, we manage the environment variables of each photographing. As most of the current mobile phones have GPS sensor and magnetometer, we utilize these variables to narrow down the matching candidates in the database. The proposed method was applied to outdoor localization, and simulation results show that the proposed method reduced the computational cost for image matching.

Keywords: Image matching, Feature Descriptors, Pre-extraction, GPS, magnetometer, Database filtering

1 Introduction

Image matching refers to a technique that finds the best matching image among a variety of images to an input image. Image matching is very important and it has been a fundamental technique to various applications. In 3D reconstruction, where a 3D model of an object or scene is generated from multiple images, image matching is utilized to generate 3D points from input images [1, 2]. For panoramic image generation, image matching is utilized to align input images [3, 4]. Image matching can also be utilized to several mobile services. One example is indoor/outdoor localization which provides the location of users by comparing the image the user currently takes with the database [5, 6].

Among various ways of image matching, feature-based image matching is widely used. In feature-based image matching, an image is represented with several feature

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_25

points inside the image (e.g. edges, lines, etc.), and the similarity between images is computed by the similarity between the extracted feature points. Since extracting and describing feature points are essential in feature-based image matching, a variety of researches have been conducted [7-11]. As the results of these researches, several features such as SIFT [7], SURF [8], BRIEF [9] has been proposed and widely used. Those features not only consider the accuracy of image matching but also the computational cost, since applications that utilize image matching normally contain lots of image and they are required to conduct the image matching process repetitively.

Even though previous works made an effort to reduce the computational cost of image matching, it is still computationally expensive for mobile services. Specifically, in case of providing localization services on a mobile phone, the increase of speed on image matching is crucial because a large number of image matching should be conducted. As described in Table 1, it takes about 205 seconds to perform a total of 170 image matching process even on the iPhone 6 which is known to have good specification. As one research reports that human beings expect the buffering of contents less than 2 seconds [12], it is important to reduce the computational cost of image matching to provide mobile services.

In this paper, we propose a method which reduces the computational cost of image matching so that image matching can be applicable to mobile services. The proposed method consists of two methods; pre-extraction of features and database filtering by environment variables. For pre-extraction of features, we extract feature descriptors from images in the database beforehand and keep the feature descriptors. In this way, the computational cost of the actual image matching is reduced since feature extraction is performed on the input image only. For database filtering, we manage the environment variables of each photographing. And by using these environment variables, we narrow down the matching candidates in the database.

This paper is organized as follows: In section 2, the proposed method is explained. In section 3, the proposed method is evaluated by several images and we conclude in section 4.

2 Proposed method

2.1 System overview

Fig. 1 illustrates the concept of a mobile application to which the proposed method is applied. The application assumes that users want to identify the building they are currently facing and it provides the information of the building once users take a photo of that building. The mobile application also assumes that users are moving with their mobile phones and mobile phones of users have GPS sensor and magnetometer. To implement this functionality, a large number of images of each building are taken beforehand and the image that users currently take is compared with the images in the database.

As there are many images in the database, a large number of image matching should be conducted. To reduce the computational cost of image matching process, pre-extraction of features and data filtering are performed as Fig. 2 illustrates. Before

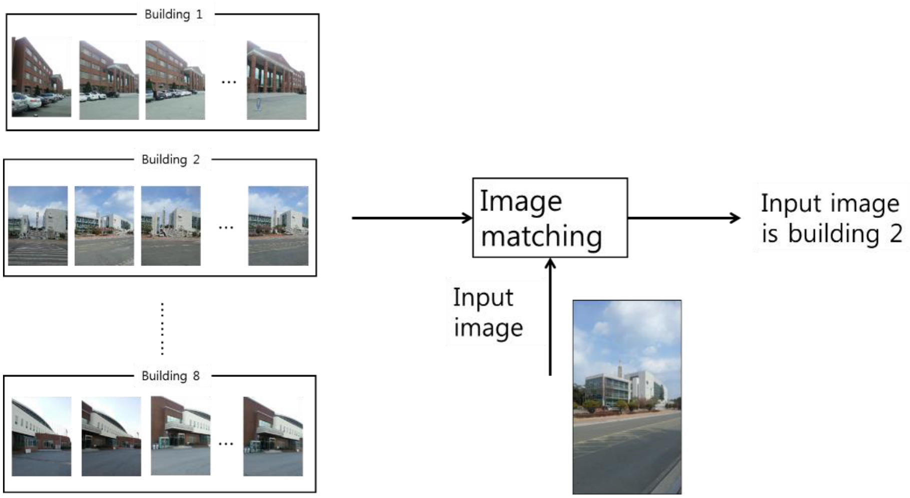


Fig. 1. The application to which the proposed method is applied

the actual image matching, instead of storing images to manage the database, we extract features from images in the database beforehand and keep the feature along with the values of GPS sensor and magnetometer of each photographing. And when the actual image matching process is necessary, the database is filtered out according to the values of GPS sensor and magnetometer of users' mobile phones. Lastly, the filtered features of images are compared with the features extracted from user's image.

2.2 Pre-extraction of features

In this paper, SURF is utilized as features for image matching. SURF is a speeded-up version of SIFT and it is widely used. Compared to SIFT which approximated Laplacian of Gaussian (LoG) with Difference of Gaussian for finding scale-space, SURF approximates LoG with Box Filter. In this way, convolution with box filter can be easily calculated with the help of integral images. And it can be done in parallel for different scales. And for orientation assignment, SURF uses wavelet responses in horizontal and vertical direction for a neighborhood. The dominant orientation is estimated by calculating the sum of all responses within a sliding orientation window of angle 60 degrees, and wavelet response can also be found out using integral images easily.

For feature description, SURF uses wavelet responses in horizontal and vertical direction. A neighborhood is taken around the keypoint, and it is divided into 4x4 subregions. For each subregion, horizontal and vertical wavelet responses are taken and a vector is formed as (1).

$$\mathrm{v} = \left(\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|\right) \quad (1)$$

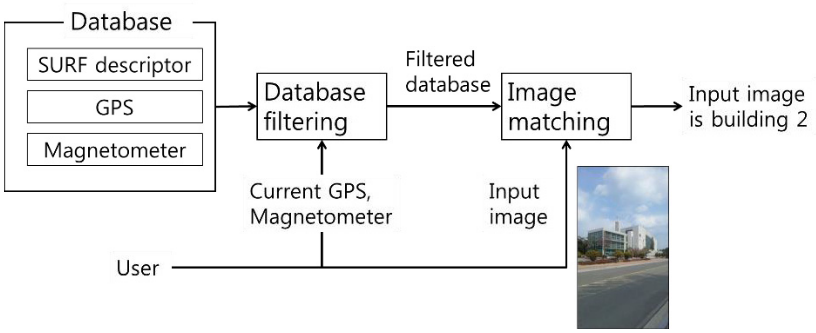


Fig. 2. The overall system of the proposed method

where d_x is wavelet response in horizontal direction. Note that each element in a vector is a floating-point value.

This paper suggest the management of database by feature points of images. Hence, when an image contains a number of 100 features, the database maintains 6400 floating values for the image. By pre-extraction of features, the computation time for feature extraction can be reduced.

2.3 Data filtering

We can additionally reduce the computation time for image matching by reducing the number of candidates in the database. For the accuracy of image-based localization, a large number of images should be taken and managed in the database. However, for images that were taken far from the user's current location, we do not need to perform image matching. Furthermore, even though some images were taken near to the user's current location, if viewing direction is completely different, image matching with those images is unnecessary.

As current mobile phones contain GPS sensor and magnetometer, we can reduce the number of candidates by utilizing these data. Fig. 3 illustrates the decrease of the candidates by using GPS sensor and magnetometer.

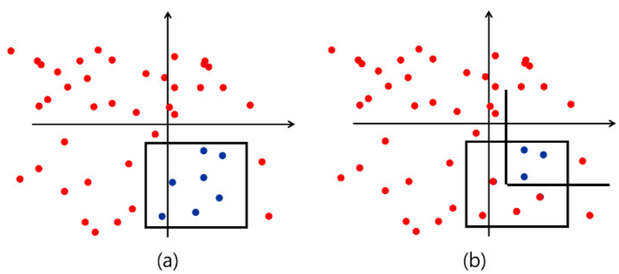


Fig. 3. The decrease of the number of candidates by GPS sensor and magnetometer. (a)By using GPS sensor (b) By using both GPS sensor and magnetometer

3 Experimental results

For the evaluation of the proposed method, we created a database for a total number of 170 images. The data for each image consists of SURF descriptors extracted in the image, and GPS/magnetometer data for the capturing environment of the image. Fig. 4 shows several examples of the created database, and the database was downloaded to the test mobile phone before the simulation. The simulation was conducted by using iPhone 6, and we tested the proposed method by performing image matching 8 times. For comparison, we compared the computation time for image matching when a) the database is maintained by images, b) the database is maintained by SURF descriptors of images without any environment variables, c) the database is maintained by SURF descriptors of images with GPS data, and d) the database is maintained by SURF descriptors of images with GPS and magnetometer data. For image matching, FLANN-based matching [13] implemented in openCV was utilized.

Table 1 describes the computation time for image matching of each cases. As Table 1 indicates, pre-computation of features and database filtering bring out the drastic decrease of the computational cost for image matching. It decreases the computation

Fig. 4. Several images in the test database

Table 1. Computation time for image matching with different database maintanent and data filtering(unit:seconds)

	Database by images	Database by descriptors	Database by descriptors with GPS	Proposed method
Test 1	208.21	21.56	2.61	1.21
Test 2	203.73	23.60	3.19	1.35
Test 3	204.26	22.17	2.85	1.42
Test 4	203.79	26.27	2.95	1.23
Test 5	203.90	22.49	2.91	1.13
Test 6	206.54	23.51	1.52	1.22
Test 7	206.38	18.55	1.77	1.11
Test 8	203.00	18.44	0.63	0.43
Average	204.98	21.95	2.30	1.14

time for image matching by a factor of 10 compared to the case when the database is maintained by images. The utilization of GPS data also decreases the computation time by a factor of 10, and the utilization of both GPS and magnetometer data decreases the computation time by a factor of 2 compared to the case when GPS data is utilized for database filtering. By utilizing the proposed method, localization can be performed less than 2 seconds, and hence, we can conclude that the application implemented by the proposed method the proposed method can satisfy user requirements [12].

The another advantage of using the proposed method is that not only it reduces the computational cost for image matching, but also it enhances the accuracy of localization. Fig. 4 shows the matching results when the database is maintained by the pre-mentioned 4 cases. This is because the proposed method filters out unnecessary images to be matched by environment variables, and as a result, it can prevent undesirable matching results.

The disadvantage of the proposed method is that, as Table 2 describes, it requires larger amount of storage than maintaining the database by images. However, this can solved by managing the database at the server and performing image matching by sending the user's image to server.

Table 2. The amount of storage required for manaing the database (unit: MB)

Database by images	Database by descriptors	Database by descriptors with GPS	Proposed method
31.2	308.01	308.09	308.10

4 Conclusion

This paper presents a method to reduce the computational cost of image matching so that image matching can be applicable to mobile services. By managing feature data extracted from images along with GPS/magnetometer data, the computational

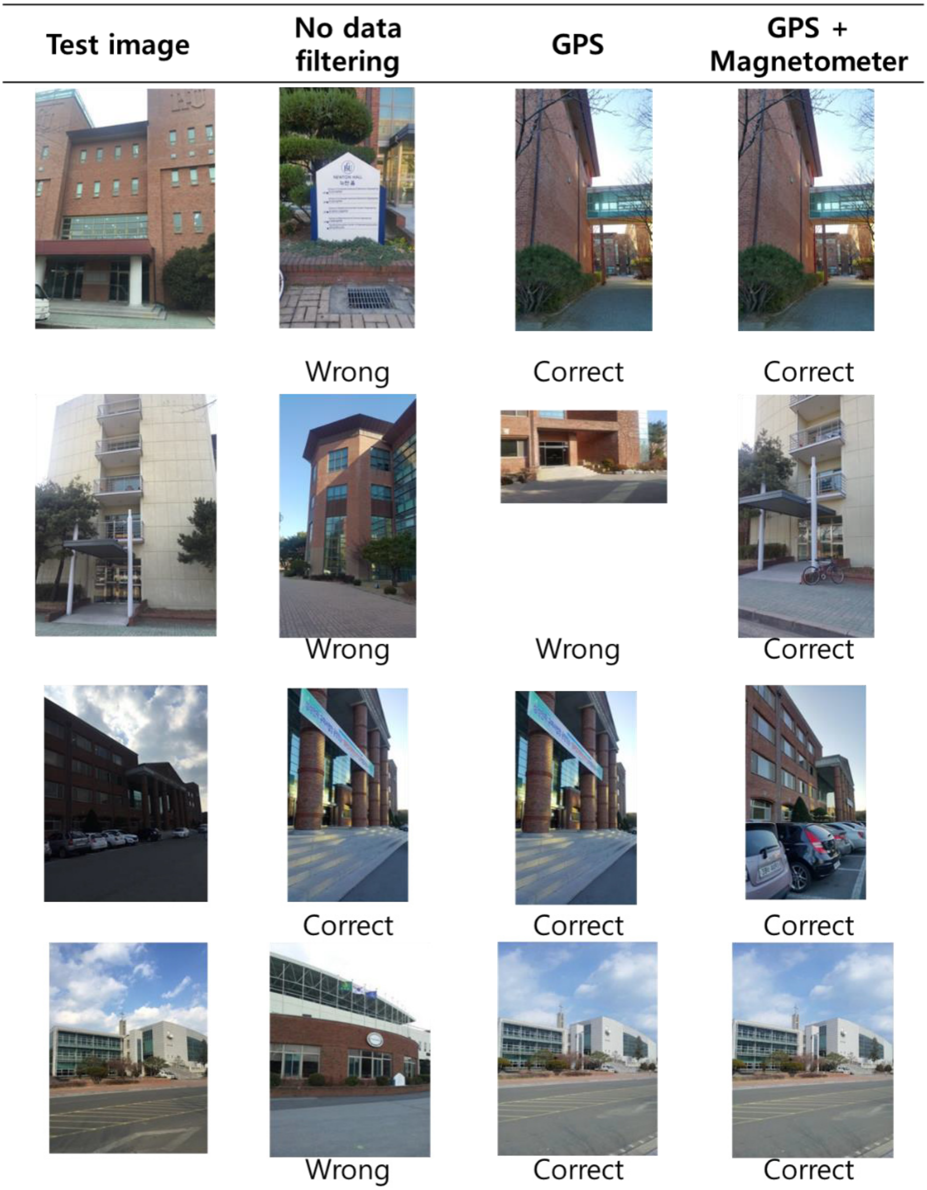


Fig. 5. Matching results with different database filtering

cost of the actual image matching is reduced. Furthermore, database filtering based on GPS/magnetometer data increased the accuracy of the overall system.

As future works, memory-efficient image matching by managing the database at the server performing image matching at the server by transmitting users' data will be studied.

Reference

1. M.P. Segundo, L. Gomes, and O.R.P. Bellon, "Automating 3D Reconstruction Pipeline by SURF-based Alignment," in *Proc. IEEE Image Processing*, pp.1761-1764, 2012.
2. M. Brown, and D.G. Lowe, "Unsupervised 3D Object Recognition and Reconstruction in Unordered Datasets," in *Proc. 3-D Digital Imaging and Modeling*, pp.56-63, 2005.
3. L. Juan, and O. Gwun, "SURF applied in Panorama Image Stitching," in *proc. Image Processing Theory and Applications*, pp.495-499, 2010.
4. Y. Li, Y. Wang, W. Huang, and Z. Zhang, "Automatic Image Stitching using SIFT," in *Proc. Audio, Language and Image Processing*, pp.568-571, 2008.
5. S. Treuillet, and E. Royer, "Outdoor/indoor Vision Based Localization for Blind Pedestrian Navigation Assistance," *International Journal of Image and Graphics*, World Scientific Publishing, 2010, pp.481-496.
6. H. Kang, A.A. Efros, M. Hebert, and T. Kanade, "Image Matching in Large Scale Indoor Environment," in *Proc. IEEE Computer Vision and Pattern Recognition*, pp.33-40, 2009.
7. D.G. Lowe, "Object Recognition from Local Scale-invariant Features," in *Proc. IEEE International Conference on Computer Vision*, pp.1150-1157, 1999.
8. H. Bay, T. Tuytelaars, and L.V. Gool, "SURF: Speeded Up Robust Features," in *Proc. ECCV*, pp.404-417, 2006.
9. M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary Robust Independent Elementary Features," in *Proc. ECCV*, pp.778-792, 2010.
10. E. Tola, V. Lepetit, and P. Fua, "DAISY: An Efficient Dense Descriptor Applied to Wide-baseline Stereo," *IEEE Trans. Pattern Anal Machine Intelligence*, pp. 815-830, 2010.
11. S. Leutenegger, M. Chli, and R.Y. Siegwart, "BRISK: Binary Robust invariant Scalable Keypoints," in *IEEE International Conference on Computer Vision*, pp.2548-2555, 2011.
12. http://www4.akamai.com/html/about/press/releases/2009/press_091409.html
13. M. Muja, and D. G. Lowe, "Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration," in International Conference on Computer Vision Theory and Applications, pp. 331-340, 2009.

Mobile Sample Allocation and Management System for BioBank

Anupat Maturongpituck[1], Natchanon Bunpetch[1], Anon Tatvichai[1], Lalita Narupiyakul[1,2]*

[1]Department of Computer Engineering, Faculty of Engineering, Mahidol University
[2]Integrative Computational Biosciece Center, Mahidol University
25/25 Phuttamonton 4 Rd., Salaya, Phuttamonton, Nakhon Pathom 73170 THAILAND
lalita.nar@mahidol.ac.th*

Abstract. The main objective for our development is to systemize and assist the process of storing samples for hospital's lab technicians through our mobile application. The system will keep tracking of all the samples along with included check-in, check-out processes. This is necessary because the current arrangement is complex, inconvenient, and very prone to errors. By developing this application on a mobile platform, it can provide the user mobility and flexibility that boosts performance in their working environment. Moreover, the application keeps data consistent with the use of a full, but light-weighted database. The application will be built using Xcode and its database will be created using SQLite. Furthermore, the application also implements a barcode scanner to further eliminate most human-errors that may happen in the process. The application is tried by lab technicians as our stakeholders. The feedbacks reported that the stakeholders are pleased with the application, and it can assist managing samples in the lab.

Keywords: Bio Sample Allocation / SQLite / Specimen Repository / Biobank

1 Introduction

During the treatment in the hospital, doctors will analyze the lab results on a case-by-case basis by gathering the data from the departments such as Pathological department, which collect the blood sample from patients. The samples come in 2 different sizes; 1.2 ml and 15 ml tubes, and stored in a box and froze in a refrigerator respectively. Staffs are responsible for samples pick-up when requested by doctors. However, one of the problems that arise is the sample arrangement is not properly managed. Poorly organized sample disrupts the workflow of staffs, and could potentially lead to a decreased in number of patients treated. Therefore, this problem is our main motivation for the development of this project. Many hospitals are currently faced with the aforementioned problem, due to the samples are all organized in a manual fashion, staffs spent most of their time on finding specific samples for the lab. Such workflow is inefficient; the time needed to find specific sample will keep increasing as number of sam-

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_26

ple increases. Therefore, our objective is to (i) design the database for the sample collection or specimen banking system, and (ii) develop a low-cost, user-friendly, and precise samples management application for mobile devices that could help the hospital's staffs reduce human errors and save time to organize all of these samples.

2 Background and Related Works

Some related works are introduced in this section. We summary about a related application called Vision Tracker and also review about the QR-code API using in this system.

2.1 Vision Tracker Sample Management Software

Vision Tracker Sample Management [1] is one of the few programs that is designed specifically to keep track of samples in a lab. It does what it states and more, the software can keep track of over 16,000,000 samples with all fields searchable. The program can also be integrated with their Vision Mate barcode reader allowing fast barcode scans to find that sample's location [1]. But that is also its downfall, the software have no other free alternative in reading the barcode, and this software can only work with their company's barcode reader. Being a sample management program only, the program doesn't offer any other services, does not keep sample information and reformat it for doctors to use. A few screenshots of Vision Tracker Management Software can be seen in Fig. 1.

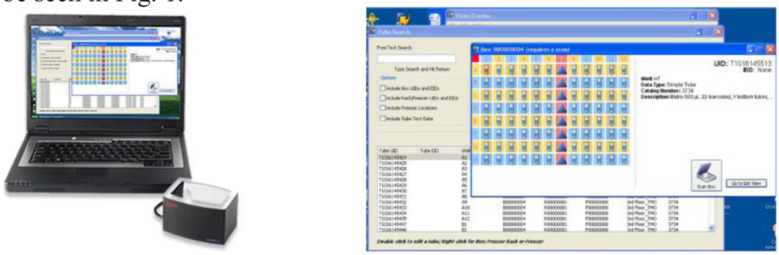


Fig. 1. Vision Tracker and Vison Mate systems [1]

2.2 Scandit

Scandit SDK [2] provides exactly what is needed for this project. It will enable our application to be able to read barcodes quickly and accurately. Scandit is able to recognize many different types of codes, including QR and UPC-A/UPC-E which are the codes that will needed to be scanned. Scandit includes features that will enable it to read low-light, damaged, blurry, or small codes in which will be needed in this project since samples' barcode will be small and hard to read due to the fact that it is frozen and covered in ice. Scandit also works on many platforms including iOS, Android, Titanium, Xamarin and many more.

2.3 Cloud Datastore

The cloud data store such as Dropbox provides four APIs for developers to integrate the Dropbox features to their application [3]. Core API is the simplest, most direct and flexible method to read and write files to Dropbox.

Dropbox for Business API is mainly use for business purpose as it features the user lifecycle management for each account and provides Core API for all members of a team. One of its function is to grant each members the permission to access each individual files. Datastore API is very useful tool to sync the structured data; for examples, contacts and to-do items, as illustrated in Fig 3.

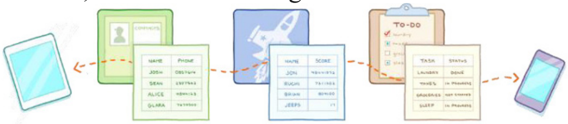


Fig. 2. Data store API for Dropbox [3]

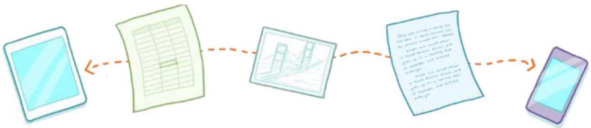


Fig. 3. Syncronized API for Dropbox [3]

3 BioSample Allocation System

According to our purpose system, this project creates a mobile application for the division of Nephrology as our example scenario. Doctors and nurses will be able to utilize a well-organized samples keeping system. The system will allow doctors to flexibly query their patients' samples with multiple constraints and retrieve it with much more convenience. The system, built on a strong database foundation, collects all the necessary information which will help keep the specimen banking organized and samples to be retrieved easily through simplified process. On the user side, staffs who are responsible for registering new biological samples and entering the specimen banking unit will find shortened amount of time needed to keep things organized. To check-in samples, staffs only need to login to an authorized user account and simply scan a barcode present on the sample. The system will read the information and fill the necessary information to check-in the sample while staffs may also add more information on the sample. When the samples are ready to check-in, the system generates an available location in the specimen banking, and the staff can just follow the location posted by the system. To validate that the sample has been checked-in to the location, addition scans of the QR codes at the location of deposit will be required. From there, querying and retrieving samples are made simple by the application. Through strong database structure, simple keywords may be used to query information such as the patientID, projectID, blood type, sample type, or any date checked-in easily. This simplified the task of doctors required to retrieve samples in order to conduct research or analysis of patient's samples. The features of this system are concluded as shown below.

List of features

- Store samples, boxes, and fridges in the system by their IDs and codes.
- Organize and keep track of samples for easier locating.
- Check-In and check-out samples with included approval system.
- Visual representation of where all the objects are at.
- Samples, boxes and racks can be added, deleted, searched, or moved.
- Device's camera can be used to scan codes on the sample and bring the information up on the application.

In this section, the system's processes are separated to 4 main parts. The first part introduces the system architecture while the second section is about the data structure and database design. The next section explains about the workflow process in the system and the last section demonstrated the example scenario.

3.1 Sample Allocation Tracking's System Architecture

In this model, the interface and database are both processed in the application. Within the application, there will be two main processes; the interface process and the database process. The interface process manages the information of the view in which user interacts with and adjust them accordingly. The interface offers users the functionality of viewing, searching, and managing samples in the system. Additionally, the interface will also provide tools which will simplify tasks to user such as the barcode reader. While the interface handles the interaction with user, the database process will regularly manage the local database stored within the application. This includes all the query, move, delete, or insert received from the interface side. Once the interface receives the data from the local database, it displays them in a user-friendly manner. For additional functionality and security, the application will prompted to synchronize the database file through iTunes, iCloud, Dropbox, or any similar services available to keep the information safe. Activities involving changes in the database will be kept in the log table to provide information when rollbacks are necessary.

3.2 Containers' Structures

Our database consisted of 18 tables and this database design mainly focuses on efficiency and ease of use. To minimize working complexity within the application, this database is designed to use as few tables as possible with each table covering all necessary data without compromising simplicity. To achieve our focus, the database is divided into 3 major parts or group of tables as follows: (i) Container information: Fridge, Rack, Box, and Sample, (ii) Environment information: Patient, Doctor, Project, and (iii) History and logs.

Container Levels.
This group of tables is the key component to our system. Most of our application functions revolve around keeping and updating containers consistently. Each table contains the specific properties to each container and their relationship to each other. Additionally, some tables are used to elaborate the details on the containers.

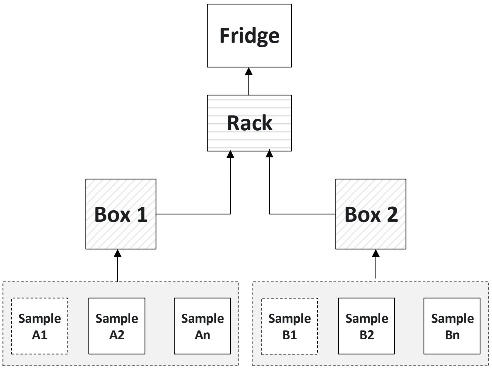


Fig. 4. The container levels

Sample (in Fig 5) table contains information about an individual sample in the system, uniquely identified by the SampleID. The unique properties to an individual sample will be the location of the sample defined by boxRow and boxCol, as well as the follow-up Type of the sample. Box (in Fig. 5) table contains information about an individual box container, which is used to contain multiple samples. Each individual box is uniquely identified by its primary key, the boxID, and further identified by the foreign keys including boxType and rackID. Rack table contains information about an individual rack container used to contain boxes and is uniquely identified by its primary key, rackID. Fridge table contains information about an individual fridge container, which is the largest container in the system, uniquely identified by its primary key, fridgeID.

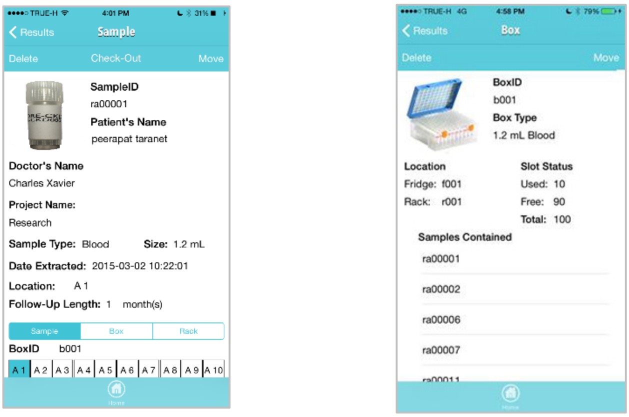


Fig. 5. Examples of containers

3.3 Application's Page Flow

The page flow diagram, in Fig. 6 depicts the flow of interface from one page to another on the application. As seen from the highlighted colors, our application is categorized into 4 main sections including search, manage, check-in/check-out, and history page. Highlighted in a red-tone gradient at the top-most part of the diagram is the search pages. The search main button is located in the main menu of the application while linking to the search option page. The search option page then provides 2 links, search by filter and search by scan. The scan search by scan page leads to the scan page. After a barcode is scanned, it brings user to the item's info page. Similarly, the search by filter page brings user to the filters page. After the filters are applied, it brings user to the item's info page.

In contrast to the search section, the manage section, highlighted in a blue-tone gradient, is more complex and involves more options than the search. From the main menu, the manage button links user to 2 options, move or create new. The move button links to the options to move items currently in storage. The options are by either searching or scanning the item. After choosing an item, it brings user to the move choose destination page, which shows the available spaces in the storage. When the user picks a certain location, it brings user to the confirmation page and returns to main menu. Another manage action is to create a new item. To create any available items the application takes user through the same process. Firstly, the application brings user to that item's create new page and lets the user fill in necessary information. Afterward, the application takes the user to choose location page. After the users choose an available location, and the item is created, the application, again, brings user to the main menu.

Similarly to the manage section, the check-in/check-out section, highlighted in a purple-tone gradient, has the similar process. Whether a user wants to check-in or check-out an item from the storage, the application brings the user to the option to check-in or checkout page. If users want to check-in a sample, the application brings the user to the scan page. After scanning a barcode, the application lets user pick a location to check-in. After confirming with the user, the application will bring the user back to the main menu. Similarly to the check-in, the checkout process lets user scans or searches for an item first, then confirms with the user by scanning the corresponding barcodes then returns to the main menu.

Lastly, the history section, highlighted in a green-tone gradient is the simplest of all the section in this application. When user choose history from the main menu, it brings user to the option to choose from manage history, or check-in/check-out history. The application will then show the responding page from user's choice.

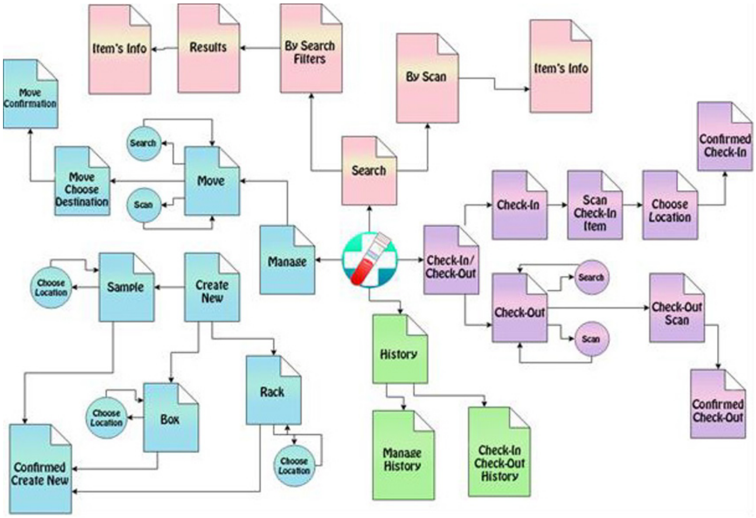


Fig. 6. The application's page flow

3.4 Process Steps

The process steps below is a brief explanation of how each section of the system architecture plays a role in which the main factors in the system architecture including the interface, database, and cloud storage will communicate and work with each other.

Step 1: The application loads and displays the login screen.
Step 2: The login screen lets user enters their username and password.
Step 3: The application checks with the database whether the user exists or/and enters the correct password upon pressing the log in button.
Step 4: After logging in successfully, the application prompts to connect to a backup storage through Internet connection for backup purposes.
Step 5: The application then displays the main menu linking them to various functions with customized user-friendly interface.
Step 6: Upon calling a function from the interface, the system queries and displays data from the local database.
Step 7: If any changes are made, they are written to database.
Step 8: If a user logout, the database are synchronized to cloud storage.
Step 9: The application returns back to the login screen.

In the next section, some examples of sample management functions such as (i) add new sample, (ii) check-out sample, and (iii) move sample are introduced into the processes.

Add New Sample Process.
The user begins by navigating to 'create new' page under 'manage', and select what would you like to create. Here we chose create new sample. Select sample type by

tapping on the picture, and fill in all the information accordingly. Type in the HN Number and press ‘OK’ will show the corresponding patient’s name. If you have the barcode in front, you can also simply scan. Final step is to “choose destination” for your item. Here we chose box007 at location C5. Press confirm when complete and check history to make sure the sample has really been created.

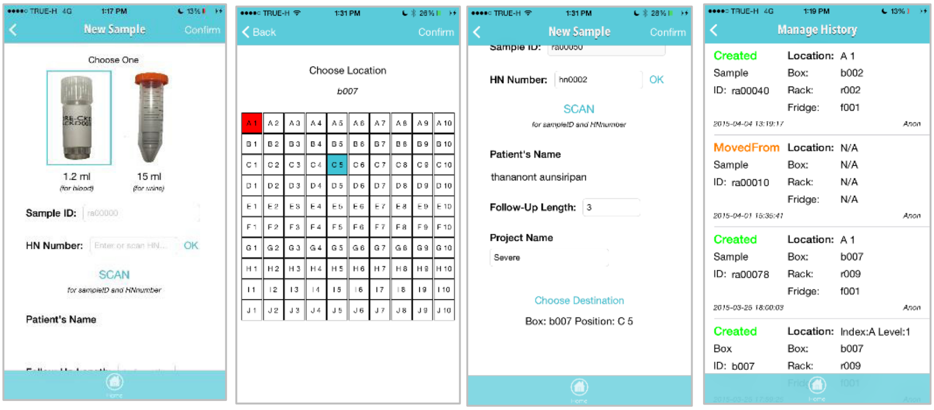


Fig. 7. Add new sample process

Check-Out Sample Process.
The user first logins using a unique ID, then the user navigates to ‘Check-Out’. Here the user’s check-out list will be presented to them and they can add samples to check-out by either scanning or manual search. If the sample’s barcode is scanned, sample needs to be valid, meaning it is a sample that exists and is not currently checked out, or it will not be added to the list. On the sample page, user can then just press ‘Check-Out’ on the sub navigation bar to add sample to list. Once the user finishes composing their list, they can press ‘Next’. In this page, user has to confirm all the samples by scanning them. This ensures that the right samples are being picked out. Then the user can ‘Submit’ and the check-out process is done.

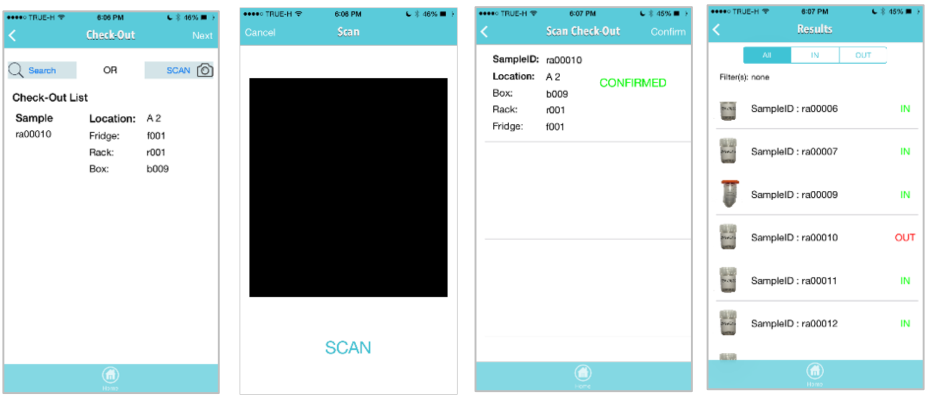


Fig. 8. Check-out sample process

Move Sample Process.
When moving, user can start the process by navigating to the 'move' feature. This will present a list of the items currently in the move list. To add an item, such as a sample, a box, or a rack; user can either search or scan. Searching will bring up a filter in which a user can find the item that they are looking for by patient's name, doctor's name, sample type, or specific ID. When the user found the item that they want to move, the user can add it to move list by tapping 'Move' in the sub navigation bar. The item will then be in the move list. When the user is done selecting all the items, they can press 'Next'. In this page, it will ask the user to 'choose destination' for all the items. Similarly to check-in, choosing a destination will bring up a list of container for the user to choose, then the location in that container. When the user completes choosing destinations for all the items, they can press 'Confirm' to complete the process. If it is all valid, the application will confirm that the samples have been moved, if not it will pop up an error message.

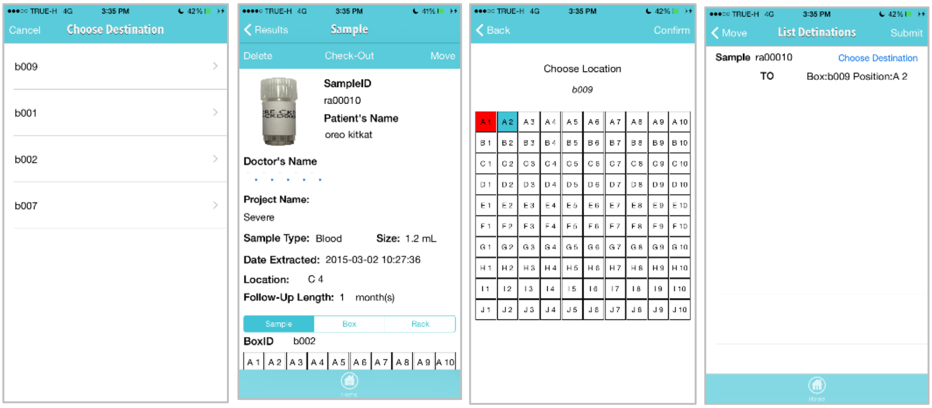


Fig. 9. Move sample process

4 System Evaluation

System evaluation is the process where users evaluate and provide feedbacks for the developed application. It is important to determine whether the application meets the requirement as stated in the objectives and scopes. Samples allocation system is designed to be used in the hospital research laboratory, which collects several samples such as blood and urine. This system is a prototype and it was demonstrated and tried only by the lab technicians from the hospital who are our stakeholders and truly understand the specimen banking concept and sample collecting process. The feedbacks are collected from them to improve the system. Some major feedbacks are concluded as follow: (i) the system is easy to use and this group of stakeholders believe that the system can help their works by reducing the human errors and save their time to manage samples (ii) scanning bar code from frozen samples at temperature below -50°C is very complicate and difficult. This is one of the limitations of mobile device's scanning which is not delicate enough to scan frozen items. Furthermore, we ran UI testing for scanning and textfield validation and process validation such as deleting, moving, checking-out, and so on. Any errors that found during this testing must be fixed to prevent any defects.

5 Conclusion

In every bio-laboratory facility lie the need for precise organization of equipment and samples stored alike. Such environment is required to perform trustworthy sample analysis. Not only a lab needs to purchase software to keep track of samples, the hardware and other needed equipment are required as well. In many cases, these costs rendered the lab impossible to operate. As a result, many labs choose alternatives such as deploying half-developed software, recording on excels, or just labeling. Such practice would, over time, cause severe disorganization, contamination, as well as disrupting the lab's workflow.

The main objective for our development is to assist the process of storing samples for lab technician through our application. By developing this application on a mobile platform, it can provide the user mobility and flexibility that boosts performance in their working environment. Moreover, the application keeps data consistent with the use of a full, but light-weighted database without any direct user interaction. Furthermore, the application also implements a barcode scanner to further eliminate most human-errors that may happen in the process. Completing our main objective, this system is a prototype that have the functions and abilities that may be use to fully implement a well-organized sample lab environment. The application can, instead of manual actions, keep track of samples stored, containers involved, the precise location of each sample, changes in location occurred, sample check in, sample check out, patients information, doctors information, project information, as well as a full history log of all the actions done in the application. This application is simulate tested in a real lab environment to help establish a well-organized lab environment as our objective stated. The test run help us to improve and develop our mobile application further through trial and error. The advantage of this the application is assist the hospital staffs into creating a more advanced and well-organized sample keeping environment. Finally we would like to express our gratitude to Assoc. Prof. Dr.Chagriya Kitiyakara, Dr. Nuankanya Sathirapongsasuti and Dr. Natini Jinawath from Ramathibodi Hospital who allowed us to observe the specimen banking and provided us with useful information.

Reference

1. Thermo Scientific. Vision Tracker Sample Management Features. Retrieved Febuary 2016, from http://www.thermoscientific.com/content/tfs/en/product/visiontracker-sample-management-software.html
2. SCANDIT. Enterprise-Grade Barcode Scanner for Mobile Apps. Retrieved Febuary 2016, from http://www.scandit.com/products/barcode-scanner/
3. Dropbox. Build the power of Dropbox into your app. Retrieved Febuary 2016 from https://www.dropbox.com/developers

MDA Approach to Automate Code Generation for Mobile Applications

Hanane BENOUDA[1], Mostafa AZIZI[1], Redouane ESBAI[1], Mimoun MOUSSAOUI[1]
[1] MATSI Laboratory, EST Oujda, Mohammed First University (UMP), Oujda, Morocco
{benouda.89, es.redouane}@gmail.com, {azizi.mos, m.moussaoui}@ump.ma

Abstract. The development of mobile applications is increasingly growing up in our environment due to the intensive use of applications in mobile devices with touch screens. However, the diversity of mobile operating systems (iOS, Windows Phone, Android, etc.), requires additional cares like code efficiency, interaction with device resources, as well as short time to market. What makes developers facing a big challenge of developing the same application for different platforms. Considering this development complexity, we directed our work to platforms using Model Driven Architecture (MDA) combined with UML, as previously used in software engineering, can offer abstraction and automation for mobile applications developers. This paper presents a MDA approach for mobile applications development including UML based modeling and code generation according to the principal (Develop Once, Use Everywhere) in order to facilitate and accelerate the development of mobile applications.

Keywords: Software engineering, MDA, UML, QVT, Acceleo, code generation, mobile application development.

1 Introduction

Currently, mobile devices become an alternative to personal computer, which are ubiquitous and accompany their users everywhere and every time. The diversity of mobile devices makes the development and maintenance of mobile applications a challenging task. Moreover, the software product diffusion is influenced by the number of platforms on which it functions. Hence portability is an essential factor for software companies in both mobile and desktop context. The variety of mobile platforms, like Android, Windows Phone and iOS, force software engineers to develop the same application targeting various Operating Systems (OS) and technologies often changing languages (such as Java, C# or Objective-C). This requires implementing different low level design for a given application specification that should be developed by different teams. All these tasks produce the increased costs of maintenance and time.

In this context some help may result from Model Driven Engineering (MDE), which provides abstraction through high level models and allows the use of modeling languages to automate the generations of applications from the model when

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3_27

variability and customizability are the inevitable requirements. Models are considered as a principal key for all project lifecycle, from requirements capture, by the modeling and developing stage, until testing.
The MDE approach is based on transformations between models, from the most abstract to the most specific one, and code generation step in the end which emits source code running with little customization. The interest for the MDE approach was increased towards the end of the last century, when the Object Management Group (OMG) had made public its initiative MDA as a restriction of the MDE.
This paper proposes the Model Driven Architecture (MDA) [1] approach for mobiles applications development. Our approach includes Model to Model transformations using QVT and Model to Text transformations using Acceleo with the aim to accelerate and facilitate the development of Android applications.
The remaining of the paper is organized as follows: the second section defines the MDA approach, while the third presents some related works. MDA for Android applications is the main topic of the fourth section. In the fifth section, we discuss the applicability of our approach through a case study.

2 Model Driven Architecture

The OMG group made public the Model Driven Architecture in 2001 as an approach for Model Driven Engineering. The principal concept of MDA [1] is to make every effort to separate functional specification of a system from its implementation on given platform. The MDA approach considers the models to transform as a productive element to be used to automatically generate the application source code.

OMG defines some terms around the models named model, meta-model and meta-meta-model, and MDA defines three levels of models (see Fig. 1), Computation Independent Models (CIM), Platform Independent Models (PIM), and Platform Specific Models (PSM) for modeling the application and the successive transformations to generate automatically source code.

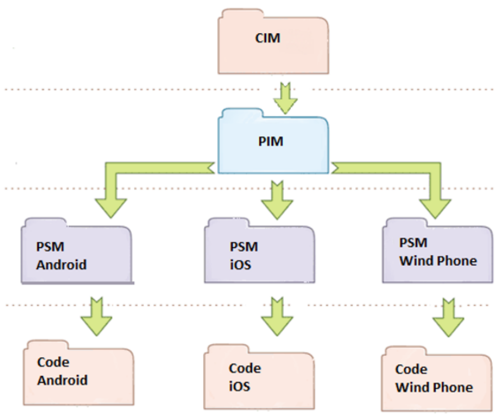


Fig. 1. MDA Models [14]

MDA approach allows the generation of target models from source models and allows for the same model to be implemented on different platforms through standardized projections. It allows interoperation of applications with models and support the development of new techniques and platforms. According to the MDA standard, the use of CIM, PIM and PSM models provides a mean to model the application and then by successive transformations to generate source code. Model transformations are composed of constraints and rules that transform elements defined in a source metamodel into other elements of target metamodel. The standard transformation language MOF QVT (Query/ View/ Transformation) [2] elaborated by OMG group has been developed to define a metamodel that allows the elaboration of model transformation. Therefore, transformations are applied to source model conforming to the source metamodel and generate a target model conforming to the metamodel. See Fig. 2.

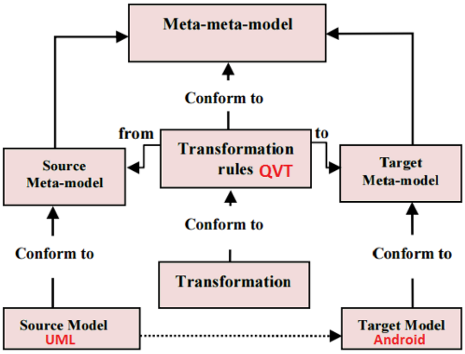


Fig. 2. Transformation Process in the MDA approach [14]

3 Related Works

MDA approach has demonstrated itself for the development of enterprise applications and can also bring a lot for mobile application development. The variety of existing mobile platforms, diversity of devices and their multiple features; gives the growing demand for better mobile devices and the difficulty in developing any type of mobile applications in such environments. The MDA approach ensures the gain productivity and sustainability of expertise while addressing the issues of fragmentation of mobile platforms. As matter of the fact, in the last years different studies have been interested in this direction.

The authors in [11] improved the generation of graphical user interfaces (GUI) for different platforms as Java Server Page Standard Tag Library (JSTL) and Java Server Faces (JSF) using the AndroMDA open source Framework. Their paper describes the approach based on the analysis and design of the PIM model by the use of the UML diagram, and then enriched with stereotypes to get the model PSM, when the PIM to PSM transformation is implemented. Then, the specific code of the target platform is generated by AndroMDA.

[14], proposes an approach to design of the user interface of mobile applications. The authors use the MDA approach to provide a PIM model under textual format and the M2M and M2T transformations are applied to generate the GUI for a specific platform; in order to do this, the researchers utilize Xtext to define a DSL and Xtend 2 to make different transformations.

In [18], paper presents a comparison of existing cross-platform tools that relieve the development efforts, rejecting the web apps which are too naïve. The solution supporting the architecture of the Xmob cross-compile is started with Xmob and ended to native code, each transformation chain contains two M2M transformations and one M2T transformations:

- PIM to PSM: transformations of high-level Domain Specific Language (DSL) conception into equivalent.
- PSM to PSM (optional): iterative introduction of good practices (various refactoring, naming convention, etc.).
- PSM to Code production of source code which corresponds to the elements in the PSM.

[21], presents an UML profile for the modeling of user interfaces. The authors use the MDA approach to define a PIM and the QVT to produce the different transformations.

In [7], authors use the implementation of the MDA standard in order to model and generate graphical interfaces for mobile platforms. This method presents the benefit of automatically generating graphical interfaces for various mobile platforms from a UML model, and the MDA transformation rules are expressed in the ATL (Atlas Transformation Language).

In our paper, we present a new approach based the MDA approach to model and generate automatically mobile applications. MDA provides a PIM model and transforms it into PSM models then generates code source for our mobile applications. The MDA transformations rules are expressed in the QVT [2] (Query/View/Transformation) and the Model to Text transformation using Acceleo [5].

4 MDA approach for Android Applications Development

Android is an OS based on the Linux Kernel designed for mobile devices like smartphones and tablets. The android applications are developed in Java, but run on a specific virtual machine named Dalvik. An android application consists of one or more activities. An activity usually presents a single visual user interface in android. It’s like window or frame of Java. Service is a particular type of activity without a visual user interface and generally run in the background for an indefinite period of time [6]. The 7 lifecycle method of Android Activity represent how activity will behave at different states. Fig. 3 presents the 7 lifecycle methods of Android activity.

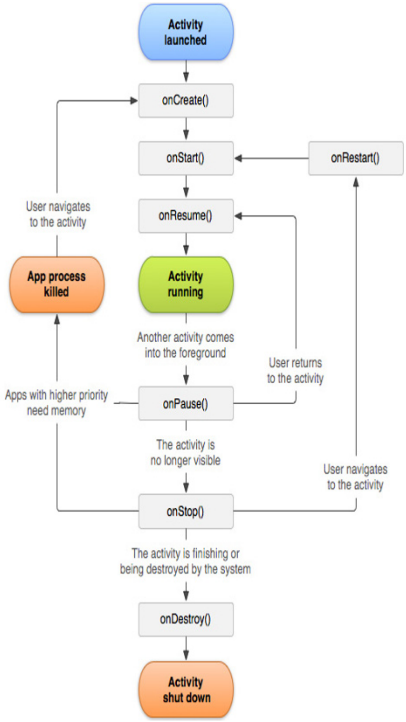


Fig. 3. Android Activity Life Cycle [6]

MDA approach for Android application development supports UML modeling and automatic code generation. This paper presents how model the structure and behavior of an Android Apps using UML class diagram. Our modeling is based on MVC (Model View Controller) pattern for implementing the android application. The Fig. 4 shows the MVC pattern which divides a software application into three interconnected parts: Model, View and controller.

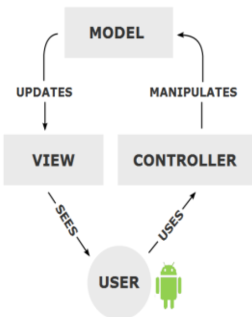


Fig. 4. Mobile MVC Architecture

The main advantage of MDA in the development of Android apps is the automation. This way, to demonstrate the automation support provides by our MDA approach. The QVT [2] tool is used to do PIM to PSM transformation and the Acceleo [5] tool is used to generate the Android code. Our android code generation is based on class diagrams, which are used by Acceleo to generate the application structure. The relationship between classes or interfaces, as association and inheritance, are respected during the code generation.

5 CASE STUDY: ERP Mobile Applications

Due to the growing pressure on IT and standardization and cost, an increasing number of enterprises have turned to Enterprise Resource Planning (ERP) systems to build their core IT system. The phenomenon is not restricted to big companies, but the Small and Medium Industries (SMI) are reaching. The companies are setting to build responsive and prompt functions at all levels. An ERP system integrates all factors of business activities like human resources, sales and distribution… and on the other hand, Customer Relationship Management (CRM) systems allows the manage and track the customer relationship. Now these days, the next generation business is transitioning to commerce on mobile devices and smart phones.

Our case study adopts a single pattern: the CRUD-pattern, which allows creating, reading, updating and deleting instances of some entity. The next figure (Fig. 5) demonstrates our Android application generated for a customer management application on an android smartphone. It represents a list of customers, where you can also create new customer, update existing ones, or delete customers which are not required.

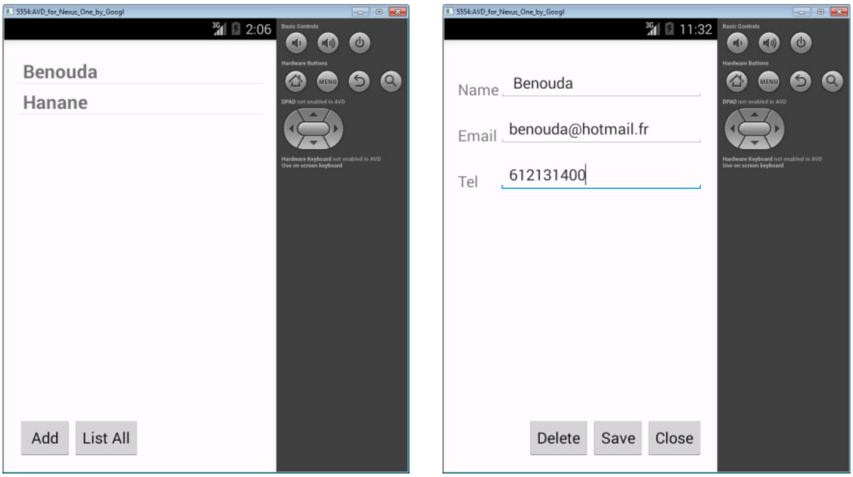


Fig. 5. Examples of screens involved with a classic CRUD-pattern for an entity

We present here the different metaclasses forming the UML source metamodel (see Fig.6) and the Android target metamodel (see Fig. 7), used to develop the algorithm of transformation between the source and target model to generate the Android source code.

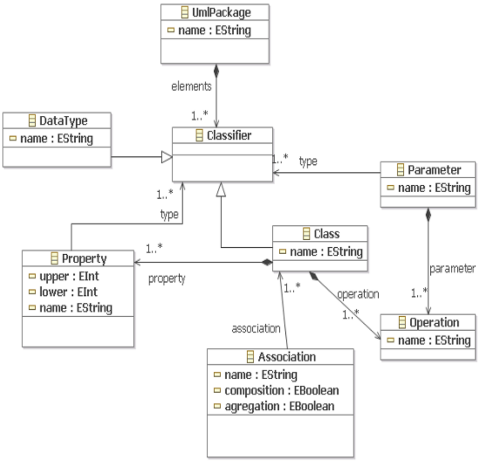


Fig. 6. Simplified UML source metamodel

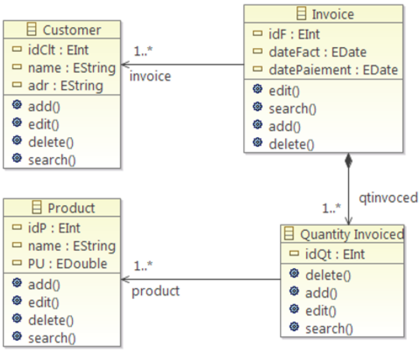


Fig. 7. Instance of UML Model

The Android metamodel facilitate the creation of PSMs. Its objective is to help the creation of models that represent applications for an Android platform. The next figure (Fig. 8) shows the Android target metamodel.

The development of many metamodel requires multiple model transformations. The M2M and M2T transformations are required to generate the necessary source code for an application that run on an Android device.

The M2M transformations are implemented with QVT. The M2T transformations are implemented using Acceleo.

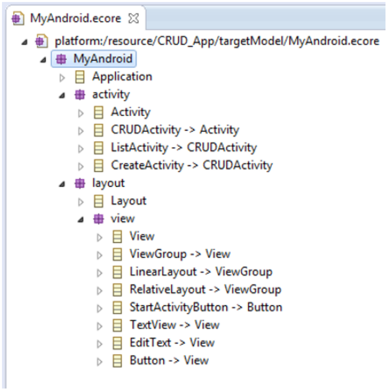


Fig. 8. Excerpt of Android target metamodel

Our application is constituted by 2 screens; first screen will display all customers in the database and the second screen is detail screen. We can automatically generate the source files in Android project using our approach, and the writing of the transformation rules itself does not produce any problems in practice. The following figure (Fig. 9) shows the transformation rules to generate the Android Manifest File. The manifest file presents essential information concerning your application to the Android system; information the system must have before it may run any of the application code.

```
[template public pack2Manifest(aPack : Package)]
<?xml version="1.0" encoding="utf-8"?>
<manifest xmlns:android="http://schemas.android.com/apk/res/android"
   package="com.hanane. [aPack.name/]"
   android:versionCode="1"
   android:versionName="1.0" >
   <uses-sdk
      android:minSdkVersion="8"
      android:targetSdkVersion="21" />
   <application
      android:allowBackup="true"
      android:icon="@drawable/ic_launcher"
      android:label="@string/app_name"
      android:theme="@style/AppTheme" >
                  [for (c:Class|aPack.elements)]
       <activity
       android:name=".MainActivity[c.name.toUpperFirst()/]"
         android:label="@string/app_name" >
                  [if (c.kind='main')]
                              <intent-filter>
```

```
            <action android:name="android.intent.action.MAIN" />
            <category android:name="android.intent.category.LAUNCHER" />
         </intent-filter>
                  [/if]
      </activity>
      <activity
         android:name=".[c.name.toUpperFirst()/]Detail"
android:label="@string/title_activity_[c.name.toLower()/]_detail" >
      </activity>
                  [/for]
   </application>
</manifest>
    [/template]
```

Fig. 9. Code generated for the AndroidManifest

5 Conclusion

Considering the variety of mobile technologies, the development of the same application for different platforms becomes a complicated and exhausting task. That's because each platform uses different tools and programming languages, what makes hard developing cross-platform applications. Thus, it requires software engineers to create applications over platforms, while ensuring the largest dissemination. The present paper proposes a MDA approach to generate Android applications based on UML class diagrams. The transformations rules were developed using the approach by modeling, QVT to browse the class diagram and then automatic code generation using Acceleo with the goal to accelerate and makes easy the development of Android applications. For future studies we will complete this code generator to automate the generation of source code for all mobiles platforms (Windows Phone, iOS …).

References

1. MDA- The Architecture of Choice for Changing World. http://www.omg.org/mda/.
2. Meta Object Facility (MOF) 2.0 Query/View/Transformation, V1.1, (2011). http://www.omg.org/spec/QVT/1.1/PDF/
3. Object Constraint Language (OCL), Version 2.4, (2014). http://www.omg.org/spec/OCL/2.4/PDF
4. Unified Modeling Language (UML), Version 2.4.1, (2011). http://www.omg.org/spec/UML/2.4.1/
5. Acceleo, http://www.eclipse.org/acceleo/.
6. Android Developers. http://developer.android.com/index.html.
7. A. Sabraoui, M.E. Koutbi and I. Khriss, "GUI Code Generation for Android Applications Using a MDA Approach", Mobile Intell. Syst. Team (MIS), Ecole Nat. Super. d'Inf. et d'Anal. Des Syst. (ENSIAS), Rabat, Morocco, IEEE, (2012) November 5-6.

8. B. Min, M. Ko, Y. Seo, S. Kuk, and H. S. Kim, “A UML metamodel for smart device application modeling based on Windows Phone 7 platform”, TENCON 2011 - 2011 IEEE Region 10 Conference, (2011) November 21-24, Bali, pp. 201-205 .
9. C. Atkinson and T. Kuhne. “Model-driven development: a metamodeling foundation”. Software, IEEE, 20(5):36 – 41, (2003) september.-october.
10. F. A. Kraemer et al., “Engineering android applications based on uml activities,” Proceedings of the 14th international conference on Model driven engineering languages and systems (MODELS’11), Jon Whittle, Tony Clark, and Thomas Khne (Eds.). Springer-Verlag, Berlin, Heidelberg, (2011), pp. 183– 197.
11. J. A. Monte-Mor, E. O. Ferreira, H. F. Campos, A. M. da Cunha, and L. A. V. Dias, “Applying MDA Approach to Create Graphical User Interfaces”, Eighth International Conference on Information Technology: New Generations, Las Vegas, NV, IEEE, (2011) April 11-13 , p.p. 766-771.
12. R. Esbai, M. Erramdani, S.Mbarki, I. Arrassen, A. Meziane, M. Moussaoui, “Transformation by modeling MOF 2.0 QVT: from UML to MVC2 web model”, INFOCOMP, v. 10, no. 3, p. 01-11, (2011) September.
13. M. Lachgar, A. Abdali, “Generating Android graphical user interfaces using an MDA approach,” Information Science and Technology (CIST), 2014 Third IEEE International Colloquium in , vol., no., pp.80,85, (2014) October 20-22.
14. M. Lachgar, A. Abdali, “MODELING AND GENERATING THE USER INTERFACE OF MOBILE DEVICES AND WEB DEVELOPMENT WITH DSL,” Journal of Theoretical and Applied Information Technology, 10th February 2015, Vol. 72. No. 1 – (2015).
15. Parada, G. Abilio, and Lisane B. de Brisolara. “A model driven approach for Android applications development.” Computing System Engineering (SBESC), 2012 Brazilian Symposium on. IEEE, (2012).
16. Son, Hyun Seung, Woo Yeol Kim, and Robert Young Chul Kim. “MOF based Code Generation Method for Android Platform.” International Journal of Software Engineering & Its Applications 7.3 (2013).
17. Yoon, Hyeon-Ju. “A study on the performance of Android platform.” International Journal on Computer Science and Engineering 4.4 (2012): 532-537.
18. Olivier Le Goaer , Sacha Waltham, “Yet another DSL for cross-platforms mobile development”, Proceedings of the First Workshop on the Globalization of Domain Specific Languages, p.28-33, (2013) July 01-01 , Montpellier, France [doi>10.1145/2489812.2489819].
19. I. Madari, L. Lengyel, and T. Levendovszky, “Modeling the User Interface of Mobile evices with DSLs”, 8th International Symposium of Hungarian Researchers on Computational Intelligence and Informatics, Budapest, Hungary, (2007) November 15-17, pp. 583-589.
20. K. Minhyuk, S. Yong-Jin, M. Bup-Ki, K. Seunghak, and K. Hyeon Soo, “Extending UML Meta-model for Android Application”, 2012 IEEE/ACIS 11th International Conference on Computer and Information Science, (2012) May 30 -June 1, Shanghai, China, pp.669-674.
21. R. Mannadiar, and H. Vangheluwe, “Modular synthesis of mobile device applications from domain-specific models”, Proceedings of the 7th International Workshop on Model-Based Methodologies for Pervasive and Embedded Software, New York, USA, (2010), pp. 21-28.
22. S. Link, T. Schuster, P. Hoyer, and S. Abeck, “Focusing Graphical User Interfaces in ModelDriven Software Development", First International Conference on Advances in Computer-Human Interaction”, Sainte Luce, (2008) February 10-15, pp. 3-8.

Author Index

K.J. Kim et al. (eds.), *Mobile and Wireless Technologies 2016*,
Lecture Notes in Electrical Engineering 391, DOI 10.1007/978-981-10-1409-3